



**COORDINACIÓN DE DESARROLLO
PROFESIONAL E INSTITUCIONAL**

SUBCOORDINACIÓN DE POSGRADO

T E S I S

**DISEÑO E IMPLEMENTACIÓN DE UNA GEOBASE DE DATOS
DISTRIBUIDA, EN UN AMBIENTE VIRTUALIZADO, ACOPLADA A UN
SISTEMA DE INFORMACIÓN GEOGRÁFICA**

que para obtener el grado de

Doctor en

Ciencias y Tecnología del Agua

presenta

M. en C. Carlos Raúl Montaña Espinosa

Tutor: Dr. Julio Sergio Santana Sepúlveda

Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimoniales de la obra titulada "**Diseño e implementación de una geobase de datos distribuida, en un ambiente virtualizado, acoplada a un sistema de información geográfica**", otorgo de manera gratuita y permanente al Instituto Mexicano de Tecnología del Agua, autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre su personal, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación.

M. EN C. CARLOS RAÚL MONTAÑO ESPINOSA

Jiutepec, Morelos a de diciembre de 2014.

Firma

Lo que sabemos es una gota de agua; lo que ignoramos es el océano.

Isaac Newton

Excelente maestro es aquel que, enseñando poco, hace nacer en el alumno un deseo grande de aprender.

Arturo Graf (1848-1913) Escritor y poeta italiano

Visión del Sector Hídrico

“Ser una nación que cuente con seguridad en el suministro del agua que requiere para su desarrollo, que la utilice de manera eficiente, que reconozca su valor estratégico y económico, que proteja los cuerpos de agua y preserve el medio ambiente para las futuras generaciones”

Agradezco al Dr. Julio Sergio Santana Sepúlveda, su invaluable ayuda en la dirección de este trabajo y su gran apoyo para verlo culminado.

Agradezco al Dr. Héctor Sanvicente Sánchez, al Dr. Iván Rivas Acosta y al Dr. Víctor Manuel Arroyo Correa sus comentarios y correcciones al contenido de este trabajo.

Hago un agradecimiento especial a las autoridades de la Comisión Nacional del Agua por su invaluable ayuda en la realización de mis estudios de Posgrado.

Dedico el presente trabajo:

A mis padres, L.A.E. Carlos Raúl Montaña Hernández y Sra. Lucila Espinosa de Montaña, quienes siempre me han brindado todo su apoyo y amor para lograr esta nueva meta, ver terminado mi Doctorado en Ciencias.

Con amor a mi esposa, Lic. Angélica Isabel Jover de Montaña, quien siempre ha estado a mi lado brindándome amor y fortaleza para lograr mis objetivos. Gracias por tu paciencia, tu gran comprensión y apoyo.

A mi hermano M. en F. Roberto y mis hermanas, Lilian y Gisela, gracias por su apoyo y amor fraternal.

ÍNDICE

Resumen.....	1
Introducción	5
Capítulo 1.0. PLANTEAMIENTO DEL PROBLEMA Y CONCEPTOS BÁSICOS	
1.1. Planteamiento del problema	9
1.2. Justificación.	10
1.3. Objetivos	15
1.4. Hipótesis	16
1.5. Marco teórico-conceptual y estado del conocimiento	
1.5.1. Marco teórico-conceptual.....	16
1.5.2. Virtualización de hardware.....	17
1.5.3. Virtualización en el Sistema Operativo.....	20
1.5.4. Virtualización parcial (Address Space Virtualization).	21
1.5.5. Virtualización de red.....	22
1.5.6. Virtualización de aplicaciones.....	27
1.5.7. Virtualización de Bases de Datos.....	28
1.5.8. Almacenamiento de grandes volúmenes de datos(Big Data). ...	31
1.5.9. Geobases de datos distribuidas.	34
1.5.9.1. Bases de datos habilitadas geoespacialmente.....	34
1.5.9.2. Enfoque al problema de diseño de la base de datos distribuida	40
1.5.9.3. Tipos de geobases de datos distribuidas	41
1.5.9.4. Implementación de cada tipo de arquitectura	48
1.5.9.5. Geobase de datos distribuidas operando en servidores virtuales.....	52
1.5.9.6. Ventajas y desventajas de cada tipo de arquitectura	54
1.5.9.7. Plataformas sobre las que operan las bases de datos distribuidas.....	58

1.5.9.8 Ventajas de utilizar ArcGIS Server en un sistema virtualizado.	65
1.5.9.9. Estudios de caso.	66
1.6. Identificación de elementos relacionales	
1.6.1. Diseño metodológico	67
1.6.2. Medición.	69
1.6.3. Justificación técnica	70
1.6.4. Limitantes contextuales	72
Capítulo 2.0. ANTECEDENTES, FUNDAMENTOS LEGALES Y ETAPAS DEL PROYECTO	
2.1. Antecedentes para la creación del Comité Técnico Estadístico y de Información Geográfica del Sector Medio Ambiente y Recursos Naturales.	75
2.2. Fundamentos legales, modificaciones y reformas a la Ley de Aguas Nacionales.	78
2.2.1. El Grupo Temático del Agua.	80
2.3. Vinculación entre bases de datos geográficas.	84
2.4. Proyectos piloto.	85
Capítulo 3.0. IMPLEMENTACIÓN DE LA GEOBASE DE DATOS INSTITUCIONAL DEL AGUA EN UN AMBIENTE DISTRIBUIDO OPERANDO SOBRE SERVIDORES VIRTUALIZADOS	
3.1. Método a implementar.	93
3.2. Implementación del método	
3.2.1. Recopilación de requerimientos.	94
3.2.2. Análisis de requerimientos.	97
3.2.3. Uso empresarial de la Geobase de Datos.	103
3.3. Implementación de la primera propuesta de solución	
3.3.1. Modelo conceptual.	112
3.4. Replicación de geobases de datos entre Unidad Central y Organismos de Cuenca.	113
3.5. Medición de la funcionalidad de la alternativa de solución.	114

Capítulo 4.0. LINEAMIENTOS, APORTACIONES, MEJORAS Y ACCIONES A FUTURO

4.1. El cambio que se requiere para la interoperabilidad para el trabajo de geobases de datos distribuidas.....	117
4.2 Políticas de geobases de datos distribuidas, estándares, retorno de inversión y geoinnovaciones. Iniciativas a futuro.	120
4.2.1. Estándares en Geobases de Datos Distribuidas.	120
4.2.2. El retorno de la inversión ROI.....	123
4.2.3. Geoinnovaciones.....	125
4.3. Iniciativa de Geoinformación Gubernamental Sostenible.....	126

Capítulo 5.0. CARACTERÍSTICAS, ANÁLISIS DEL RENDIMIENTO, BENEFICIOS Y ESPECIFICACIONES DEL SISTEMA PROPUESTO

5.1. Software de prueba utilizado para el análisis del rendimiento.	127
5.1.1. Pruebas de carga.....	134
5.1.2. Pruebas de estrés.....	137
5.2. Sobre la geobase de datos.....	144
5.2.1. Métodos de revisión y validación requeridos para cada tipo de información e implementados en las ETL (Extracción-Transformación-Carga)	148
5.2.2. Mejores indicadores de que la geobase está funcionando adecuadamente	150
5.2.3. Procedimientos de actualización y mantenimiento de la geobase para garantizar su actualización a lo largo del tiempo	164

Capítulo 6.0. CONCLUSIONES Y RECOMENDACIONES

6.1. Sobre el tema de la tesis.....	167
6.2. Conclusiones generales.....	173
6.3. Conclusiones particulares.....	178
Glosario.....	181
Lincografía	187
Bibliografía	189

Resumen

La Comisión Nacional del Agua (CONAGUA) es una dependencia desconcentrada del Gobierno Federal, cuya misión es “administrar y preservar las aguas nacionales y sus bienes inherentes, para lograr su uso sustentable, con la corresponsabilidad de los tres órdenes de gobierno y la sociedad en general”¹.

Al trabajar con un recurso natural con una amplia y heterogénea distribución sobre el territorio nacional, la Institución requiere conocer la cantidad, calidad y usos del agua en cada una de las zonas geográficas donde esta se encuentra, permitiéndole una mejor administración del vital líquido. Entre las herramientas que la Institución emplea para cumplir con su objetivo, está la cartografía digital a través de la cual se identifican y clasifican las distintas regiones hidrológicas, administrativas, cuerpos de agua y demás recursos hídricos existentes en el país, así como la información documental, estadística y financiera. Estos tres tipos denominados alfanuméricos.

Para contar con información geográfica vinculada a las grandes bases de datos alfanumérica de la Institución; desde hace más de una década, la Institución se dio a la tarea de generar un acervo cartográfico digital y almacenarlo en una geobase de datos² la cual es factible vincularla con bases de datos alfanuméricas, implementadas en distintas fuentes. Esta actividad recayó en la Subgerencia denominada Sistema de Información Geográfico del Agua (SIGA); que ha tenido como tareas primordiales la recopilación, manejo, gestión y análisis de datos espaciales. Su misión es “administrar y normar en Geomática, así como en Sistemas de Información Geográfica, a fin de permitir el análisis y consulta de información relacionada a la ubicación de los recursos hidrológicos dentro de su ámbito geográfico, considerando tanto entidades gráficas como estadísticas, apoyando a todas las áreas que conforman la Comisión Nacional del Agua para lograr el uso sustentable del recurso”³. No obstante la misión del SIGA, algunos de sus objetivos específicos principales son la administración del acervo cartográfico digital, servicios de consulta y visualización de datos georreferenciados⁴, mediante herramientas de consulta de cartografía digital

¹<http://www.conagua.gob.mx/Contenido.aspx?id=5582b1c3-9e5c-49ba-ad38-2f2a3b30e7a8|CONÓCENOS|1|0|0|0|0>

² Base de datos implementada sobre un sistema administrador de base de datos relacional, el cual permite, a través de un motor de consulta geográfica (ArcSDE), almacenar datos cartográficos y vincularlos tanto con tablas de atributos como con otras bases de datos relacionales las cuales pueden ser geoespaciales o alfanuméricas.

³ <http://siga.conagua.gob.mx/Contenido/Acerca/MisionSIGA.aspx>

⁴ Cualquier elemento hídrico puede ser asociado a sus coordenadas geoespaciales (longitud y latitud) a fin de

interactiva vinculados a muy diversas fuentes y temas hídricos. Es importante mencionar que para lograr la adecuada administración de esta acervo cartográfico de una forma eficiente, y con base en el artículo 12 bis 6 de la LAN⁵, el SIGA debe apoyarse en el trabajo realizado a nivel regional por los Organismos de Cuenca⁶ (OC).

Para hacer frente a estas actividades a nivel nacional, según lo señalado en el párrafo anterior, el SIGA ha tenido que enfrentar un problema complejo debido a la falta de personal en los OC, lo cual ha hecho difícil la administración de la información geográfica en ámbitos locales. Es por esta razón que ha surgido la necesidad de contar con geobases de datos distribuidas que permitan brindar acceso a este tipo de información al personal que constituye cada uno de los OC. Cabe señalar que la administración de estas geobases de datos, requiere de personal con la suficiente capacidad técnica tanto en el manejo de los sistemas de información geográfica como en la operación de servidores geoespaciales; sin embargo, es difícil contar con tales personas debido a la falta de presupuesto y a las políticas internas de la Institución para la contratación de personal todo ello asociado al problema que implica la adquisición o renta de equipo de cómputo nuevo dado que, debido a que las políticas administrativas de TICs⁷ en la CONAGUA deben alinearse a lo dispuesto en el Decreto de Austeridad en materia de TICs⁸ derivado del Decreto que establece las medidas de austeridad y disciplina del gasto de la Administración Pública Federal⁹. Así, con la ayuda de servidores virtualizados¹⁰ es como se pretende apoyar en el manejo de la información geográfica administrada por la CONAGUA dentro del ámbito nacional. De esta forma, se contaría con una red “distribuida” de servidores geoespaciales vinculados a la Geobase de Datos central permitiendo la migración a servidores físicos conforme se cuente con el personal necesario para su operación y puesta en marcha que, en caso de no poder operar en forma física, se pueden seguir trabajando desde el entorno central en forma remota.

ubicarlo de forma precisa sobre la superficie terrestre y, en particular, sobre la República Mexicana.

⁵cf. Ley de Aguas Nacionales.

⁶ Ley de Aguas Nacionales (LAN), artículo 12 bis 1. Los Organismos de Cuenca, en las regiones hidrológico - administrativas son unidades técnicas, administrativas y jurídicas especializadas, con carácter autónomo que la LAN les confiere, adscritas directamente al Titular de la CONAGUA, cuyas atribuciones, naturaleza y ámbito territorial de competencia se establecen en la LAN y se detallan en sus reglamentos, cuyos recursos y presupuesto específicos son determinados por la CONAGUA.

⁷ Tecnologías de Información y Comunicaciones

⁸ <http://dgi.sedesol.gob.mx/index/index.php?sec=40>

⁹ <http://www.funcionpublica.gob.mx/unaopspfd/doctos/adquisiciones/auste041206.pdf>

¹⁰ cf. pág. 16

Nota.

Las referencias bibliográficas se presentarán en formato estándar. Las lincográficas se presentarán en formato [ref. n] en donde n hace referencia al número del índice con el que se encuentra registrado el elemento en el apartado respectivo.

Introducción

En el periodo comprendido entre los años de 1990 y 2000 surgieron nuevas tendencias en la forma de utilizar tanto la información alfanumérica como los mapas para la evaluación de recursos, en particular del agua y la planificación del uso de la tierra¹¹. Al comenzar a gestarse el cambio de paradigma hacia el análisis de información vinculada al ámbito geográfico y al observar que no eran independientes entre sí, se empezó a reconocer la necesidad de evaluarlos de una forma integrada y multidisciplinaria. Durante este periodo la tecnología del uso de computadoras progresó rápidamente en cartografía y se perfeccionaron cientos de sistemas informáticos para distintas aplicaciones cartográficas y de bases de datos. Al mismo tiempo, se estaba avanzando en una serie de sectores conexos, entre ellos: la edafología, la topografía, la fotogrametría y la percepción remota. En un principio, este rápido ritmo de desarrollo provocó una gran duplicación de esfuerzos en las distintas disciplinas conexas así como entre los bancos de información de aquellas dependencias que administran el sector de recursos naturales, pero a medida que se multiplicaban los sistemas y se adquiría experiencia, surgió la posibilidad de articular los distintos tipos de elaboración automatizada de datos espaciales, reuniéndolos en verdaderos sistemas de información geográfica para fines generales. Para principios del nuevo milenio, los sistemas de información y, en particular los sistemas de información geográfica (SIG) se habían convertido en herramientas fundamentales en el análisis de información relativa a la administración del recurso hídrico. Actualmente los sistemas de información alfanumérica como los SIG se implementan rápidamente en los organismos públicos, los laboratorios de investigación, las instituciones académicas, la industria privada y las instalaciones militares. La idea de la implementación de una de estas herramientas de análisis geoespacial, dentro de la CONAGUA, se hizo realidad en el SIGA. Dadas las atribuciones que le confiere el Reglamento Interior de CONAGUA¹², en su artículo 81 inciso V y lo señalado en el artículo 82 inciso XII, XI y XVIII; el SIGA está encargado de “Administrar y normar en Geomática, así como en Sistemas de Información Geográfica, a fin de permitir el análisis y consulta de información relacionada a la ubicación de los recursos hidrológicos dentro de su ámbito geográfico, considerando tanto entidades gráficas como

¹¹ <http://revistas.ucm.es/index.php/RGID/article/viewFile/RGID9797120093A/10990>

¹² <http://www.conagua.gob.mx/CONAGUA07/Noticias/ReglamentoInteriorCONAGUA.pdf>

estadísticas, apoyando a todas las áreas que conforman la Comisión Nacional del Agua para lograr el uso sustentable del recurso”.¹³

Partiendo de los objetivos e iniciativas plasmadas en la Agenda del Agua 2030¹⁴, de los objetivos planteados en la Ley de Agua Nacionales (LAN), en particular en sus artículos 7 bis inciso II y IV, 9 inciso XLVII, 12 bis 1, bis 4 y bis 6 inciso XXIX así como apoyo a los referido en el párrafo primero del artículo 14 bis; para su cabal cumplimiento, se enuncian las siguientes iniciativas las cuales sustentan la elaboración del presente trabajo:

En la Agenda del Agua 2030:

Para el logro de cuencas y acuíferos en equilibrio es necesario:

- 1) Robustecer las funciones de gobierno de CONAGUA y su organización, en particular las asociadas a los Organismos de Cuenca.
- 2) Iniciativas de carácter general:
 - a) Aplicar la evaluación orientada a resultados a todos los programas públicos que incidan o afecten el logro de la sustentabilidad hídrica.
 - a) Desarrollar sistemas regionales de información para reforzar la gestión del agua por cuenca y acuífero.

En la LAN:

- 3) Se declaran de interés público:
 - a) La descentralización y mejoramiento de la gestión de los recursos hídricos por cuenca hidrológica, a través de Organismos de Cuenca de índole gubernamental.
 - b) El mejoramiento permanente del conocimiento sobre la ocurrencia del agua en el ciclo hidrológico, en su explotación, uso o aprovechamiento y en su conservación en el territorio nacional, y en los conceptos y parámetros fundamentales para alcanzar la gestión integrada de los recursos hídricos, así como la realización periódica de inventarios de usos y usuarios, cuerpos de agua, infraestructura hidráulica y equipamiento diverso necesario para la gestión integrada de los recursos hídricos.
- 4) Atribución de la CONAGUA, en su Nivel Nacional:
 - a) Integrar el Sistema Nacional de Información sobre cantidad, calidad, usos y conservación del agua, con la participación de los Organismos de Cuenca, en coordinación con los gobiernos de los estados, el Distrito Federal y con los Consejos de Cuenca en concordancia con la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental.

¹³ <http://siga.cna.gob.mx/Contenido/Acerca/MisionSIGA.aspx>

¹⁴ <http://www.conagua.gob.mx/CONAGUA07/Temas/AgendadelAgua2030.pdf>

En particular la iniciativa 3.a) implica contar con una infraestructura de TI¹⁵ la cual permita el manejo integrado del agua, desde el ámbito de los OC hacia unidad central de CONAGUA así como a la inversa. Para ello es necesario el trabajo coordinado, en equipo, entre éstos y las gerencias que se ubican en oficinas centrales para que, con base en un análisis detallado de la información que poseen, se implementen líneas de acción así como estrategias enfocadas a contribuir al bienestar social y al desarrollo sustentable del recurso hídrico sin descuidar la preservación del medio ambiente. Para el efecto es necesario el intercambio de información entre los OC y oficinas centrales de una forma automática, expedita, segura, confiable y de calidad. Para avanzar en este sentido es que surge la necesidad de generar una red que vincule a los principales bancos de información, tanto alfanuméricos¹⁶ como geográficos¹⁷, de cada una de las áreas que conforman a la CONAGUA.

Esta implementación permitirá distribuir la información referente al recurso hídrico a través de medios de TI para presentar un conjunto único de datos relativos a la cantidad, calidad, usos y conservación del agua dentro del ámbito nacional. Esto facilitará el fortalecimiento de los vínculos de trabajo entre los OC y oficinas centrales de la CONAGUA; propiciando la confianza de la sociedad en esta última al contar con la seguridad de que la información consultada es única, oficial y con un alto grado de seguridad de que el sistema opera exitosamente en el ambiente distribuido; trascendiendo límites políticos y geográficos permitiendo con ello cumplir con compromisos de desarrollo, mejorando la cooperación, colaboración y el compromiso de la gestión integrada del agua.

El volumen, la diversidad y la complejidad de datos básicos que se requieren para la administración del recurso hídrico exigen del uso de sistemas de información eficientes capaces de apoyar en forma sustantiva el proceso de tomas de decisión. Un sistema de información hídrico unificado es, desde la perspectiva del INEGI, "...un conjunto de hardware, software, información y procedimientos computarizados, que permite y facilita el análisis, gestión o representación de datos del recurso hídrico a las Instituciones de gobierno que gestionan el recurso. Este sistema unificado debe funcionar como si fuese una única base de datos con información tanto alfanumérica como geográfica en donde se encuentre representado cada elemento por un identificador único pero común a cada Institución" (ref. [5]). De esta forma, señalando un objeto se deben conocer sus atributos, y a la inversa, preguntando por un registro de la base de

¹⁵ Tecnologías de Información.

¹⁶ Datos en forma de texto o números organizados en forma tabular.

¹⁷ Gráficos de un mapa en formato digital.

datos se debe poder saber su localización en la cartografía. Al asociar los datos a un sistema de información alfanumérica y geográfica es posible separar la información en diferentes capas temáticas y almacenarlas en forma independientemente, permitiendo trabajar con ellas de manera rápida y sencilla facilitando al profesional la posibilidad de relacionar la información existente a través de la topología de los objetos.

Es muy importante señalar que la iniciativa para el desarrollo de este proyecto surge debido a que actualmente se pueden encontrar múltiples versiones no compatibles de la misma información dentro de la Institución y, por ende, surge la necesidad de contar con un banco de único de datos que permita contar con toda la información geoespacial en el ámbito de su área de competencia¹⁸ pero que al mismo tiempo forme parte de la información geoespacial oficialmente aceptada dentro de la CONAGUA soportada en un modelo cliente-servidor virtualizado lo cual ofrecerá la posibilidad de compartir la información dentro de la organización sobre una Infraestructura de Datos Espaciales (IDE).

¹⁸ En el caso de CONAGUA la administración de los recursos hídricos.

Capítulo 1. PLANTEAMIENTO DEL PROBLEMA Y CONCEPTOS BÁSICOS

Resumen del capítulo.

Soportado sobre la infraestructura de cómputo con la que cuenta la CONAGUA establecida en la Red Nacional de Datos (RND), una red digital que utiliza enlaces para interconectar los diversos Organismos de Cuenca y Direcciones Locales, hace factible la interoperabilidad de aplicaciones y vinculación entre bases de datos distribuidas permitiendo un amplio acceso a sus bancos de información. Este logro importante permitirá la conectividad instantánea entre un gran número de personas al interior de CONAGUA. La interconexión de bases de datos geoespaciales de OC con la Geobase de Datos del Agua (GeoAgua) será un claro ejemplo de interoperabilidad entre sistemas distribuidos sobre Sistemas de Información Geográfica (SIG) en un ambiente basado en los protocolos de la Internet. Este trabajo pretende ser una posible solución, definitiva, al problema de interconexión entre bases de datos geoespaciales operando en la RND de CONAGUA.

En el presente capítulo, después de una justificación y descripción del marco teórico, se da una delineación del proyecto así como la presentación del diseño metodológico. Se provee al lector con una referencia inicial de lo que representa una geobase de datos; adicionalmente se definen algunos conceptos básicos, a fin de facilitar su ulterior lectura. Posteriormente, se discute brevemente el estado del conocimiento relativo a la virtualización y geobases de datos distribuidas en sistemas de información geográfica así como algunos estudios de casos en los cuales la virtualización se constituyó como elemento esencial para permitir el éxito de algunas iniciativas de e-gobierno¹⁹. Finalmente, se presenta una justificación técnica de este trabajo así como una descripción de las limitantes del mismo.

1.1. Planteamiento del problema

Las atribuciones y responsabilidades conferidas a la CONAGUA por la LAN, la gran cantidad de información ligada a su ubicación geográfica, la complejidad de los análisis requeridos para la toma de decisiones y la necesidad de conocer mejor las fases del ciclo hidrológico en los que la componente geográfica juega

¹⁹ El Gobierno Digital, conocido como e-gobierno, es un componente del Sistema Nacional e-México, que promueve el uso intensivo de sistemas digitales, en especial de *Internet*, como la herramienta principal de trabajo de las unidades que conforman la APF, a través de siete líneas de acción [ref. http://www.oas.org/juridico/spanish/mesicic2_mex_VIIG_bg_gd_sp.pdf]

un papel primordial, hacen indispensable la vinculación de las bases de datos de la CONAGUA con los respectivos de OC los cuales incluyan la dimensión geográfica a fin de unificar datos y permitir el desarrollo en forma modular distribuida de la consulta de datos a nivel nacional.

Los sistemas implementados en la CONAGUA se han adquirido, generalmente, determinados por la necesidad de contar la mejor implementación para satisfacer un objetivo específico. Como resultado de esto, se crea una amplia gama de islas de datos sin que, aparentemente, exista una manera sencilla de desbloquear los valiosos activos de información que todas ellas contienen en su conjunto, para dar soporte a procesos más útiles y productivos.

Por otro lado, la distribución de los OC dentro del territorio nacional implica el uso de un conjunto de medios que permitan la transmisión remota y almacenamiento de la información de forma ágil y confiable.

Un acercamiento bien estructurado a la virtualización y las geobases de datos distribuidas podría ser útil para vincular almacenes de datos dispersos en forma geoespacial y permitir que la información sea intercambiada más fácilmente. Así, las aplicaciones de gestión de la información podrían entonces aprovechar toda aquella información integrada para proporcionar una mejor percepción, un mayor control y una eficiencia operacional mejorada. Como resultado final, el usuario podría estar mejor informado, facilitando la toma de decisiones haciéndolas más oportunas, todo ello relacionado con una mejor eficacia en los costos.

El equipo de cómputo y la infraestructura de telecomunicaciones, con las que cuenta CONAGUA, hoy en día, hacen que los sistemas de información avancen sustancialmente permitiendo obtener la información proveniente de cualquier parte del país con mayor oportunidad. La CONAGUA cuenta, en la actualidad, con medios de comunicación y procesamiento de datos mediante los cuales se puede consultar y analizar de forma eficaz, mediante programas de cómputo, la información tanto geográfica²⁰ como la alfanumérica en formatos digitales.

1.2. Justificación.

A principios del año 2013 el directorado de CONAGUA planteó la posibilidad de la creación de un sistema de información que permitiese de una forma sencilla y amigable realizar consultas relativas a toda la información que se encuentra disponible a través de los diferentes bancos de datos de diversas áreas que

²⁰ Entendido como la ubicación de los recursos hidrológicos dentro de su entorno geográfico.

conforman la CONAGUA a fin de que éste sirviese de apoyo a la toma de decisiones así como a la atención y seguimiento en caso de que se presenten, en cualquier parte del territorio nacional, alguna contingencia debido a la afectación de algún fenómeno hidrológico extremo. Dado este compromiso, en la Subdirección General Técnica se planteó la creación del prototipo del sistema y/o mecanismos de vinculación de datos. Este trabajo fue solicitado al autor de este documento quien generó el modelo conceptual del arquetipo antes señalado así como para la implantación del mismo. Como primer propuesta se creó una presentación la cual consistió en la explicación preliminar del estado actual del manejo de la información así como de los objetivos a lograr en el futuro con el desarrollo de un sistema geoespacial para la toma de decisiones (SDSS) el cual opere como eje central en la integración de toda la información con la que cuenta CONAGUA; toda vez que el manejo del recurso hídrico, por propia naturaleza, debe estar basado en un manejo geoespacial. En la Figura 1 se presenta el diagrama del SDSS como parte de la presentación generada. Con la implementación de un vínculo de bases de datos y sistemas de información entre OC y la unidad central de CONAGUA para la consulta de información geoespacial, será posible que una mayor cantidad de personas que laboran en esta Institución puedan contar con información automática, expedita, segura, confiable y de calidad del acervo de datos, a nivel nacional, con el que se cuenta.

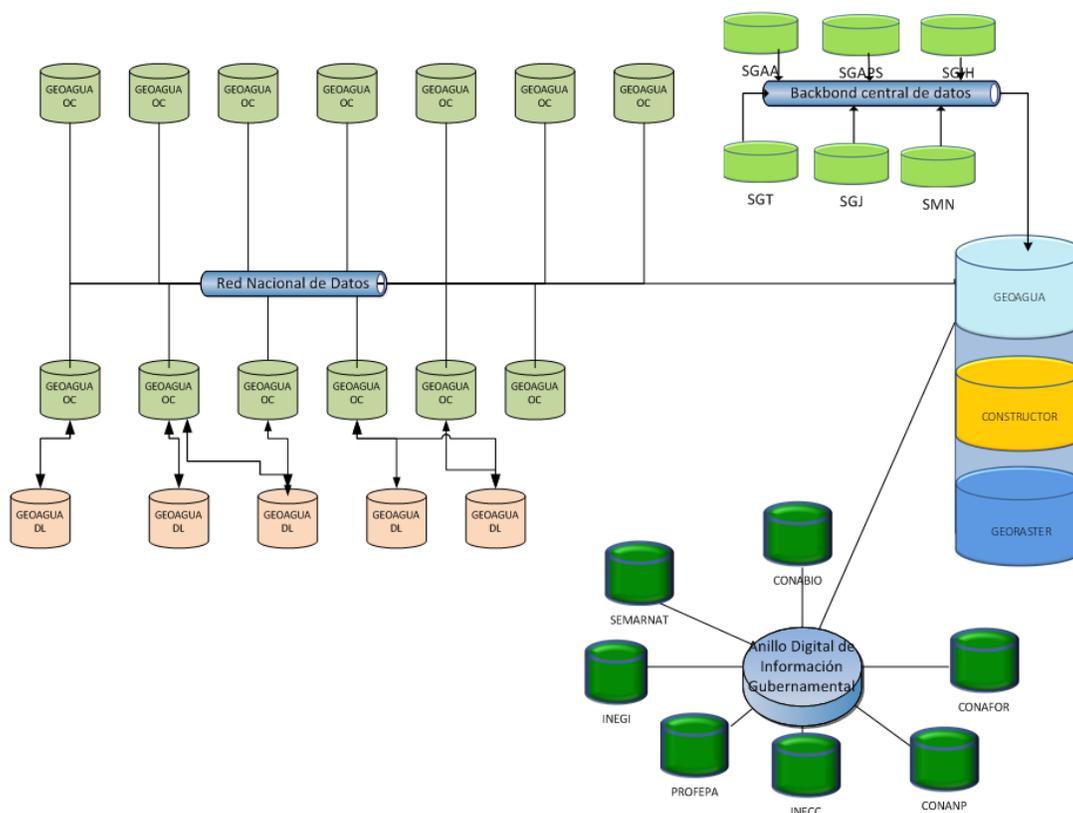


Figura 1. Diagrama del un sistema geoespacial para la toma de decisiones (SDSS)

Mediante el aprovechamiento de las mejores prácticas de virtualización de geoservidores y geobases de datos distribuidas, la CONAGUA será capaz de alcanzar un intercambio de datos mejorado, así como un acceso integrado a sistemas y aplicaciones. El resultado, será el empleo más eficaz de la información geoespacial y alfanumérica dentro de la propia Institución.

El emplear este tipo de vínculo de datos, hará posible la mejora de las acciones de planeación y programación que realiza tanto la unidad central de CONAGUA así como los OC distribuidos en el territorio nacional, cuyo resultado final será la administración de la información geoespacial de una mejor manera mediante su gestión a través de bases de datos virtualizadas distribuidas que permitan su uso como si estuviesen operando en el ámbito local. Entre las acciones benéficas destacan las relacionadas con la administración de infraestructura hidráulica existente en cualquier región del país, permitiendo con ello ubicarla con mayor exactitud sobre la superficie terrestre y apreciar su interrelación con otros aspectos geográficos importantes de la zona.

Otra bondad de la implementación será la estandarización de los aspectos tanto cartográficos como de datos alfanuméricos, es decir el uso de geobases de datos distribuidas operando sobre geoservidores virtualizados permitirá combinar cartografía digital y datos de naturaleza numérica, de texto o multimedia cuyas características geospaciales estén almacenadas en una base de datos geoespacial. En la Tabla 1 se muestra un comparativo entre la situación imperante antes y después de la implementación del sistema.

Tabla 1. Comparación de situación anterior y posterior a la implantación del sistema

Situación	Antes de la existencia del sistema	Después de la existencia del sistema
<p>Beneficios en el desarrollo de las tareas diarias de la Institución</p>		<ol style="list-style-type: none"> 1. Existencia de vínculo entre las bases de datos geospaciales regionales con la geobase a nivel central lo cual facilitará el despliegue de toda la cartografía básica generada por el SIGA. 2. Acceso a otros bancos de datos intrainstitucionales mediante llamadas SQL, con la facilidad de poder enviar los resultados de la consulta a usuarios remotos mediante acceso vía herramientas API. Casos concretos: SEMARNAT, INEGI y Secretaría de Salud en Mérida. 3. Intercambio de información entre los OC y oficinas centrales de una forma automática, expedita, segura, confiable y de calidad.

		<ol style="list-style-type: none"> 4. Alto grado de seguridad de que el sistema opera exitosamente en un ambiente distribuido; trascendiendo límites políticos y geográficos permitiendo con ello cumplir con compromisos de desarrollo, mejorando la cooperación, colaboración y el compromiso de la gestión integrada del agua. 5. Mejor desempeño y balance de cargas. 6. Reducción de carga de red. 7. Particionamiento de datos geoespaciales. 8. Cómputo desconectado. 9. Uso empresarial de la geobase de datos. 10. Consolidación de servidores (reorganización) 11. Arquitectura de alta disponibilidad.
<p>Perjuicios en el desarrollo de las tareas diarias de la Institución</p>	<ol style="list-style-type: none"> 1. Inexistencia de vínculos entre las bases de datos geoespaciales regionales y a nivel central lo cual impide el despliegue de toda la cartografía básica generada por el SIGA. 2. No hay acceso a otros bancos de datos intrainstitucionales, a 	

	pesar de estar operando en plataformas compatibles mediante llamadas SQL, con la facilidad de poder enviar los resultados de la consulta a usuarios remotos mediante acceso vía herramientas API	
--	--	--

1.3. Objetivos

En la CONAGUA, el uso de bases de datos geoespaciales y sistemas de información geográficos aumenta diariamente, ya que a través de ellos es posible tener un panorama más cercano a la realidad de la situación que presenta el agua en México; empero las herramientas de este tipo requieren un nivel de conocimientos profundos en ellas, así como de una adecuada distribución de la información a fin de poder garantizar su calidad, actualidad, veracidad y oportunidad para la adecuada implementación en la toma de decisiones.

Objetivo general.

Implementar, mediante un marco metodológico y de actuación, una estructura, en red, de un sistema de bases de datos distribuido con al menos 3 geobases de datos para los respectivos Organismos de Cuenca sobre el mismo número de servidores virtuales, a nivel central o local; vinculando, adicionalmente, la geobase de datos central con al menos 3 de los sistemas de información más importantes para la Subdirección General Técnica a fin de aprovechar toda aquella información dispersa para integrarla y proporcionar una mejor percepción, un mayor control así como una eficiencia operacional, mejorada, en la gestión de las fuentes de datos.

Como resultado final, se pretende mejorar la consulta geoespacial tomándola como eje central para la búsqueda de la mayor parte de la información Institucional, con el objetivo de facilitar la toma de decisiones.

Los objetivos específicos del presente trabajo son los siguientes:

1. Identificar adecuadamente la problemática de la gestión de la información tanto de formato alfanumérico como de referencia geoespacial al interior de la Subdirección General Técnica.

2. Evaluar los aspectos organizacionales de gestión de la información.
3. Diseñar, configurar e instalar una red de geoservidores virtualizados que permitan operar a la GeoAgua en un entorno distribuido.
4. Implantar, a través de un plan de trabajo preciso y realista, el mecanismo de vinculación entre los bancos de información oficiales de las diferentes gerencias que conforman la Subdirección General Técnica con la GeoAgua en un entorno interoperable.

Así, el vínculo entre las bases de datos geoespaciales facilitará el despliegue de toda la cartografía básica generada por el SIGA, además permitirá el acceso a otros bancos de datos intrainstitucionales, mediante llamadas SQL, con la facilidad de poder enviar los resultados de la consulta a usuarios remotos mediante acceso vía herramientas API (c.f. Glosario).

Con la implementación y desarrollo de un vínculo de bases de datos y sistemas de información entre diversas áreas de la COANGUA, será posible que una mayor cantidad de personas que laboran en esta Institución así como personal del sector medio ambiente, y en un futuro próximo el público en general, puedan contar con información automática, expedita, segura, confiable y de calidad de todo el acervo de datos con el que cuenta la CONAGUA que sea de utilidad para las actividades que realizan.

1.4. Hipótesis.

La implementación de una red de geoservidores virtualizados distribuidos, que posibilite el acceso automatizado entre los diversos bancos de datos institucionales de la COANGUA y estos geoservidores, a nivel nacional, permitirá un manejo más rápido, sencillo, ágil, dinámico y robusto para la obtención de información estratégica a fin de facilitar la toma de decisiones.

1.5. Marco teórico-conceptual y estado del conocimiento.

1.5.1. Marco teórico-conceptual.

La virtualización se está convirtiendo en una práctica estándar de TI. Su uso es cada vez es más popular en los SIG. Estas soluciones están siendo ampliamente utilizadas tanto en entornos de prueba y desarrollo como en el de producción, proporcionando grandes beneficios a las aplicaciones en ambientes SIG.

El concepto de virtualización es un término amplio que se refiere a la abstracción de los recursos informáticos llamada Hypervisor²¹ que crea una capa entre el hardware de la máquina física (“host”) y el sistema operativo de la máquina virtual (“guest”)²², siendo un medio para crear una versión virtual de un dispositivo o recurso, como un servidor, un dispositivo de almacenamiento, una red o incluso un sistema operativo, donde se divide el recurso en uno o más entornos de ejecución. El Hypervisor [ref. 3] maneja, gestiona y arbitra los cuatro recursos principales de una computadora: CPU, memoria, tarjeta de red y almacenamiento local; permitiendo repartir dinámicamente dichos recursos entre todas las máquinas virtuales definidas en el “guest”; de este modo, permite tener varias computadoras virtuales ejecutándose sobre la misma computadora física.

La virtualización se encarga de crear una interfaz externa que esconde una implementación subyacente mediante la combinación de recursos en localizaciones físicas diferentes o por medio de la simplificación del sistema de control. Un avanzado desarrollo de nuevas plataformas y tecnologías de virtualización han hecho que se vuelva a prestar atención a este importante concepto. De modo similar al uso de términos como “abstracción” y “orientación a objetos”, la virtualización es usado en muchos contextos diferentes.

La máquina virtual, en general, es un sistema operativo completo que corre como si estuviera instalado en una plataforma de hardware autónoma. Típicamente muchas máquinas virtuales son simuladas en una computadora física central. Para que el sistema operativo del “guest” funcione, el sistema físico debe ser lo suficientemente grande, siempre dependiendo del tipo de virtualización.

Existen tres formas de virtualización para servidores, de las cuales dos están presentes en el mercado, a saber: la Hypervisor y la OS virtualization.

1.5.2. Virtualización de hardware.

La **virtualización de hardware (completa) o Hypervisor**, virtualiza a nivel de hardware creando una réplica de todos los recursos del sistema tales como el sistema operativo, CPU, memoria y archivos de configuración. Esto genera el punto común de referencia del hardware. Este tipo de virtualización, se lleva a cabo en una plataforma de hardware mediante un software “host”, que es un programa de control que simula un entorno computacional (máquina virtual) para su software “guest”. Este software “guest”, que generalmente es un sistema operativo completo, se ejecuta como si estuviera instalado en una plataforma de hardware autónoma. Muchas máquinas virtuales pueden ser simuladas en una

²¹ También conocida como Virtual Machine Monitor(VMM)

²² Virtual machine

máquina física dada. Para que el sistema operativo “guest” funcione, el sistema físico debe ser lo suficientemente grande como para soportar todas las interfaces externas de los sistemas “guest”, las cuales deben incluir los drivers del hardware simulado. La plataforma Institucional, en CONAGUA, es la de Microsoft. Con base en esto enseguida se indican las recomendaciones mínimas, sobre esa plataforma, para la instalación de un servidor tipo Hypervisor, que contendrá máquinas virtuales²³:

- Procesador:
 - x64
 - Multi-socket, multi core
 - CPU de alta velocidad
- Virtualización:
 - Virtualización asistida por hardware. Está disponible para procesadores que incluyen una opción de virtualización.
 - AMD Virtualization (AMD-V)
 - Intel Virtualization Technology (Intel VT)
 - Traducción de direcciones de Segundo nivel (Second Level Address Translation SLAT)
 - AMD Nested Page Tables (NPT) o Rapid Virtualization Indexig (RVI)
 - Intel Extended Page Tables (EPT)
 - Hardware con Data Execution Prevention (DEP), Prevención de ejecución de datos, habilitada.
 - AMD (NX no execute bit)
 - Intel (XD execute disable)
 - Concretamente, debe habilitar el bit XD de Intel (bit ejecutar deshabilitado) o el bit NX de AMD (bit no ejecutar).
- Almacenamiento local
 - RAID 5, RAID 6 o RAID 10 para la partición del Sistema Operativo.
 - 120 GB como mínimo
- RAM
 - 64 GB o más
- Se recomienda el uso de servidores en cluster para soportar alta disponibilidad y tolerancia a fallas.

La virtualización de Hypervisor tiene una capa base, generalmente un kernel o núcleo²⁴ que se carga directamente en el servidor base, conforme se muestra en

²³ <http://technet.microsoft.com/es-es/library/cc731898.aspx>

²⁴ Es un software que constituye una parte fundamental del sistema operativo y se define como la parte que se ejecuta en modo privilegiado, conocido también como modo núcleo. Es el principal responsable de facilitar a los distintos programas acceso seguro al hardware de la computadora o en forma básica, es el

la Figura 2. Para asignar hardware y recursos a las máquinas virtuales, es recomendable que todo el hardware del servidor esté virtualizado. La siguiente capa superior muestra cada chip o placa que debe virtualizarse para que así pueda ser asignado a las máquinas virtuales; una vez en la máquina virtual, hay una copia completa de un sistema operativo y finalmente la aplicación según se indica en la Figura 3. A su vez, existen dos modelos básicos para este sistema de virtualización: la virtualización total y la paravirtualización²⁵. La primera ofrece una simulación completa del hardware fundamental, la segunda permite un modelo “similar” del hardware fundamental. La virtualización total está implementada como hypervisor tipo 1, la cual corre directamente en el hardware y la paravirtualización, que es un hypervisor tipo 2, la cual corre a nivel superior de un sistema operativo tradicional.

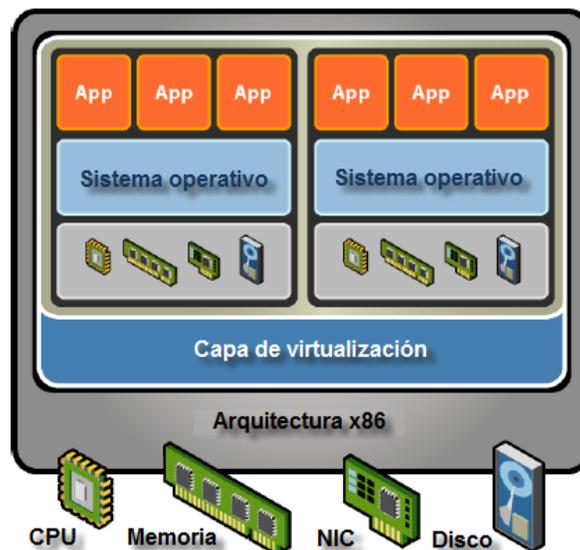


Figura 2. Esquema de virtualización tipo Hypervisor

Es de mencionar que múltiples copias de los recursos del sistema producen una sobrecarga en el servidor físico. Esta sobrecarga reduce las capacidades de los servidores virtuales por servidor físico así como el desempeño de sí mismo.

encargado de gestionar recursos, a través de servicios de llamada al sistema. Linux y Hyper-V son ejemplos de sistemas operativos con kernel.

²⁵http://www.vmware.com/files/pdf/VMware_paravirtualization.pdf, “Understanding full virtualization, paravirtualization and hardware assist”, White paper.

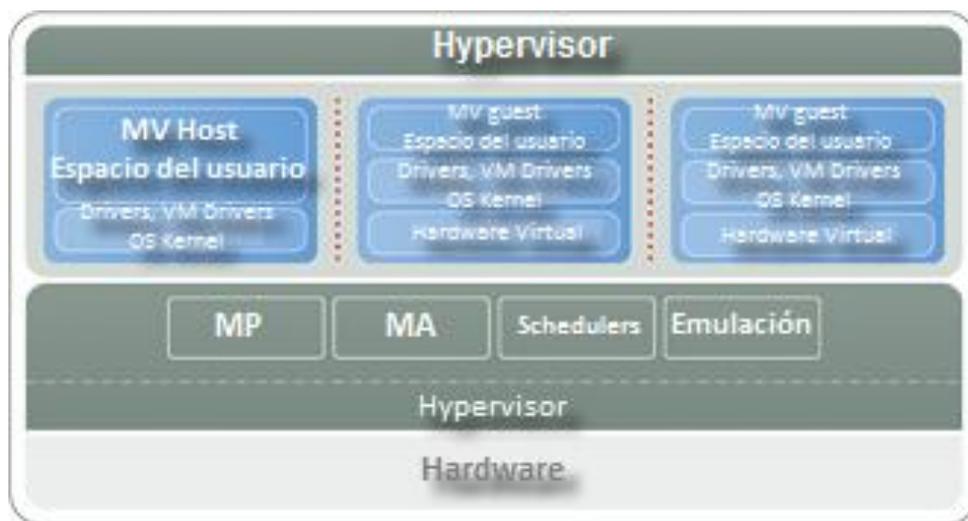


Figura 3. Diagrama de operación de máquinas virtuales en un Hypervisor. MV= memoria virtual, MP=memoria principal, MA=Memoria Aleatoria, MV Host= Máquina virtual anfitriona, MV guest=Máquina virtual invitada, OS=Sistema operativo

1.5.3. Virtualización en el Sistema Operativo.

El segundo tipo es la **virtualización en el sistema operativo del host (OS Virtualization)**, la cual provee una capa de virtualización común que distribuye los recursos del sistema físico entre todos los servidores virtuales, llamados **contenedores**. El resultado es una capa de virtualización más eficiente. Esto se traduce en mejores desempeños por parte de los servidores virtuales operando sobre los servidores físicos, acercando el desempeño virtualizado al de los servidores nativos; ventajas únicas en el manejo de los ambientes virtualizados como se muestra en la Figura 4. Algunas de estas ventajas incluyen la relocalización dinámica de los recursos y la habilidad para crear ligas al sistema operativo del servidor creando operaciones de manejo más rápidas y ligeras. Este tipo de virtualización consiste en instalar un sistema operativo dentro de otro al que se le llama “guest”, mediante el uso de una máquina virtual. Este tipo de proceso permite la virtualización de servidores en la capa del sistema operativo (kernel). Este método de virtualización crea particiones aisladas o entornos virtuales (Ves) en un único servidor físico e instancia de SO para así maximizar los esfuerzos de administración del hardware, software y centro de datos. El uso de virtualización de nivel-SO o partición²⁶ en las arquitecturas de nube pueden ayudar a resolver algunos de los temas de seguridad central,

²⁶ Tal y como LPARs, VPARS, NPARS y dominios del Sistema Dinámico

privacidad y regulación que pudieran de otra manera dificultar la adopción del cómputo de nube.

1.5.4. Virtualización parcial (Address Space Virtualization).

El tercer tipo de virtualización es la llamada **virtualización parcial (Address Space Virtualization)**. La máquina virtual simula múltiples instancias de gran parte, pero no de todo, del entorno subyacente del hardware, particularmente del address spaces. Tal entorno acepta compartir recursos y alojar procesos, pero no permite instancias separadas de sistemas operativos “guest”. Aunque no es vista como dentro de la categoría de máquina virtual, históricamente²⁷ éste fue un importante acercamiento, y se usó en sistemas como TSS, el IBM M44/44X experimental y podría mencionarse en sistemas operativos como OS/VS1, OS/VS2 y MVS.

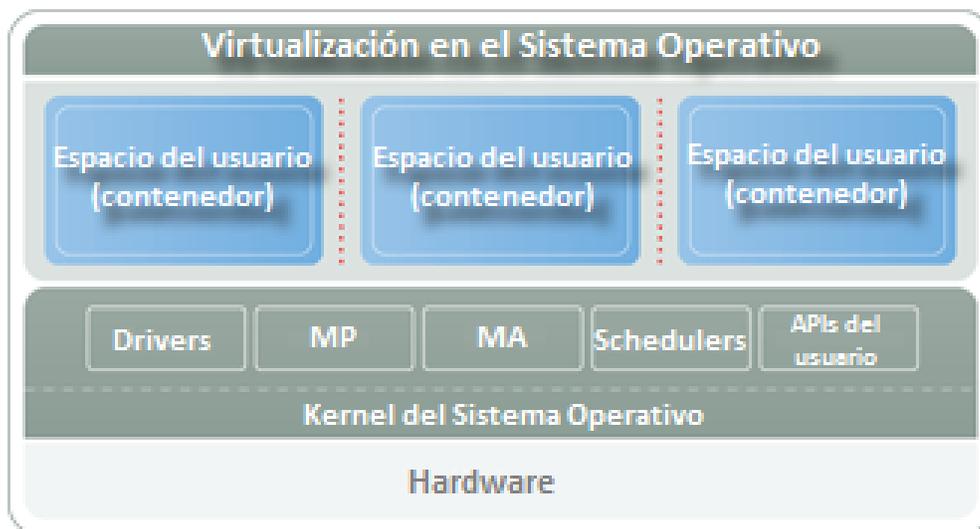


Figura 4. Diagrama de operación de máquinas virtuales en el Sistema Operativo

1.5.5. Virtualización de Red

Las técnicas de balanceo de carga han sido un tema muy importante en el cómputo de nube porque, conforme se escalan los sistemas físicos y virtuales dentro de la nube, lo mismo hace la complejidad del manejo de la carga de trabajo que es ejecutada para entregar el servicio.

Los balanceadores de carga agrupan múltiples servidores y servicios detrás de las direcciones IP virtuales. Ellos proveen una guía basada en recursos de

²⁷ <http://www.kernelthread.com/publications/virtualization/>

requerimientos de servicio y recuperación de falla automática cuando el nodo falla. Mientras los balanceadores del hardware puedan ejecutar balanceadores basados en el software, su flexibilidad es siempre limitada.

Un reto significativo en la red de cómputo de nube no es sólo la provisión de interfaces individuales de la red virtual a un ambiente virtual dado, sino también la necesidad creciente de las infraestructuras de nube de ofrecer más centros de datos virtuales más complejos, los cuales provean un conjunto de diferentes roles de sistemas y la interconexión lógica entre esos roles.

Los componentes de una red virtual incluyen switches virtuales, adaptadores virtuales de red, un servidor DHCP virtual y un dispositivo para NAT.

Switches Virtuales.

Como un switch físico, un switch virtual conecta varios componentes de la red. Los switches virtuales, son llamados Vmnet0, Vmnet1, Vmnet2, etc., en los sistemas virtuales de VMWare²⁸. Unos pocos switches virtuales son mapeados, por defecto, a redes específicas. En VMWare es posible crear redes personalizadas solamente en las máquinas virtuales que fueron creadas en la Workstation.

Existen tres formas de configurar las redes virtuales en VMWare, como se indica en la Tabla 2.

Tabla 2. Tipos de switches en redes virtuales

Tipo de red	Nombre de switch
Bridged	Vmnet0
NAT	Vmnet8
Host-only	Vmnet1

Puente (Bridged). Esta configuración conecta o “*puentea*” una máquina virtual a una red física mediante el uso de la NIC²⁹ física del sistema host. Si el sistema host está conectado a una red, la opción bridged networking es, por lo general,

²⁸ VMware Inc., (VM de *Virtual Machine*) es una filial de EMC Corporation que proporciona software de virtualización disponible para computadoras y servidores. Entre sus principales productos se incluyen VMware Workstation y los gratuitos VMware Server y VMware Player. El software de VMware puede funcionar en Windows, Linux y en la plataforma Mac OS X que corre en procesadores INTEL, bajo el nombre de VMware Fusion.

²⁹ Acrónimo, en inglés, utilizado para hacer referencia a una tarjeta de red (Network Interface Card)

la forma más fácil de darle acceso a la máquina virtual a la red física conforme se muestra en la Figura 5.

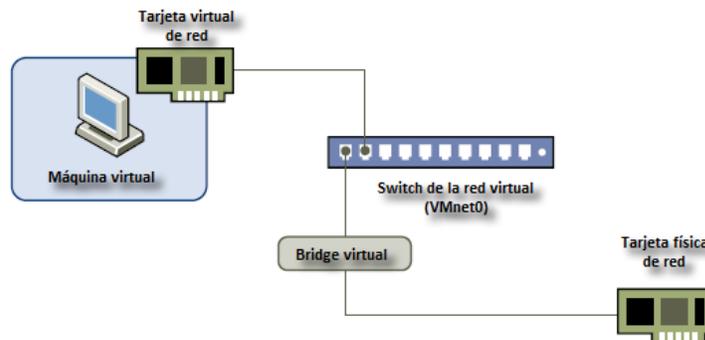


Figura 5. Diagrama de configuración de red virtual tipo bridge

NAT. Con la opción NAT, una máquina virtual no tiene su propia dirección IP en la red externa. En su lugar, se crea una red privada virtual separada dentro del sistema host. En la configuración que se indica en la Figura 6, por default, una máquina virtual obtiene una dirección IP, dentro de la red virtual, a partir del servidor DHCP virtual; así, la máquina virtual y el sistema host comparten una sola identidad en la red física que no es visible en esta red externa. Cuando se usa el wizard para generar una nueva máquina virtual y se selecciona la configuración típica, el wizard configura la máquina virtual para usar la red NAT por default. Solamente es factible tener una red NAT.

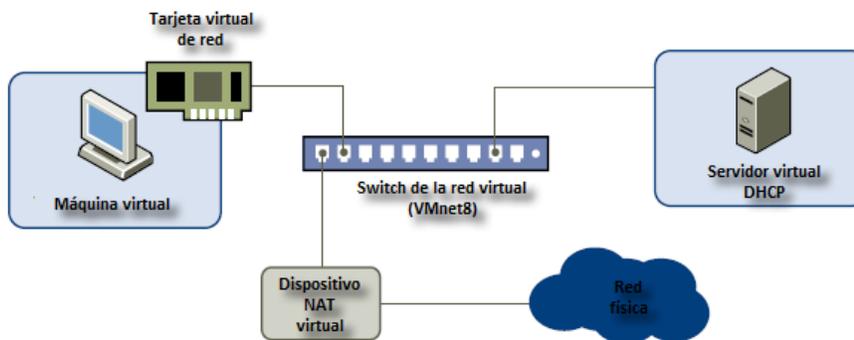


Figura 6. Diagrama de configuración de red virtual tipo NAT

Host-Only. Esta alternativa, mostrada en la Figura 7, crea una red que está completamente contenida dentro de la computadora host; provee una conexión de red entre la máquina virtual y el sistema host mediante el uso de un adaptador virtual de red que solamente es visible en el sistema operativo del equipo host.

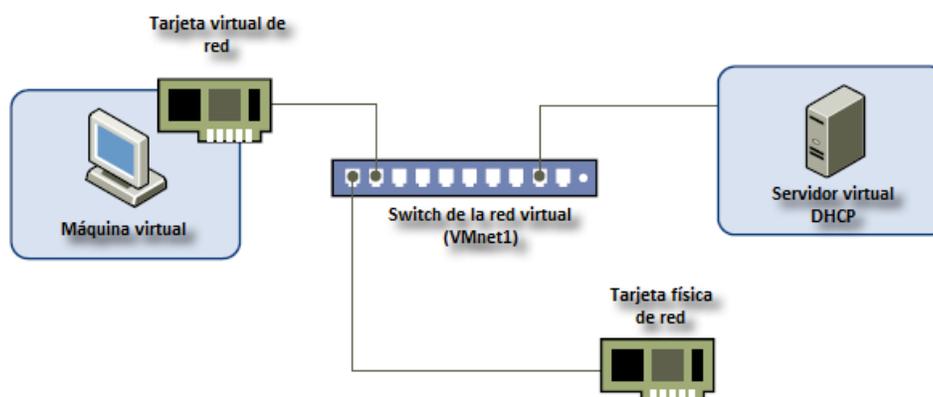


Figura 7. Diagrama de configuración de red virtual tipo host only

En la Figura 9 se puede observar que el switch generado por default está vinculado a la tarjeta física del equipo. Por su parte Hyper-V permite crear switches virtuales adicionales al que se genera en forma automática. Existen, en VMWare, tres tipos o clases diferentes de redes virtuales³⁰ que se pueden crear, según se indica en la Figura 8, a saber:

Externa (External). Esta opción crea un switch virtual a través del cual las máquinas virtuales pueden acceder a la red física completa y a Internet, suponiendo que se cuenta con la infraestructura necesaria en sitio. Adicionalmente, cada adaptador físico de red puede ser usado solamente para una red virtual. Por lo tanto si se está creando una red virtual secundaria externa; entonces será necesario contar con una NIC, ver nota anterior al respecto, secundaria la cual se puede vincular a la nueva red generada.

Interna (Internal). En este caso se crea un switch virtual interno; este dispositivo no es capaz de acceder la red privada como un todo. Sirve, en primera instancia, como un mecanismo para permitir la comunicación entre las máquinas virtuales que se encuentran como huéspedes (*hostead*³¹) en el server. Una conexión de

³⁰ Se debe tener cuidado en no confundir la forma de configurar las redes virtuales, la cual hace referencia a los elementos que conforman la arquitectura con el tipo de la red virtual.

³¹ El término host es usado en cómputo para hacer referencia a las computadoras conectadas a una red, que proveen y utilizan servicios o de esta. Los usuarios deben utilizar equipos que trabajen como anfitriones para tener acceso a la red. En general, los anfitriones son computadores monousuario o multiusuario que ofrecen servicios de transferencia de archivos, conexión remota, servidores de base de datos, servidores web o cualquier otro servicio. Los usuarios que hacen uso de los anfitriones pueden a su vez pedir los mismos servicios a otras máquinas conectadas a la red. De forma general un anfitrión es todo equipo informático que posee una dirección IP, interconectado con uno o más equipos, que brinde servicios a un área o persona que no sea propietaria del equipo en sí mismo. Un host o anfitrión es una computadora que funciona como el punto de inicio y final de las transferencias de datos.

este tipo facilita la comunicación entre el sistema operativo del *host* y el del *guest*.

Privada (Private). Una red privada virtual debe ser usada únicamente para facilitar la comunicación entre las máquinas virtuales que están siendo *hosted* en el servidor. Este tipo de redes no puede tener acceso a la red externa ni al sistema operativo del sistema *host*.

Adaptadores virtuales de red

Cuando se crea una máquina virtual nueva el *wizard*, en este caso de VMWare, crea un adaptador virtual de red nuevo para esa máquina virtual en particular. El adaptador de red virtual aparece en el sistema operativo invitado como una tarjeta AMD PCNET PCI o una Intel Pro/1000 MT Server. En los sistemas operativos *guest* Windows Vista y Windows 7, aparece como una tarjeta Intel Pro/1000 MT Server.

Desde la versión de VMWare Player 3.x y posteriores cada máquina virtual puede tener hasta 10 adaptadores de red virtuales.

Servidor Virtual DHCP

El servidor virtual Dynamic Host Configuration Protocol (DHCP) provee direcciones IP a las máquinas virtuales en configuraciones que no están vinculadas a una red externa. Este servidor solamente asigna direcciones IP y configuraciones NAT a máquinas virtuales que se encuentren dentro del equipo *host*.

NAT Device

En una configuración tipo NAT, el dispositivo NAT pasa datos entre una o más máquinas virtuales y la red física externa, identificando paquetes de datos entrantes para cada una de las máquinas virtuales y enviándolos al destino correcto. Es posible configurar una conexión *bridged networking*, NAT y *host-only networking* para máquinas virtuales. Adicionalmente es posible usar los componentes de redes virtuales para crear sofisticadas redes virtuales personalizadas. Cabe señalar que la creación de redes personalizadas solamente está disponible en máquinas virtuales creadas en la versión Workstation de VMWare.

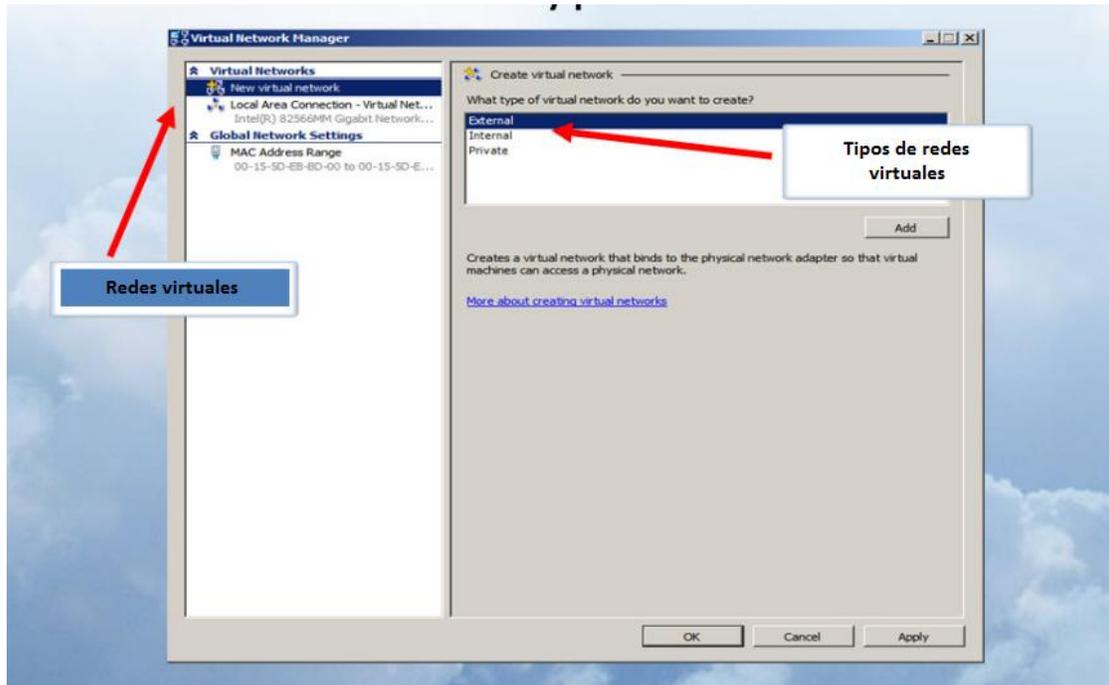


Figura 8. Tipos de redes virtuales en plataforma Hyper-V

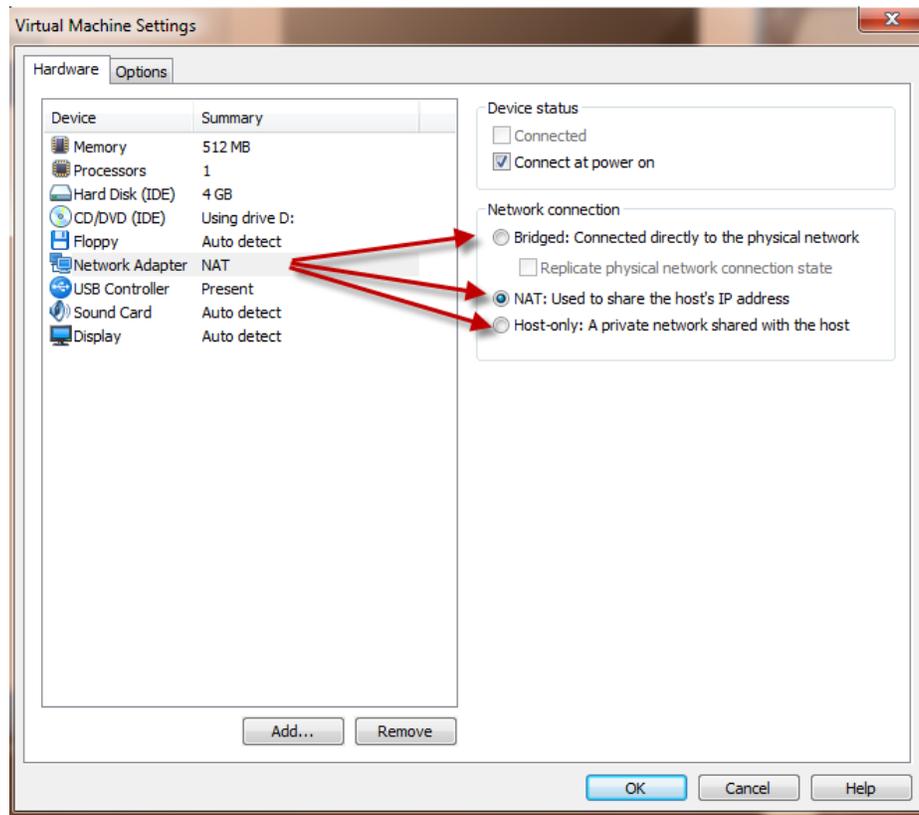


Figura 9. Tipos de redes virtuales en plataforma VMWare

1.5.6. Virtualización de Aplicaciones

El contenedor web es la parte del servidor de aplicación que maneja *servlets*³², archivos JavaServer Page³³ y otros componentes Web-tier. Pero no todas las tecnologías del contenedor Web son creadas de la misma forma. Apache Tomcat³⁴, por ejemplo, es una tecnología de contenedor Web de código abierto, pero tiene limitaciones para los desarrolladores quienes quieren ir más allá de las aplicaciones *Web-tier*³⁵. Si una aplicación necesita hacer uso de persistencia, *clustering*, mensajería o Enterprise Java Beans³⁶ estas capacidades tienen que

³² Clase en el lenguaje de programación Java, utilizada para ampliar las capacidades de un servidor. Aunque los *servlets* pueden responder a cualquier tipo de solicitudes, éstos son utilizados comúnmente para extender las aplicaciones alojadas por servidores web, de tal manera que pueden ser vistos como applets de Java que se ejecutan en servidores en vez de navegadores web. Este tipo de programación son la contraparte Java de otras tecnologías de contenido dinámico Web, como PHP y ASP.NET. La palabra *servlet* deriva de otra anterior, *applet*, que se refiere a pequeños programas que se ejecutan en el contexto de un navegador web. El uso más común de los *servlets* es generar páginas web de forma dinámica a partir de los parámetros de la petición que envíe el navegador web.

³³ JSP por sus siglas en inglés.

³⁴ <http://tomcat.apache.org/>

³⁵ <http://javablog.eliumontoya.com/home/webtierexplicacion>

³⁶ EJB por sus siglas en inglés.

ser añadidas a Tomcat, una por una, mientras que GlassFish Project provee una colección integrada de contenedores Java EE que entrega todas esas capacidades.

Existe un problema angular de software para los “contenedores” dentro de la nube. La tecnología de contenedor web implementada en la nube impacta, de manera significativa, a la productividad y flexibilidad del desarrollador.

Hoy en día, la mayoría de las ofertas del cómputo de nube se concentran en la virtualización de plataformas completas en donde el desarrollador escoge el SO y la plataforma de desarrollo. Pero las crecientes nubes públicas, así como las nubes privadas ofrecerán abstracciones de una programación de ambiente de desarrollo al más alto nivel. Con el tiempo, podríamos esperar que el nivel de abstracción con el que el desarrollador tiene interface se mueva gradualmente hacia arriba mientras que una mayor funcionalidad se filtra hacia adentro de la plataforma.

1.5.7. Virtualización de Bases de Datos

En el contexto de geoservidores, la virtualización se utiliza generalmente para la virtualización de plataforma, que permite que múltiples sistemas operativos y/o las aplicaciones se ejecuten en una máquina física; esto se hace a través del concepto de máquinas virtuales.

Los beneficios de la virtualización son muy variados. Estos les permiten a las organizaciones reducir:

1. El tiempo de configuración³⁷ y puesta a punto de servidores nuevos.
2. La necesidad de nuevos servidores así como el número de los servidores existentes mejorando el uso de estos últimos, es decir, permite la implementación de una estrategia de consolidación de servidores.
3. Los costos de propiedad tales como consumo de energía eléctrica, enfriamiento, espacio y mantenimiento; por mencionar algunos.
4. Prever el impacto de una aplicación sobre otra cuando se efectúa alguna actualización o cambio de plataforma, por ejemplo, correr diferentes versiones de ArcGIS sobre el mismo servidor físico.
5. Incrementar la continuidad de la operación de los servicios a través de la reducción del tiempo de inactividad mediante la recuperación rápida de interrupciones no planificadas con la capacidad de copia de seguridad y migración de todos los entornos virtuales sin interrupción.

³⁷ Configuración significa la instalación y adecuación del sistema operativo y determinadas aplicaciones, basadas en requisitos organizacionales, para el óptimo desempeño del servidor al tiempo de ejecución.

En la Tabla 3 se presenta una comparación de algunos de los productos para la implementación de tecnologías de virtualización disponibles en el mercado, de los cuales los más importantes, por sus ventas y buen desempeño, son:

Vmware

Es, actualmente, una de las plataformas de virtualización más populares entre los desarrolladores y usuarios de SIG. La utilizan, esencialmente, para virtualizar entornos de ArcGIS Server. Sin embargo, también la utilizan para virtualizar aplicaciones ArcGIS Desktop³⁸.

Microsoft® Hyper-V

Es la solución de virtualización incluidas en Microsoft Windows Server 2008. Está ha presentado una aceptación que ha crecido muy rápidamente entre los usuarios de ArcGIS³⁹.

Sun Microsystems Solaris Containers

Es la solución de virtualización nativa de Solaris 10. Los usuarios de ArcGIS lo utilizan, principalmente, para virtualizar servidores de base de datos⁴⁰.

Citrix XenServer y Citrix Essentials

Citrix ofrece XenServer de forma gratuita e incluye la más avanzada solución de gestión de virtualización de Citrix Essentials, que ofrece una funcionalidad adicional, incluyendo la automatización y la integración con Microsoft Hyper-V⁴¹.

³⁸ Para mayores detalles, consultar <http://www.vmware.com>

³⁹ Para obtener más información, consultar <http://www.microsoft.com/windowsserver2008>

⁴⁰ Para mayores detalles, consultar: <http://www.sun.com/solaris/containers>

⁴¹ Para obtener más información, consultar <http://www.citrix.com/xenserver>

Tabla 3. Comparación entre las diferentes plataformas de virtualización disponibles en el mercado

Plataforma en hardware físico	Plataforma en software físico	Software
x86 y AMD64/Intel64	<ul style="list-style-type: none"> • Windows 7 (32/64-bit) • Windows 2000 • Windows XP (32/64-bit) • Windows Server 2008 (32/64-bit y R2) • Debian 5.0 (64-bit) • Debian 5.0 (64-bit) • SUSE 9/10.0 • Mandriva 2008 • Mandriva 2009.0/2009.1 • Mandrake 10.1 • Fedora 7, 8, 9, 10, 11 • Red Hat Enterprise Linux 5 • Red Hat Linux 9 • Xandros 4 • Mac OS X Server (Leopard, Snow Leopard) Macintosh • OpenSolaris 	 <p>VirtualBox (Hypervisor Virtualization)</p>
Cualquier hardware estándar tipo x86, incluyendo sistemas de virtualización asistida Intel y AMD. También soporta sistemas de doble procesador Virtual SMP, permitiendo que una sola máquina virtual comparta múltiples procesadores físicos.	Incluye soporte para Windows Server 2008, Windows Vista Business Edition y Ultimate Edition, Red Hat Enterprise Linux 5 y Ubuntu 8.04.	 <p>Vmware Server (Operating System Virtualization) Vmware vSphere</p>
Ambientes Linux: X86, IA64, AMD64, EM64T Ambientes Windows: Cualquier plataforma x86 o x64 soportada por Windows Server 2008 y Windows Server 2003 (Intel, AMD o similar) así como IA64 para Windows Server 2003	<p>Linux: CentOS 4, 5; Fedora 7, 8; Red Hat Enterprise 4, 5; SuSE Enterprise Server 10</p> <p>Windows: Windows Server 2008 (32 & 64-bit) Windows Server 2003 (32 & 64-bit)</p>	 <p>Parallels Virtuozzo Containers (Operating System Virtualization)</p>
4-bit x86 server-class system CPU: 1.5 GHz mínimo, 2 GHz o superior se	X86-based system Microsoft® Windows® 2000, Windows XP, Windows Server® 2003, Windows Server 2008,	 <p>XenServer</p>

<p>recomienda Intel® VT multi-core o AMD-V™</p> <p>De 1GB a 256GB en memoria física</p> <p>Más de 64 procesadores lógicos</p> <p>Disco de boot con 16 GB de espacio mínimo, 60 GB o más recomendado.</p>	<p>Windows Vista o Windows 7</p> <p>.NET Framework 2.0 SP1 o posterior</p>	<p>(Advance, Enterprise y Platinum)</p> <p>(Hypervisor Virtualization)</p>
<p>Microprocesador x64 compatible con tecnología Intel VT o AMD-V. Velocidad mínima de CPU: 1.4 GHz; Recomendada: 2 GHz o superior.</p> <p>RAM: Mínima: 1 GB</p> <p>Recomendada: 2 GB o superior (se requiere memoria adicional para cada sistema operativo invitado (guest)); Máximo: 1 TB.</p> <p>Espacio disponible en disco duro: Mínimo: 8 GB; Recomendado: 20 GB o superior (se requiere espacio adicional para cada sistema operativo invitado (guest)).</p>	<p>Windows Server 2008 R2</p>	<p>Microsoft®</p> <p>Hyper-V™ Server 2008 R2</p> <p>Microsoft Hyper-V</p> <p>(Hypervisor Virtualization)</p>

1.5.8. Almacenamiento de grandes volúmenes de datos (Big Data)

Con relación al almacenamiento de grandes volúmenes de datos, NetApp dirigió a sus clientes la propuesta de incrementar el ROI⁴² (*Return On Investment*) del centro de datos a fin de impulsar mayores eficiencias para aplicaciones de base de datos con la promesa de cero inversiones.

La herramienta FlexClone puede crear clones de bases de datos en minutos usando un espacio de almacenamiento mínimo, lo que resulta en un desarrollo y despliegue acelerados.

⁴² Una medida de desempeño usada para evaluar la eficiencia de una inversión o para comparar la eficiencia de un número de diferentes inversiones. Para calcular el ROI, el beneficio (retorno) de una inversión está dividido por el costo de la inversión; el resultado está expresada como un porcentaje o un cociente.

El aprovisionamiento ligero⁴³ mejora el uso de almacenamiento al acumular capacidad sin usar amplios volúmenes de almacenamiento y al compartirlos dinámicamente en todas las aplicaciones conforme los requerimientos cambian.

La presión por la reducción de costos ha llevado a la rápida adopción de tecnologías de almacenamiento eficiente en espacio como el Snapshot⁴⁴, la creación virtual de clones, el aprovisionamiento ligero, la deduplicación y todos los estándares en los sistemas NetApp. Algunas de las cuales se describen brevemente más adelante.

Con las tecnologías de eficiencia de almacenamiento de NetApp, se reducen los requerimientos de almacenamiento en el ambiente de desarrollo de Microsoft SQL Server. Las bases de datos de producción tienen una capacidad de al menos 10 TB y requerimientos para crear copias de bases de datos que tienen al menos 40 TB en tamaño. El ahorro potencial proviene de las eficientes características en el almacenamiento, incluidas en la deduplicación, FlexVol, copias Snapshot, replicación SnapMirror, protección de datos SnapVault y copias virtuales FlexClone de NetApp.

Así NetApp provee mayor eficiencia en sistemas de almacenamiento SAN, mediante la recuperación de capacidades a fin de que sean usadas ineficazmente y mediante la reducción de consumo de espacio de almacenamiento.

Estos programas trabajan sobre ambientes de prueba y desarrollo de bases de datos de Microsoft SQL Server. Esto permite acelerar el proceso de patchit⁴⁵,

⁴³ El aprovisionamiento ligero (thin provisioning TP) es un método de optimización de la eficiencia con la que se utiliza el espacio disponible en las redes de área de almacenamiento (SAN). TP opera mediante la asignación de espacio de almacenamiento de disco de una manera flexible entre múltiples usuarios, basado en el espacio mínimo requerido por cada usuario en cualquier momento dado. En el modelo de aprovisionamiento de almacenamiento convencional, también conocido como aprovisionamiento grueso (fat provisioning FP), el espacio de almacenamiento se asigna más allá de las necesidades actuales, en previsión de la necesidad cada vez mayor y el aumento de la complejidad de los datos. Como resultado, la tasa de utilización es baja. Las grandes cantidades de espacio de almacenamiento se pagan, pero nunca pueden ser utilizados. En TP, estos problemas se eliminan mientras se mantiene bajo costo operativo. Los beneficios adicionales incluyen la reducción del consumo de energía eléctrica, las necesidades de espacio de hardware más pequeños y menor generación de calor en comparación con los sistemas de almacenamiento en red tradicionales. (ref. <http://searchstorage.techtarget.com/definition/thin-provisioning>)

⁴⁴ Es una copia instantánea de volumen del estado de un sistema en un momento determinado. El término fue acuñado como una analogía a la de la fotografía. Puede referirse a una copia real del estado de un sistema o de una capacidad que ofrecen los sistemas de copia de seguridad.

⁴⁵ Un parche es un software diseñado para solucionar problemas con un programa o una actualización del mismo en una computadora o en sus datos de apoyo; esto incluye la reparación de la vulnerabilidad de seguridad y otros errores de programación también conocidos como "bugs", la mejora de la usabilidad y el rendimiento. Aunque la intención es solucionar los problemas, parches mal diseñados a veces pueden introducir nuevos problemas.

actualización y desarrollo mediante el proceso de creación de copias de bases de datos en un espacio eficiente sin la pérdida de sobrecarga requerida por los sistemas tradicionales. Así, se incrementa la velocidad de desempeño y ciclos de prueba.

1.5.9. Geobases de datos distribuidas.

Los modelos de Sistemas de Información Geográfica (SIG) han evolucionado gradualmente en estructuras georrelacionales donde sus atributos relacionales pueden ser almacenados en una base de datos relacional (RDBMS)⁴⁶ la cual ha sido vinculada a características basadas en archivos geoespaciales. El formato georrelacional tiene una muy buena escalabilidad y una estructura de datos dual⁴⁷ lo cual implica que la Geobase de Datos puede tomar completa ventaja de las características de las RDBMS tales como respaldo, recuperación, replicación y tolerancia a fallas. Adicionalmente, permite el soporte a grandes capas de información lo cual requiere del uso de estructuras complejas a fin de mantener el desempeño y compartir la información geoespacial con otras aplicaciones.

1.5.9.1. Bases de datos habilitadas geoespacialmente.

A mediados de los años noventa, la nueva tecnología emergente permitió almacenar datos espaciales en los RDBMS⁴⁸, abriendo una nueva era de gran escalabilidad y soporte para grandes capas de los datos geográficos continuos (Arctur, 2004). Cuando las nuevas bases de datos espacialmente habilitadas se combinaron con ambientes de desarrollo de cliente-servidor que podrían asociarse con aplicaciones comerciales, la vinculación de datos espaciales con aplicaciones esenciales de la Institución tales como sistemas ejecutivos, fue posible. Además, estas bases de datos espacialmente habilitadas permitieron dar los primeros pasos hacia SIG empresariales y la eliminación de “islas de los datos espaciales”.

Quizás en forma no coincidente, el movimiento de los SIG abiertos fue gestando poco después la llegada de los primeros modelos totalmente-relacionales capaces de almacenar tanto datos espaciales como de tablas de atributos en una sola base de datos relacional cuando las organizaciones de estándares tales como el *Open GIS Consortium* (OGC), la Organización Internacional para Estandarización y el Comité Federal Americano para Datos Geográficos (FGDC), comenzaron a promover la idea de que los datos espaciales se deben compartir a través de normas (Camps, 2005). El trabajo inicial de estas organizaciones se

⁴⁶ Un sistema de bases de datos relacionales (relational database management system RDBMS) es un sistema de gestión de bases de datos que se basa en el modelo relacional. Una breve definición de un RDBMS es: un sistema de gestión de bases de datos en la que éstos se almacenan en tablas y las relaciones entre ellos también se almacenan en tablas. Los datos pueden ser accedidos o vuelto a montar de muchas maneras diferentes sin tener que cambiar las formas de las tablas.

⁴⁷ Características espaciales almacenadas en un formato basado en archivos con atributos almacenados en una base de datos relacional.

⁴⁸ También referido como habilitación espacial de base de datos relacionales.

centró en compartir los atributos espaciales simples en una base de datos relacional, mientras habilitaban la interoperabilidad entre los programas de diferentes vendedores de SIG. OGC, un consorcio de la industria internacional de compañías privadas, agencias gubernamentales y universidades, publicó una norma espacial abierta que llamó la *OpenGIS Simple Features Specification* [ref. 14].

ESRI participó activamente en la definición de las *OpenGIS Simple Features Specification* y fue la primera compañía con productos que cumplían con esa norma (ESRI, 2003). ESRI cuenta con productos cliente y servidor que cumplen con las especificaciones *OpenGIS Simple Features* para SQL. ArcSDE⁴⁹, utiliza el esquema binario original para Oracle y SQL Server, el cual es totalmente compatible con las *OpenGIS Simple Feature Specification* para la geometría binaria de SQL. ArcSDE también provee soporte para tipos adicionales de datos SIG tales como valores z, medidas, anotaciones, soporte para imágenes raster y datos de estudio que cumplen ampliamente las especificaciones de la OGC.

Como solución al problema de interoperabilidad, ESRI cuenta con los servicios Web⁵⁰ (Amutio, 2007 y ESRI, 2003). Los servicios Web evitan los problemas y complicaciones inherentes a aplicaciones de SIG que se atan al esquema espacial de un vendedor particular de RDBMS y les permite a otros SIG manejar sus propios datos los cuales hacen uso de métodos y formatos adecuados a sus herramientas en un ambiente de base de datos cualquiera. Además, los servicios de Web permiten acceso a datos, servicios y consulta en un esquema servidor-a-servidor, opuesto al caso de una integración que sólo pasa al nivel del cliente. ESRI, utiliza un RDBMS con el esquema y métodos que óptimos para sus herramientas (ESRI, 2005). Otros usan sistemas de archivo⁵¹. Los servicios Web permite el manejo de datos propietarios de un SIG puede proporcionar rápidamente servicios consulta SIG⁵² a un público más grande en un ambiente común. Los servicios de Web constituyen, en sí mismos, toda una nueva estructura de normas computacionales. Los servicios de Web conforman todo una red de nodos distribuidos los cuales pueden incluir servidores, estaciones de trabajo, clientes del escritorio y “clientes ligeros”⁵³. Los servicios Web implican

⁴⁹ Solución de la compañía ESRI para el manejo y administración de información vectorial y raster en un RDBMS

⁵⁰ Web services

⁵¹ Estos sistemas de archivos se refieren al uso de archivos tipo “shape” el cual, a pesar de ser un estándar, no permite la facilidad de acceso hacia un banco de datos geográficos de gran tamaño, tampoco permite trabajar con “versiones” de la misma información y no propicia el trabajo participativo entre otros inconvenientes.

⁵² Datos, mapas y geoprocésamiento.

⁵³ Teléfonos celulares, PDAs, GPS, Palms, etc.

normas que proporcionan los elementos para que estos dispositivos actúen recíprocamente para formar una red informática de mayor entorno. También es importante reconocer que estos servicios Web representan una arquitectura poderosa para todos los tipos de informática distribuida.

Los servicios Web proveen la base para la difusión de la información vía redes abiertas⁵⁴. En los servicios Web, cada nodo puede tener cualquiera de los tres papeles siguientes: cliente⁵⁵, servicio⁵⁶ y “broker”⁵⁷. Un proceso del cliente hace una solicitud de un servicio a la red informática y recibe los resultados para cada una de éstas. Cualquier cliente de la red puede buscar un servicio requerido en el “broker”. Por requerimientos de la CONAGUA, para este trabajo se usaron servidores y TI en plataforma Windows publicando información mediante el MetadataExplorer⁵⁸.

Los servicios Web pueden apoyar la integración de información y servicios que se encuentran disponibles en una red distribuida. Esto ocurre en las grandes organizaciones, como CONAGUA la cual tiene varias entidades que administran datos geoespaciales en forma independiente⁵⁹. Al mismo tiempo, muchas de las funciones de CONAGUA exigen integrar estos conjuntos de datos. El uso de servicios de Web⁶⁰ como una herramienta acoplada al SIG⁶¹ puede atender esta necesidad eficazmente. El resultado es que se pueden consultar varias capas de información dinámicamente en forma integral, mientras al mismo tiempo los propietarios de los datos pueden mantener esta información en un ambiente cómputo distribuido. En la actualidad la CONAGUA hace uso de servicios de publicación WMS⁶² para la publicación de información cartográfica interactiva en acuerdo con varias dependencias gubernamentales, conforme se indica en la Figura 10. Entre los servicios que destacan se encuentran: “Espacio Digital Geográfico, Visor de mapas en línea, SEMARNAT”⁶³, “Atlas Nacional Digital

⁵⁴ Internet, redes inalámbricas y locales.

⁵⁵ Un cliente es cualquier computadora que tiene acceso a los recursos y funciones de uno o más nodos de la red informática. Los clientes típicos incluyen computadoras de escritorio, exploradores Web, subprogramas de Java y dispositivos móviles.

⁵⁶ Un servicio es un proceso de la red informática que espera las solicitudes, responde a cada una de ellas y devuelve un conjunto de resultados.

⁵⁷ Un bróker es esencialmente un portal de servicios de metadatos.

⁵⁸ <http://siga.cna.gob.mx/metadataexplorer/>

⁵⁹ Por ejemplo caminos, estudios, archivos de uso de suelo, límites administrativos, etc.

⁶⁰ Una herramienta que une.

⁶¹ Una tecnología integradora.

⁶² <http://www.opengeospatial.org/standards/wms>

⁶³ <http://infoteca.semarnat.gob.mx/website/geointegrador/mviewer/viewer.htm>

Interactivo de México (ANIM). INEGI⁶⁴, “Mapas de México. SEGOB⁶⁵”, este último se muestra en la Figura 11.

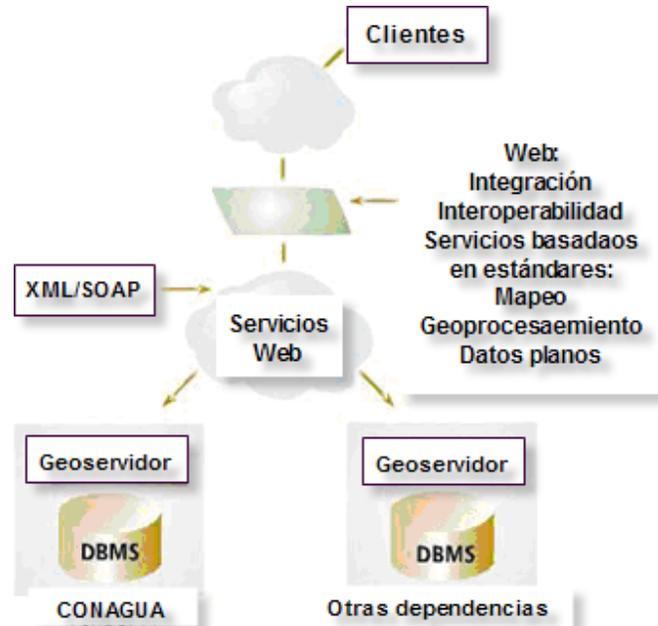


Figura 10. Integración e interoperabilidad basados en servicios Web.

⁶⁴ <http://www.atlasdemexico.gob.mx/>

⁶⁵ <http://www.mapas.gob.mx/home.do?newsession=ok>

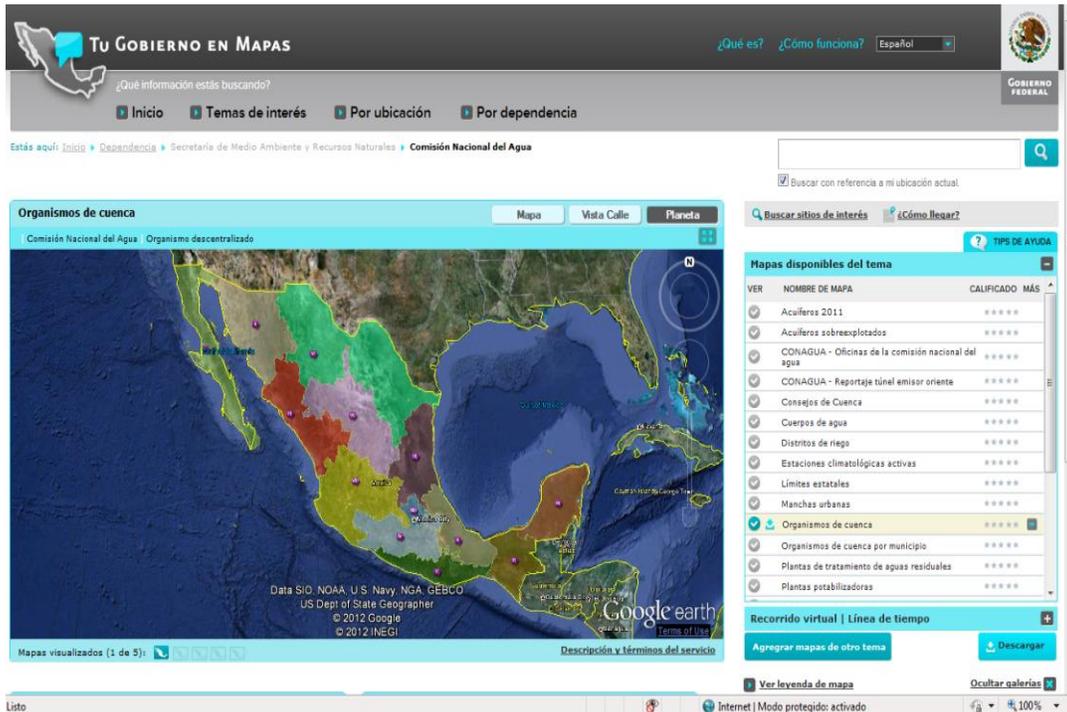


Figura 11. Página de publicación de servicios WMS.

Los estándares básicos utilizados para servicios Web son una serie de protocolos, como XML, Simple Object Access Protocol (SOAP), Web Services Description Language (WSDL) y el Universal Description, Discovery and Integration (UDDI) los cuales dan soporte a sofisticados modelos de comunicación entre varios nodos de la red. Estos protocolos permiten un proceso de comunicación más inteligente y colaborativo entre nodos.

Una base de datos distribuida (BDD) es un conjunto de múltiples bases de datos lógicamente relacionadas las cuales se encuentran distribuidas en diferentes espacios lógicos (por ejemplo un servidor corriendo 2 máquinas virtuales) e interconectados por una red de comunicaciones. Dichas BDD tienen la capacidad de realizar procesamiento autónomo, esto permite realizar operaciones locales o distribuidas. Un sistema de bases de datos distribuidas (SBDD) es un sistema en el cual múltiples sitios de bases de datos están ligados por un sistema de comunicaciones de tal forma que, un usuario en cualquier sitio puede acceder los datos en cualquier parte de la red exactamente como si estos fueran accedidos de forma local, según se muestra en la Figura 12.

Un sistema distribuido de bases de datos se almacena en varias computadoras; lo cual implica que existe un administrador del sistema responsable de la seguridad global y, a su vez, en cada una de las otras bases de datos hay un

administrador local quien llega a un nivel de autonomía local. Los principales factores que distinguen un SBDD de un sistema centralizado son los siguientes:

- Hay múltiples computadores, llamados sitios o nodos.
- Estos sitios deben de estar comunicados por medio de algún tipo de red de comunicaciones para transmitir datos y órdenes entre los sitios.

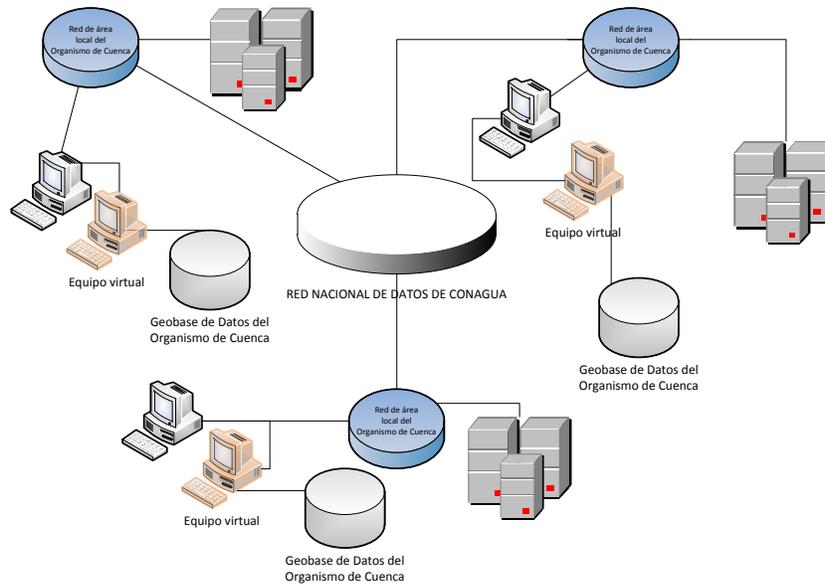


Figura 12. Diagrama de geobase de datos distribuida

Antes que se pensara en el cómputo distribuido se almacenaba la información de manera centralizada, pero con el paso del tiempo las necesidades aumentaron y esto produjo ciertos inconvenientes que no era posible solucionarlos o volverlos eficientes de la forma centralizada. Estos problemas impulsaron la creación de almacenamiento distribuido, los cuales hoy en día proveen características indispensables en el manejo de información; es decir, la combinación de las redes de comunicación y las bases de datos.

Hay varios factores que han hecho que las bases de datos evolucionen a bases de datos distribuidas. En el mundo de los negocios se ha dado una globalización y a la vez las operaciones de las empresas son cada vez más descentralizadas geográficamente. También el poder de las computadoras personales aumentó. Además la necesidad de compartir datos ha hecho que crezca el mercado de las bases de datos distribuidas. Diferentes usuarios tienen acceso sin interferir unos con otros; sin embargo, el SBDD debe sincronizar periódicamente las bases de

datos dispersas, para asegurar una uniformidad en la información (Castillo, 2002).

Este sistema está formado por las transacciones y los administradores de la base de datos distribuidos. Un SBDD implica un conjunto de programas que operan en diversas computadoras; así que se debe permitir que cada sitio almacene y mantenga su propia base de datos, facilitando de esta forma el acceso inmediato y eficaz a sus datos, mejorando con ello la fiabilidad si la computadora de un sitio no funciona, el resto de la red seguirá operando. De esta manera se logrará un control local de los datos en un sitio mejorando el grado de satisfacción de los usuarios con relación al SBDD.

El SBDD recibe las solicitudes de procesamiento de los programas de consulta o transacciones y las traduce en acciones para los administradores de la base de datos. El soporte completo para las bases de datos distribuidas requiere que una sola aplicación permita operar de forma independiente, ejecutada en varias máquinas diferentes, conectadas a través de varias redes diferentes, como si los datos estuviesen integrados en un solo RDBMS y ejecutados en una sola máquina; un usuario debe poder realizar operaciones sobre los datos desde su equipo local, como si el sitio no participase en el sistema distribuido.

1.5.9.2. Enfoque al problema de diseño de la base de datos distribuida.

Enfoque de arriba hacia abajo (top-down): Este enfoque es utilizado para aplicaciones nuevas y sistemas homogéneos. Parte desde el análisis de requerimientos a fin de definir el estilo conceptual y las vistas del usuario. A partir de esto se define un esquema conceptual global y los esquemas externos necesarios. Se prosigue con la fragmentación del sistema y enseguida con la localización de los fragmentos del sitio, creando las imágenes físicas. Finalmente se aplica el diseño físico en cada uno de los sitios, conforme se muestra en la Figura 13.

Diseño de abajo hacia arriba (bottom-up): Este diseño se utiliza para bases de datos existentes. Requiere de la selección de un modelo de datos común para describir el esquema global de la base de datos. Esto es resultado a que es posible que se utilicen diferentes RDBMS para su operación. Posteriormente se hace una traducción de cada esquema local en el modelo de datos común y por último se integra el esquema local al global.

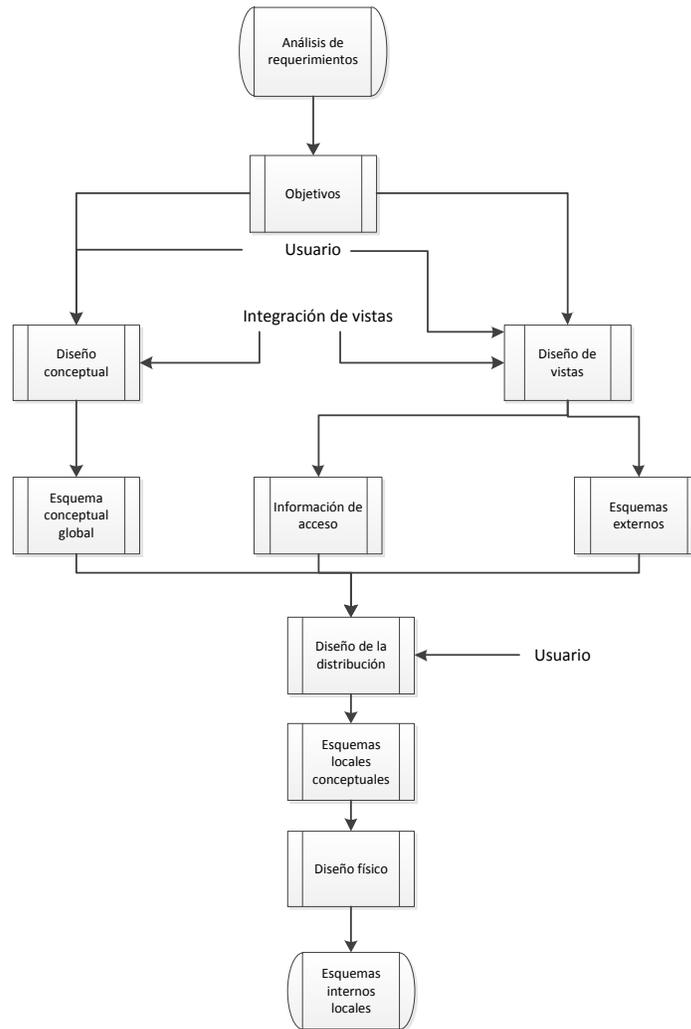


Figura 13. Enfoque top-down para el diseño de bases de datos distribuidas

1.5.9.3. Tipos de geobases de datos distribuidas (arquitectura).

Una base de datos distribuida es un conjunto de múltiples bases de datos lógicamente relacionadas las cuales se encuentran distribuidas entre diferentes sitios interconectados entre sí mediante una red de comunicaciones, los cuales tienen la capacidad de procesamiento autónomo lo cual implica que puede realizar operaciones locales o distribuidas; lo anterior de forma tal que un usuario en cualquier sitio puede acceder a los datos en cualquier parte de la red como si los tuviesen almacenados de forma local.

Una de las decisiones más importantes que el diseñador de bases de datos distribuidas debe tomar en cuenta es el posicionamiento de los datos en el sistema y el esquema bajo el cual lo desea hacer. Para esto existen cuatro alternativas principales: centralizada, replicada, fragmentada, e híbrida.

Centralizada: es similar al modelo de Cliente/Servidor en el sentido que el SBDD está centralizado en un lugar y los usuarios están distribuidos. Este modelo solo brinda la ventaja de tener el procesamiento distribuido ya que en sentido de disponibilidad y fiabilidad de los datos no se gana nada.

Replicada: consiste en que cada nodo debe tener su copia completa de la base de datos. Es fácil ver que este esquema tiene un alto costo en el almacenamiento de la información. Debido a que la actualización de los datos debe ser realizada en todas las copias, también tiene un alto costo de escritura, pero todo esto vale la pena si tenemos un sistema en el que se va a escribir pocas veces y leer muchas, y dónde la disponibilidad y fiabilidad de los datos sea de máxima importancia. Esta técnica permite la distribución de datos a través de dos o más geobases de datos de forma que las ediciones puedan ser efectuadas independientemente y sincronizadas periódicamente. Cuenta con seguridad interconstruida contra pérdida de datos, redundancia de datos y estabilidad sistémica. La replicación de geobases de datos requiere al menos de una geobase de datos, ArcSDE, versionada (ESRI, 2005).

Este esquema está estructurado de forma que cada nodo cuenta con su copia completa de la base de datos. Se intuye que este esquema tiene un alto costo en el almacenamiento de la información pues la base de datos se encuentra completa en n sitios. Adicionalmente, debido a que la actualización de los datos se debe realizar en todas las copias también tiene un alto costo de escritura, estos elementos deben de tomarse en cuenta en la parte de la implementación debiendo considerar si se tiene un sistema en el que se van a escribir pocas veces y leer muchas; pues ello depende de que la disponibilidad y fiabilidad de los datos sea de importancia primordial

Las geobases de datos empresariales operando sobre ArcSDE, están construidas sobre un RDBMS el cual incluye la tecnología para la replicación a nivel de base de datos relacional. Las geobases de datos, como cualquier otra aplicación construida sobre un RDBMS, pueden ser usadas para vincularse con otros sistemas que cuenten con soporte de bases de datos relacionales. El uso

de la replicación RDBMS en una geobase de datos requiere del conocimiento de cómo están implementadas las estructuras a nivel de la base de datos.

La creación de réplicas involucra desde la definición de los datos a replicar en una geobase de datos fuente hasta la corrida de un proceso para la creación de la réplica. El proceso de copia de datos en la geobase de datos fuente a la geobase de datos objetivo, crea una réplica en cada geobase. La réplica describe que dato ha sido replicado y contiene la información necesaria para sincronizar los cambios. La réplica, en la geobase de datos fuente, se denomina la “réplica padre” y la réplica, en la geobase objetivo, se llama “réplica hija”. Cada combinación de una réplica hija y una réplica padre es llamada “réplica par”, conforme a lo presentado en la Figura 14.



Figura 14. Replicación de Geobases de Datos

La replicación de geobases de datos permite replicar “datasets” específicos dentro de la geobase de datos. Así mismo permite aplicar filtros para definir cuáles “features” dentro de estos “datasets” están involucrados en la réplica.

La replicación de geobases de datos está construida sobre versiones. La réplica “padre” debe estar almacenada en una geobase de datos de tipo ArcSDE y los datos asociados deben estar “versionados”. En función del tipo, las réplicas “hijas” podrían requerir ser geobases de datos versionadas de ArcSDE. Cuando se crea una réplica, también se define la versión. Durante el proceso de sincronización, las ediciones aplicadas a esta versión pueden ser enviadas a la réplica relativa. Cambios recibidos de la réplica relativa también son importados en esta versión.

Fragmentada: Este modelo, mostrado en la Figura 15 para el caso CONAGUA, consiste en que solo hay una copia de cada elemento, pero la información está distribuida a través de los nodos. En cada nodo se aloja uno o más fragmentos disjuntos de la base de datos. Como los fragmentos no se replican esto disminuye el costo de almacenamiento, pero también sacrifica la disponibilidad y fiabilidad de los datos. Algo que se debe tomar en cuenta cuando se desea implementar este modelo es la granularidad de la fragmentación. La fragmentación se puede realizar también de tres formas:

- **Horizontal:** Los fragmentos son subconjuntos de una tabla; es decir, divide la base asignando a cada división a uno o varios fragmentos.
- **Vertical:** Los fragmentos son subconjuntos de los atributos con sus valores, implica la descomposición en varios subconjuntos.
- **Mixto:** La base de datos se divide por medio de una serie de relaciones de fragmentos, donde cada uno de estos tiene como resultado los procedimientos de las fragmentaciones horizontales y mixtas.

Una ventaja significativa de este esquema es que las consultas (SQL) también se fragmentan por lo que su procesamiento es en paralelo y más eficiente.

Esta técnica, para la distribución de datos, involucra la copia simple de datos de una geobase a otra. Esta técnica es utilizada para sistemas con requerimientos simples y puede ser utilizada donde los datos no estén versionados. No obstante, esta técnica no cuenta con seguridad interconstruida contra pérdida de datos, redundancia de datos y estabilidad sistémica.

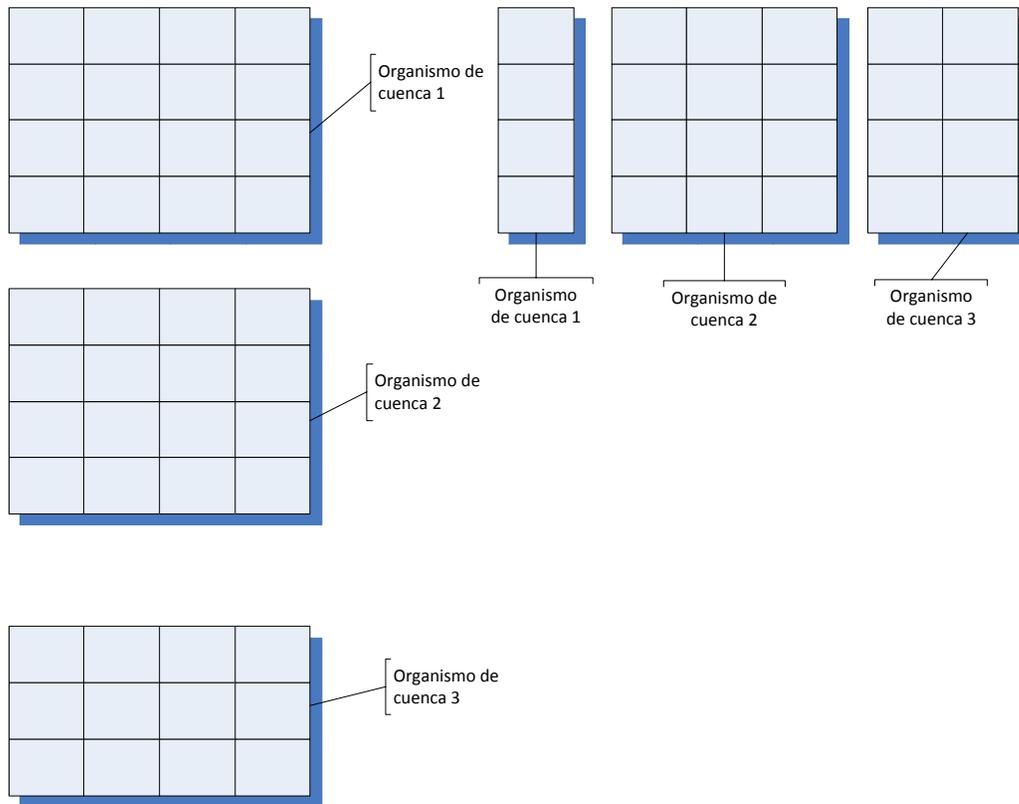


Figura 15. Fragmentación de geobases de datos

Para que una fragmentación sea correcta esta debe cumplir con las siguientes tres reglas:

1. **Compleción:** Si una relación R se fragmenta en R_1, R_2, \dots, R_n , cada elemento del conjunto R debe estar en algún R_i ; es decir R_n es completa si y solo si cada elemento de datos de datos R se encuentra en alguno de los R_i
2. **Reconstrucción:** Si una relación R se descompone en una serie de fragmentos R_1, R_2, \dots, R_n , se puede definir un operador relacional ∇ tal que

$$R = \nabla R_i \quad \forall R_i \in F_R.$$

Donde F_R es el fragmento que contiene a la relación R . El operador ∇ será diferente dependiendo de la forma de fragmentación. La reconstrucción de la relación a partir de sus fragmentos asegura la preservación de las restricciones definidas sobre los datos en forma de dependencias.

3. **Disyunción:** Si la relación R se descompone horizontalmente en una

serie de fragmentos R_1, R_2, \dots, R_n , entonces si un elemento está en R_i este elemento no puede estar en ningún $R_k \forall k \neq i$. En el caso de fragmentación vertical es necesario que se repitan las llaves primarias y esta condición solo se debe cumplir para el conjunto de atributos que no son llave primaria.

La primera regla es extremadamente importante para asegurar que los datos de la relación global se proyectan sobre los fragmentos sin pérdida alguna. Cabe mencionar que en el caso de fragmentación horizontal el elemento es una tupla; mientras que en el caso vertical se trata de un atributo.

La tercera regla asegura que los fragmentos horizontales sean disjuntos. Si una relación R se descompone verticalmente, sus atributos primarios clave normalmente se repiten en todos sus fragmentos.

Réplica y fragmentación de datos: esta se aplica de manera continua, es decir de una pequeña división que anteriormente se ha fragmentado también se puede hacer una réplica y estos fragmentos también se pueden volver a fragmentar.

Asignación de cada fragmento de datos

Supongamos que tenemos un conjunto de fragmentos $F = \{F_1, F_2, \dots, F_n\}$ y una red que consiste en un conjunto de sitios $S = \{S_1, S_2, \dots, S_m\}$. El problema de asignación determina la distribución óptima de F en S ; es decir, la asignación de cada fragmento entre los distintos sitios de la red. El atributo de optimicidad puede ser definido de acuerdo a dos medidas:

1. **Costo mínimo.** Consiste en el costo de la comunicación de datos, costo del almacenamiento y el costo de procesamiento. El objetivo es encontrar una función que minimice el costo.
2. **Rendimiento.** La estrategia de asignación se diseña para mantener una métrica de rendimiento. Las dos métricas más utilizadas son el tiempo de respuesta y el “throughput” (productividad).

Cuando una serie de datos se asignan, estos pueden replicarse para mantener una copia o varias idénticas. Por tanto, respecto a la replicación, en la asignación de fragmentos existen tres estrategias:

1. No soportar replicación. Cada fragmento reside en un solo sitio.

2. Soportar replicación completa. Cada fragmento reside en cada uno de los sitios.
3. Soportar replicación parcial. Cada fragmento reside en alguno de los sitios.

Se considera de gran utilidad la replicación cuando el número de consultas de solo lectura es mucho mayor que el número de consultas de actualizaciones.

Híbrida

Esta arquitectura es la combinación de las arquitecturas de partición y replicación. La relación se particiona y, a la vez, los fragmentos se replican en forma selectiva a través del sistema de base de datos.

Transparencia de la red: Es de suma importancia que los usuarios no requieran saber cómo se almacenan los datos en la base de datos; así el sistema debe poder ser capaz de ocultar esta característica de la distribución de datos. Así, todos los elementos de datos, las divisiones y las copias deben tener nombres únicos pero asegurándose que dos procedimientos no utilicen el mismo nombre para distintos elementos de datos. El sistema debe ser transparente en cuanto a la repetición y fragmentación de las bases de datos donde se ocultará al usuario; esto se logra creando un conjunto de claves para cada usuario.

Transacciones en base de datos distribuidas.

Qué ocurre cuando dos consultas tratan de actualizar el mismo elemento de datos o si el sistema falla durante la ejecución de una consulta. Se puede pensar que el concepto principal que debe manejar la base de datos es el de ejecución consistente de consultas. Por eso es que se introduce el concepto de una transacción que se utiliza dentro del área de bases de datos como una unidad básica de cómputo consistente y confiable.

El objetivo en el manejo de transacciones es por un lado tener una transparencia adecuada de las acciones concurrentes a una base de datos y, por otro, tener una transparencia adecuada en el manejo de las fallas que se pueden presentar en una base de datos. Las propiedades de una transacción son las siguientes⁶⁶:

1.-**Atomicidad.** Una transacción se trata como una unidad de operación. Así que, o todas las acciones de la transacción se realizan o ninguna de ellas se lleva a cabo.

⁶⁶ ACID por sus siglas en inglés (atomicity, consistency, isolation, durability)

2.-**Consistencia.** Es la correcta ejecución de la transacción. En otras palabras, una transacción es un programa correcto que lleva la base de datos de un estado consistente a otro con la misma característica.

3.-**Aislamiento.** Una transacción en ejecución no puede revelar sus resultados a otras transacciones concurrentes antes de su *commit*⁶⁷. Más aún, si varias transacciones se ejecutan concurrentemente, los resultados deben ser los mismos que si ellas se hubiesen ejecutado de forma secuencial.

4.-**Durabilidad.** Es la propiedad de las transacciones que asegura que una vez que una transacción hace su *commit*, sus resultados son permanentes y no pueden ser borrados de la base de datos.

Las transacciones proporcionan una ejecución atómica y confiable ante la presencia de fallas, una ejecución correcta cuando existen accesos de usuarios múltiples y un manejo correcto de réplicas.

1.5.9.4. Implementaciones de cada tipo de arquitectura de base de datos distribuida

Multi base de datos distribuida

Cuando una base de datos distribuida es muy homogénea se dice que es multi base de datos distribuida.

Base de datos Federada

Cuando una base de datos distribuida tiene mucha autonomía local se dice que es federada.

Objetivos comunes en la implementación de una BDD

Al implementar una base de datos distribuida se tienen ciertos objetivos comunes:

- **Transparencia de ubicación.** Permite a los usuarios tener acceso a los datos sin que tenga conocimiento de la ubicación de éstos. Se puede conseguir este nivel de transparencia al utilizar los administradores de transacciones distribuidas, los cuales son capaces de determinar la localización de los datos y de emitir acciones a los temporizadores

⁶⁷ Una sentencia COMMIT en SQL finaliza una transacción de base de datos dentro de un RDBMS y pone visibles todos los cambios a otros usuarios. El formato general es emitir una sentencia BEGIN WORK, una o más sentencias SQL, y entonces la sentencia COMMIT. Alternativamente, una sentencia ROLLBACK se puede emitir, la cual deshace todo el trabajo realizado desde que se emitió BEGIN WORK. Una sentencia COMMIT (ref. Rehimi) publicará cualquiera de los savepoints (puntos de recuperación) existentes que puedan estar en uso. En términos de transacciones, lo opuesto de commit para descartar los cambios "en tentativa" de una transacción, es un rollback.

apropiados, lo cual puede ejecutarse cuando los administradores de transacciones distribuidas poseen acceso a los directorios de localizaciones de los datos.

- **Transparencia de duplicación.** Para que la transparencia de duplicación sea posible, los administradores de transacciones deben traducir las solicitudes de procesamiento de transacción en acciones para el administrador de datos. Para las lecturas el administrador de transacciones selecciona uno de los nodos que almacena los datos y ejecuta la lectura. Para optimizar el proceso, el administrador de transacciones necesita información sobre el rendimiento de varios nodos respecto al sitio de consulta, así podrá seleccionar el nodo de mejor rendimiento. La actualización y escritura de datos duplicados suelen ser más complicadas, ya que el manejador de transacciones debe emitir una acción de escritura para cada uno de los temporizadores que almacena una copia de los datos.
- **Transparencia de concurrencia.** Cuando varias transacciones se ejecuten al mismo tiempo, los resultados de las transacciones no deberán afectarse. La transparencia de concurrencia se logra si los resultados de todas las transacciones concurrentes son consistentes de manera lógica con los resultados que se habrían obtenido si las transacciones se hubieran ejecutado una por una, en cualquier orden secuencial.
- **Transparencia de fallas.** Significa que a pesar de fallas las transacciones sean procesadas de un modo correcto. Frente a una falla, las transacciones deben ser atómicas, significa que se procesen todas o ninguna de ellas. Para este tipo de problemas es importante tener resguardo de la base de datos, y así poder restaurarla cuando sea conveniente. El sistema debe detectar cuándo falla una localidad y tomar las medidas necesarias para recuperarse del fallo. El sistema no debe seguir utilizando la localidad que falló. Por último, cuando se recupere o repare esta localidad, debe contarse con mecanismos para reintegrarla al sistema con el mínimo de complicaciones.
- **Localidad del procesamiento.** Los datos se deben distribuir lo más cerca posible de las aplicaciones que los usan para maximizar la localidad del procesamiento, este principio responde a minimizar el acceso remoto a los datos. Diseñar una distribución que maximice localidad del procesamiento puede hacerse añadiendo la cantidad de referencias locales y remotas correspondientes a cada fragmentación candidata y asignar la fragmentación eligiendo la mejor solución. Independencia de

configuración. La independencia de configuración permite añadir o reemplazar hardware sin tener que cambiar componentes de software existentes en el sistema de base de datos distribuida.

- **Particionado de la base de datos.** La base de datos se distribuye de modo que no haya traslape o duplicación de los datos que se encuentran en las diferentes localidades, como no hay duplicaciones de los datos, se evitan los costos asociados con el almacenamiento y mantenimiento de datos redundantes. Si un mismo segmento de datos se usa en más de una localidad se ve limitada la disponibilidad de los datos. La fiabilidad también puede verse limitada cuando se produce un fallo en el sistema de cálculo de una localidad se afecta la disponibilidad de los datos de esa localidad no estén disponible para los usuarios en cualquier parte del sistema.
- **Fragmentación de datos.** Consiste en subdividir las relaciones y distribuir las entre los sitios de la red, tiene como objetivo buscar formas alternativas de dividir una las instancias (tablas) de relaciones en otras más pequeñas. La fragmentación se puede realizar por tuplas individuales (fragmentación horizontal), por atributos individuales (fragmentación vertical) o una combinación de ambas (fragmentación híbrida). El principal problema de la fragmentación radica en encontrar la unidad apropiada de distribución. Una relación no es una buena unidad por muchas razones. Normalmente las vistas de una relación están formadas por subconjuntos de relaciones. Además, las aplicaciones acceden localmente a subconjuntos de relaciones. Por ello, es necesario considerar a los subconjuntos de relaciones como unidad de distribución. Al descomponer una relación en fragmentos, tratados cada uno de ellos como una unidad de distribución, permite el proceso concurrente de las transacciones. El conjunto de estas relaciones, provocará la ejecución paralela de una consulta al ser dividida en una serie de subconsultas que operará sobre los fragmentos. Cuando las vistas definidas sobre una relación son consideradas como unidad de distribución que se ubican en diferentes sitios de la red, podemos optar por dos alternativas diferentes:
 1. La relación no estará replicada y se almacena en un único sitio
 2. Existe réplica en todos o algunos de los sitios en los cuales reside la aplicación.

Las consecuencias de esta estrategia son la generación de un volumen de accesos remotos que pueden ser innecesarios con un mal manejo de estas réplicas. Además, las réplicas innecesarias pueden causar problemas en la ejecución de las actualizaciones y puede no ser deseable si el espacio de almacenamiento es limitado. Los inconvenientes de la fragmentación se deben a que si alguna de las consultas está definida por fragmentos mutuamente exclusivos; al recuperar los datos de dos fragmentos situados en sitios diferentes será necesario transmitir los datos de un sitio a otro y realizar sobre ellos la operación de unión (“join”) lo cual puede ser costoso; esto se debe al control semántico; es decir, cuando los atributos implicados en una relación se descomponen en diferentes fragmentos, y estos se ubican en sitios diferentes, es necesario hacer búsquedas en un gran número de sitios.

Sistemas comerciales. Inicialmente se consideraron equipos con hardware dedicado y entorno centralizado, en un área geográficamente distribuida y sistemas distintos. Los ejemplos son: sistemas de reservación de camiones y líneas aéreas, aplicaciones bancarias, gestión de grandes almacenes. En cuanto al manejo de una WAN, Internet es un claro ejemplo de una muy amplia gama de servicios como: correo electrónico, servicios de noticias, transferencia de archivos, entre muchos otros. En cuanto a aplicaciones multimedia se refiere se pueden encontrar ejemplos como videoconferencia, tele vigilancia, juegos multiusuario, enseñanza asistida por computadora.

En un sistema de bases de datos distribuidas, existen varios factores que deben ser tomados en cuenta pues estos definen la arquitectura del sistema que se debe de utilizar:

- **Distribución:** Si los componentes del sistema estarán localizados en la misma computadora o no.
- **Heterogeneidad:** Existen en él componentes que se ejecutan en diversos sistemas operativos o en diferentes fuentes de datos.
- **Autonomía:** Se puede presentar en diferentes niveles:
 - 1) **Autonomía de diseño:** Capacidad de un componente del sistema para decidir cuestiones relacionadas a su propio diseño.
 - 2) **Autonomía de comunicación:** Capacidad de un componente del sistema para decidir cómo y cuándo comunicarse con otros Sistemas de Bases de Datos.
 - 3) **Autonomía de ejecución:** Capacidad de un componente del sistema para ejecutar operaciones locales como quiera.

Aplicaciones específicas.

Ejemplo 1. Considérese la operación de la CONAGUA que, por simplicidad, se supondrá que tiene tres Organismos de Cuenca (OC); en cada OC, una computadora controla las terminales de la misma y el sistema de cuentas de usuarios. Cada computadora con su sistema de cuentas local en cada OC constituye un “sitio” de la base de datos; las computadoras están conectadas por una LAN. Durante las operaciones normales, las aplicaciones en las terminales de los OC que solo necesitan acceder a la base de datos del mismo OC. Como solo accesan la misma red local, se les llaman aplicaciones locales.

Desde el punto de vista tecnológico, aparentemente lo importante es la existencia de algunas transacciones que accedan a información en más de una OC. Estas transacciones son llamadas transacciones globales o transacciones distribuidas. La existencia de transacciones globales es considerada como una característica que ayuda a discriminar entre las bases de datos y un conjunto de base de datos locales.

Una típica transacción global sería una consulta de datos relativos a una región hidrológica y otra adyacente. Esta aplicación requiere la actualización de datos en dos diferentes OC y se debe asegurar la real actualización en ambos sitios o en ninguno. Asegurar el buen funcionamiento de aplicaciones globales es una tarea difícil.

1.5.9.5. Geobases de datos distribuidas operando en servidores virtuales

El software que se ejecuta en un entorno virtualizado afecta al rendimiento de cualquier aplicación en algún grado. Como cualquier otra aplicación, el rendimiento de ArcGIS Server recibe un impacto negativo por la virtualización. A medida que la carga de trabajo en el procesamiento del servidor se hace más pesada, la disminución en el rendimiento se hace más evidente.

Algunas aplicaciones de uso intensivo de la CPU también son afectadas negativamente en un entorno virtual. El rendimiento puede variar según el proveedor de virtualización, por la operación que se ejecuta. Es importante destacar el hecho de que una configuración deficiente de los entornos virtuales puede conducir a los resultados de rendimiento menor de hasta un 60%.

Por lo tanto, es muy importante seguir las mejores prácticas sobre la configuración de entornos de virtualización, conforme lo sugerido por cada proveedor, así como las pruebas de rendimiento. Es recomendable analizar

previamente las evaluaciones de ambientes virtualizados con el fin de diseñar correctamente el tamaño y configuración del sistema⁶⁸.

En muchos casos, el rendimiento es sólo una consideración menor, porque los beneficios de la virtualización superan las implicaciones de rendimiento. El impacto de la virtualización en el rendimiento general se toma en cuenta a la hora de dimensionar el sistema correctamente para apoyar el uso previsto.

Algunas investigaciones consideran que el precio de virtualización y soporte técnico implica una amplia gama de cuestiones y preguntas que influyen al momento de implementar una solución. A continuación se presentan algunas de las preguntas que se deben responder antes de implementar un ambiente virtualizado.

1. ¿La tecnología proporciona herramientas de conversión de físico a virtual?
2. ¿Cuenta con herramientas administrativas tales como consola de gestión?
3. ¿La tecnología puede migrar de una máquina virtual desde un host físico a otro sin tener que darlo de baja?
4. ¿La tecnología elimina el host físico como un solo punto de falla, es decir soporta alta disponibilidad?
5. ¿La tecnología puede soportar diferentes sistemas operativos con el ambiente virtualizado, es decir, permite máquinas virtuales multiplataforma?
6. ¿La tecnología soporta paravirtualización? Esto permite una ejecución optimizada.
7. ¿La tecnología soporta tarjetas de red⁶⁹, aislamiento unicast, Internet Small Computer System Interface (iSCSI), Simple Network Management Protocol y otras compatibilidades de almacenamiento y manejo de redes externas?
8. ¿La tecnología permite múltiples CPUs para trabajar en un solo proceso, es decir permite multiprocesamiento simétrico?

⁶⁸ Para ver la guía del desarrollador favor de referirse a
<http://www.vmware.com/resources/techresources/10091/>

⁶⁹ Network interface cards (NICs)

1.5.9.6. Ventajas y desventajas de cada tipo de arquitectura

En la Tabla 4 se presenta una comparación entre los diferentes tipos de redes distribuidas en donde se presentan tanto las ventajas como las desventajas de cada uno de estos tipos.

Tabla 4. Comparación entre la ventajas y desventajas de los tipos de redes distribuidas.

TIPO	VENTAJAS	DESVENTAJAS
Centralizada	<p>1. Rendimiento. Los datos de manera general están ubicados en un solo sitio que es, por lo general, el de mayor demanda, así como los sistemas tienen que trabajar solamente con una sola implementación de la base de datos.</p>	<p>1. Carencia de estándares. A la fecha no existen herramientas o metodologías específicas que ayuden a los usuarios a convertir un RDBMS centralizado en un RDBMS distribuido.</p>
Replicada	<p>1. Disponibilidad. Una falla en una parte del sistema; es decir, en alguna base de datos solo afectará a ese fragmento, en lugar de afectar a toda la base de datos.</p> <p>2. Rendimiento. Los datos de manera general están ubicados en el sitio con mayor demanda, así como los sistemas que trabajan en paralelo, lo cual permite balancear la carga en los servidores.</p> <p>3. Economía. Es más económico crear una red con varios servidores pequeños, que tener un solo servidor muy poderoso.</p> <p>4. Modularidad. Se pueden modificar, agregar o quitar sistemas de la base de datos distribuida sin afectar a los demás sistemas (módulos).</p>	<p>1. El rendimiento puede ser peor para el procesamiento distribuido que para el procesamiento centralizado. Depende de la naturaleza de la carga de trabajo, la red, el DDBMS y las estrategias utilizadas de concurrencia y de falla, así como las ventajas del acceso local a los datos y de los procesadores múltiples, ya que éstos pueden ser abrumados por las tareas de coordinación y de control requeridas. Tal situación es probable cuando la carga de trabajo necesita un gran número</p>

	<p>5. Aumento del paralelismo. Varios nodos pueden realizar consultas en paralelo sobre la misma tabla. Cuantas más réplicas existan de la tabla, mayor será la posibilidad de que el dato se encuentre en el nodo desde el que se realiza la consulta.</p> <p>6. Los datos duplicados aumentan su confiabilidad. Cuando falla una computadora, se pueden obtener los datos extraídos de otras computadoras. Los usuarios no dependen de la disponibilidad de una sola fuente para sus datos.</p>	<p>de actualizaciones concurrentes sobre datos duplicados, y que deben estar muy distribuidos.</p> <p>2. Complejidad. Se debe asegurar que la base de datos sea transparente, se debe lidiar con varios sistemas diferentes que pueden presentar dificultades únicas.</p> <p>3. El diseño de la base de datos se tiene que trabajar tomando en cuenta su naturaleza distribuida, por lo cual no podemos pensar en hacer "joins" que afecten varios sistemas.</p> <p>4. Economía. La complejidad y la infraestructura implica que se necesitará una mayor mano de obra.</p> <p>5. Seguridad. El trabajo que requiere la seguridad del sistema se y hace mayor en función al número de equipos en operación ya sea en forma síncrona o asíncrona por lo que se debe trabajar con mucho más detalle en la seguridad de la infraestructura así como cada uno de los sistemas.</p> <p>6. Falta de experiencia. Las bases de datos distribuidas son un campo relativamente nuevo y poco común por lo cual no existe, en particular en CONAGUA,</p>
--	---	--

		<p>mucho personal con experiencia o conocimientos adecuados</p> <p>7. Aumento en la sobre carga en las actualizaciones. El sistema debe asegurar que todas las réplicas de la tabla sean consistentes. Cuando se realiza una actualización sobre una de las réplicas, los cambios deben propagarse a todas las réplicas de dicha tabla a lo largo del sistema distribuido.</p>
<p>Particionada</p>	<p>1. Mejor rendimiento que el que se obtiene por un procesamiento centralizado. Los datos pueden colocarse cerca del punto de su utilización, de forma que el tiempo de comunicación sea más corto. Varias computadoras operando en forma simultánea pueden entregar más volumen de procesamiento que una sola computadora.</p> <p>2. Estructura de forma organizacional. Los fragmentos de la base de datos se ubican en los Organismos de Cuenca a los cuales tiene relación.</p> <p>3. Autonomía local. Cada Organismo de Cuenca puede controlar los datos que le pertenecen.</p> <p>4. Disponibilidad. Una falla en una parte del sistema; es decir, en alguna base de datos solo afectará a ese fragmento, en lugar de afectar a toda la base de datos.</p> <p>5. Rendimiento. Los datos de manera general están ubicados en</p>	<p>1. El procesamiento de base de datos distribuida sobre este esquema puede resultar menos confiable que el procesamiento centralizado. De nuevo, depende de la confiabilidad de las computadoras de procesamiento, de la red, del RDBMS, de las transacciones y de las tasas de error en la carga de trabajo. Un sistema particionado podría estar menos disponible que uno centralizado.</p> <p>2. Integridad. Se hace difícil mantener y aplicar las reglas de integridad a través de la red dado que resulta ser muy caro en términos de transmisión de datos.</p> <p>3. Falta de experiencia. Las bases de</p>

	<p>el sitio con mayor demanda, así como los sistemas que trabajan en paralelo, lo cual permite balancear la carga en los servidores.</p> <p>6. Economía. Es más económico crear una red con varios servidores pequeños, que tener un solo servidor muy poderoso.</p> <p>7. Modularidad. Se pueden modificar, agregar o quitar sistemas de la base de datos distribuida sin afectar a los demás sistemas (módulos).</p> <p>8. Pueden variar su tamaño de un modo más sencillo. Se pueden agregar computadoras adicionales a la red conforme aumentan el número de usuarios y su carga de procesamiento. A menudo es más fácil y más barato agregar una nueva computadora más pequeña que actualizar una computadora única y centralizada. Después, si la carga de trabajo se reduce, el tamaño de la red también puede reducirse.</p>	<p>datos distribuidas son un campo relativamente nuevo y poco común por lo cual no existe, en particular en CONAGUA, mucho personal con experiencia o conocimientos adecuados.</p> <p>4. El diseño de la base de datos se vuelve más complejo. Además de las dificultades que generalmente se encuentran al diseñar una base de datos, el diseño y la implementación de una base de datos distribuida debe considerar la fragmentación, replicación y ubicación de los fragmentos en sitios específicos.</p>
<p>Híbrida</p>	<p>1. Mayor flexibilidad en la estructura pues además de operar en una forma organizacional cuenta con la ventaja de que los datos más importantes o esenciales están ubicados en un solo sitio que es, por lo general, el de mayor demanda.</p> <p>2. Autonomía local. Cada Organismo de Cuenca puede controlar los datos que le pertenecen.</p> <p>3. Disponibilidad. Una falla en una parte del sistema; es decir, en alguna base de datos solo afectará a ese fragmento, en lugar de afectar a toda la base de datos.</p>	<p>1. Integridad. Se hace difícil mantener y aplicar las reglas de integridad a través de la red dado que resulta ser muy caro en términos de transmisión de datos.</p> <p>2. Falta de experiencia. Las bases de datos distribuidas son un campo relativamente nuevo y poco común por lo cual no existe, en particular en CONAGUA, mucho personal con experiencia o</p>

	<p>4. Rendimiento. Los datos de manera general están ubicados en el sitio con mayor demanda, así como los sistemas que trabajan en paralelo, lo cual permite balancear la carga en los servidores.</p> <p>5. Economía. Es más económico crear una red con varios servidores pequeños, que tener un solo servidor muy poderoso.</p> <p>6. Modularidad. Se pueden modificar, agregar o quitar sistemas de la base de datos distribuida sin afectar a los demás sistemas (módulos).</p>	<p>conocimientos adecuados.</p> <p>3. El diseño de la base de datos se vuelve más complejo. Además de las dificultades que generalmente se encuentran al diseñar una base de datos, el diseño y la implementación de una base de datos distribuida debe considerar la fragmentación, replicación y ubicación de los fragmentos en sitios específicos.</p>
--	--	---

1.5.9.7. Plataformas sobre las que operan las bases de datos distribuidas

Tanto las bases como las geobases de datos distribuidas requieren para su operación el uso de un RDBMS. En el mercado de software existen varias marcas que sirven para este propósito. En la Tabla 5 se muestra una descripción de los RDBMS más importantes, que sirven como plataforma base para este propósito.

Tabla 5. Descripción de algunos de las principales marcas vendedoras de software para RDBMS

Nombre del manejador	Descripción
<ul style="list-style-type: none"> • MySQL⁷⁰ 	<p>Disponible para la mayoría de las plataformas de sistemas operativos. Su bajo consumo lo hacen apto para ser ejecutado en una máquina con escasos recursos sin ningún problema. El conjunto de aplicaciones Apache-PHP-MySQL es uno de los más utilizados en aplicaciones en ambiente Web.</p>

⁷⁰ <http://www.mysql.com/>

	<p>Cuenta con una muy buena velocidad a la hora de realizar las operaciones.</p> <p>Soporta manejo de datos clusterizado (MySQL Cluster 7.2).</p> <p>My SQL ofrece el potencial de manejo de “Big Data” para que las organizaciones puedan revolucionar sus operaciones. Es un RDBMS de código abierto más popular del mundo y el más utilizado tanto en la web como en la nube. MySQL es un componente clave de muchas plataformas de datos grandes. En muchos casos se recomienda el uso de MySQL y Hadoop en una infraestructura de Big Data.</p>
<ul style="list-style-type: none"> • Microsoft SQL Server⁷¹ 	<p>Fácil de configurar e instalar, admite la administración de bases de datos distribuidas. Es utilizado por una gran cantidad de usuarios haciendo esto que sea muy popular y de fácil acceso. Diseñado para trabajar con grandes bases de datos.</p> <ul style="list-style-type: none"> • Soporta todas las funciones que se esperan de un servidor sobre un lenguaje de diseño de bases de datos muy completo (TRANSACT SQL). • SQL Server ayuda a liberar conocimiento innovador nuevo con el descubrimiento extensivo de datos a través de bancos de datos estructurados, no estructurados, y en la nube con el apoyo de la BI administrada de auto-servicio, datos fiables, consistentes con soluciones de análisis y data warehousing de gran escala. • Permite administración de

⁷¹ <http://www.microsoft.com/es-xl/sqlserver/default.aspx>

	<p>información a más de 500 TB a través de la arquitectura de procesamiento masivo en paralelo (MPP por sus siglas en inglés) con Parallel Data Warehouse.</p> <ul style="list-style-type: none"> Incrementa la escala de almacenamiento con funcionalidades como Remote Blob Storage con la que se pueden dividir las tablas para que tengan más de 15,000 particiones.
<ul style="list-style-type: none"> Oracle⁷² 	<p>Diseñado para trabajar con grandes bases de datos, es el motor de base de datos relacional más usado a nivel mundial.</p> <ul style="list-style-type: none"> Puede ejecutarse en todas las plataformas, desde una PC hasta un supercomputador. Soporta todas las funciones que se esperan de un servidor sobre un lenguaje de diseño de bases de datos muy completo (PL/SQL). Permite el uso de particiones para la mejora de la eficiencia, de replicación Admite la administración de bases de datos distribuidas. El sistema orientado a objetos ha comenzado a evolucionar en esta dirección, añadiendo tipos de clases, referencias, tablas anidadas, matrices y otras estructuras de datos complejas. Base de datos con orientación hacia INTERNET. <p>Desventajas</p> <ul style="list-style-type: none"> La implementación actual no ofrece una ventaja clara en eficiencia, como sería de esperar y provocan la incompatibilidad de los

⁷² <http://www.oracle.com/index.html>

	<p>diseños que aprovechan las nuevas características con otras bases de datos.</p> <ul style="list-style-type: none"> • Desde el lanzamiento original de la 8 se sucedieron varias versiones con correcciones, hasta alcanzar la estabilidad en la 8.0.3. El motivo de tantos fallos fue la remodelación del sistema de almacenamiento por causa de la introducción de extensiones orientadas a objetos. • El mayor inconveniente es quizás su precio. Incluso las licencias la necesidad de ajustes. • Un error frecuente consiste en pensar que basta instalar el Oracle en un servidor y enchufar directamente las aplicaciones clientes. Un Oracle mal configurado puede ser desesperantemente lento. • Elevado el costo de la formación, y sólo últimamente han comenzado a aparecer buenos libros sobre asuntos técnicos distintos de la simple instalación y administración.
<ul style="list-style-type: none"> • PostgreSQL⁷³ 	<p>Muy buen motor, es seguro y fiable, maneja integridad referencial cosa que no hacía MySQL, hasta su versión 4, es multiplataforma y de libre distribución. Es un sistema de gestión de base de datos relacional orientada a objetos y libre; funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo a la vez a el sistema. PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos</p>

⁷³ http://www.postgresql.org.es/sobre_postgresql

para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema el cual continuará funcionando. Como muchos otros proyectos de código abierto, el desarrollo de PostgreSQL no es manejado por una empresa y/o persona, sino que es dirigido por una comunidad de desarrolladores que trabajan de forma desinteresada, altruista, libre y/o apoyada por organizaciones comerciales. Dicha comunidad es denominada el PGDG (PostgreSQL Global Development Group). PostgreSQL permite que mientras un proceso escribe en una tabla, otros accedan a la misma tabla sin necesidad de bloqueos. Cada usuario obtiene una visión consistente de lo último a lo que se le hizo commit. Esta estrategia es superior al uso de bloqueos por tabla o por filas común en otras bases, eliminando la necesidad del uso de bloqueos explícitos. Mediante un sistema denominado MVCC (Acceso concurrente multiversión, por sus siglas en inglés) PostgreSQL permite que mientras un proceso escribe en una tabla, otros accedan a la misma tabla sin necesidad de bloqueos. Cada usuario obtiene una visión consistente de lo último a lo que se le hizo "commit".

Desventajas.

- Puntos de recuperación dentro de transacciones. Actualmente, las transacciones abortan completamente si se encuentra un fallo durante su ejecución.
- No soporta tablespaces para definir dónde almacenar la base de

	<p>datos, el esquema, los índices, etc.</p> <ul style="list-style-type: none"> • El soporte de orientación a objetos es una simple extensión que ofrece prestaciones como la herencia, no un soporte completo.
<ul style="list-style-type: none"> • IBM DB2⁷⁴ 	<p>El RDBMS DB2 ofrece un alto rendimiento, ofreciendo escalabilidad y la fiabilidad. Opera sobre plataformas Linux, UNIX y Windows para z / OS.</p> <ul style="list-style-type: none"> • La arquitectura física es muy similar a la de Oracle. • También puede ejecutarse en varias plataformas. • Potente. • Es similar a Oracle en el precio. • Los procedimientos almacenados de DB2 deben programarse en lenguajes externos: C, Java y Visual Basic. Esto se debe a que hace depender de otro compilador. Que casi siempre está ligado a la plataforma o sistema operativo. • DB2 está optimizado para operar en un desarrollo de cómputo de nube, sistemas de potencia⁷⁵ alta capacidad de almacenamiento y SAP
<ul style="list-style-type: none"> • IBM Informix⁷⁶ 	<p>Es una familia de productos de bases de datos que incluye un sistema de administración de bases de datos relacionales basado en SQL, un lenguaje de cuarta generación y herramientas para la inclusión de SQL en programas de aplicación. Informix es uno de los RDBMS más utilizadas en el mundo,</p>

⁷⁴ <http://www-01.ibm.com/software/data/db2/>

⁷⁵ <http://www-01.ibm.com/software/data/db2/power-systems/>

⁷⁶ <http://www-01.ibm.com/software/data/informix/>

	<p>con los usuarios que van desde las corporaciones más grandes del mundo a empresas de nueva creación; incorpora conceptos de diseño que son significativamente diferentes de las plataformas relacionales tradicionales, lo que resulta en niveles extremadamente altos de rendimiento y disponibilidad, capacidades distintivas en la replicación de datos y escalabilidad.</p>
<ul style="list-style-type: none"> • Sybase ASE⁷⁷ 	<p>Es una base de datos corporativa, programable, soporta gran cantidad de datos y usuarios. Soporta multiprocesadores. Es un software versátil, de clase empresarial RDBMS que es especialmente bueno en el manejo de cargas de trabajo. Es utilizado de forma intensiva en el mundo financiero (bancos, bolsas de valores, compañías de seguros), en el comercio electrónico, así como en el área de prácticamente todos los demás. La versión más reciente de ASE es la 15.7 (publicado en septiembre del 2011); la versión anterior es la versión 15.5. ASE 15.7 también se conoce como “el lanzamiento de SAP”, ya que esta es la versión de ASE que SAP está utilizando para apoyar el paquete de ERP Business Suite en la parte superior de Sybase ASE. Es un producto patentado, un software comercial. Sin embargo, también existen versiones libres⁷⁸.</p>

⁷⁷ Adaptive Server Enterprise

⁷⁸ <http://www.sybase.com.mx/products/databasemanagement/adaptiveserverenterprise>

1.5.9.8. Ventajas de utilizar ArcGIS Server en un sistema virtualizado.

Las organizaciones que han usado ArcGIS Server en un entorno virtualizado han detectado una gran variedad de ventajas, incluyendo la capacidad de ejecutar diferentes versiones de ArcGIS Server en una única máquina física. Esto es especialmente útil en entornos de desarrollo, así como la aplicación de estrategias de migración de versiones. La reducción de los riesgos de seguridad a la infraestructura informática en general se logra mediante el aislamiento exterior que enfrenta aplicaciones de ArcGIS Server en máquinas virtuales específicos.

El cumplimiento con el plan de la CONAGUA ante recuperación de desastres requieren de aplicaciones ArcGIS Server y servicios que puedan ser rápidamente configurados en el caso de una catástrofe. Esto puede ser efectuado mediante un desarrollo acelerado y el despliegue de equipo autónomo de las aplicaciones de ArcGIS Server.

La maximización en el uso de la infraestructura de hardware y la informática simplifica la implementación y configuración de servidores así como la ampliación del sistema. Por ejemplo, las nuevas máquinas podrían ser rápidamente configuradas como contenedores de objetos de servidor (SOC).

La virtualización en las TI de los SIG tienden a convertirse, en la práctica, en un estándar con ArcGIS Server como apoyo a todos los componentes de un entorno virtualizado sobre plataformas compatibles. La elección de un entorno completamente virtualizado depende, en gran medida, de la estrategia general de TI de la CONAGUA. La virtualización ofrece beneficios claros en sus características y funciones incluidas en las soluciones de virtualización disponibles. Las características con que cuentan las tecnologías de virtualización de ejecutar varias versiones de ArcGIS Server, la seguridad, la continuidad del negocio, el desarrollo rápido de aplicaciones y la simplificación en la configuración de servidores representan algunas de las más importantes ventajas de la virtualización.

Como sistema operativo de servidor, para el manejo de la Geobase de Datos, en estructura de cluster, se utiliza el Windows 2003 Enterprise Edition, con Service Pack 2, con MSSQL Server 2005 y ArcSDE 9.1 para Microsoft SQL Server. La descripción de este manejador se presenta en la Tabla 5.

En el presente trabajo, para generar los servidores virtuales se utilizará como máquina física, a nivel central, un servidor HP tipo Blade con sistema operativo Microsoft Windows 2008 (64-bit y R2) con Hyper V en el cual se montarán

geobases de datos sobre MSSQL Server 2005 y ArcSDE 9.1 para Microsoft SQL Server. Este último software se utilizara debido a que la Gerencia de Informática y Telecomunicaciones lo califica como oficial de acuerdo con la plataforma de TI que se encuentra instaurada en la CONAGUA. Lo anterior debido a que no se cuenta con servidores geoespaciales en cada OC.

En los Organismos de Cuenca que cuenten con personal de SIGA y con equipo desktop con suficiente capacidad se hará una configuración semejante a la señalada en el párrafo anterior solo que el servidor virtual se implementará sobre plataforma VMWare player.

1.5.9.9. Estudios de caso.

El Ministro de Ciencia, Tecnología e Innovación de Dinamarca, Helge Sander afirmó que "...la Administración Pública [danesa] en su estrategia de software ha puesto un gran énfasis sobre la necesidad de una competencia eficaz y coherente entre los sistemas de TI⁷⁹ públicos, y por lo tanto los datos y documentos pueden ser intercambiados a través de diferentes tipos de organizaciones y diferentes tipos de software." (Fishenden, 2005)

El Ministro del Interior, Otto Schily, en Alemania, señaló que cualquier compañía proveedora de sistemas "...está de acuerdo con publicar especificaciones para interfaces y formatos de datos, además de apoyar los estándares abiertos en sus productos..." estas afirmación dará la mayor flexibilidad a las agencias [alemanas] para la construcción de sus sistemas de información." (Fishenden, 2005)

El programa IDA (Intercambio de Datos entre Administraciones) de la Unión europea también ha reconocido las ventajas de la vinculación de sistemas de administración de datos, tanto ahora como en el futuro, para satisfacer las exigencias de interoperabilidad.

Dr. Dai del United States Environmental Protection Agency National Computer Center (USEPA) presentó una ponencia dentro de la ESRI International Users Conference 2010⁸⁰ en la cual mostró la forma en cómo ha implementado la tecnología Vmware.

⁷⁹ TI: Tecnologías de Información.

⁸⁰ <http://proceedings.esri.com/library/userconf/proc10/>

Adam Ziegler presentó en la ESRI International Users Conference 2010 la forma en cómo el West Chester County GIS recientemente migró de un ambiente físico a uno virtual.

Sidney Pendelberry mostró, en la ESRI International Users Conference 2010, cómo el Rochester Institute of Technology está haciendo uso de tecnología de virtualización Vmware server para desarrollar tareas ArcGIS de cómputo intensivo (navegación, geoprocésamiento, localización de direcciones, etc.) sobre un ambiente de servidores Microsoft Windows 2008 High Performance Computing (HPC) en cluster. Una hoja de cálculo, de Excel, es usada para almacenar tareas y macros a fin de enviar tareas al ambiente computacional como trabajos “tracks” de progreso y trabajo distribuido entre el ambiente de servidores virtualizado. La virtualización permite adicionar y mover nodos o servidores para una utilización máxima del servidor físico.

La gestión de emergencias en las Islas Baleares por sus peculiaridades territoriales y, específicamente, por el peso del sector turístico en su economía. El Sistema del Sistema Integral de Emergencias de las Islas Baleares (SEIB-112) administrado por Geibsa (Gestió d'Emergencies de les Illes Balears) juega con un factor de valor añadido en el desarrollo de su labor: el posicionamiento. El avance es posible gracias a la puesta en producción del módulo GIS de Storm, denominado MapViewer, que posibilita la localización gráfica de los incidentes sobre unas capas cartográfica. De hecho y como explicó Job Torres, jefe del área de Sistemas del SEIB-112, “...estamos creando una GeoBase a partir de diferentes fuentes de datos, como el INE y la CMT, pero también de otras adicionales como el proyecto Cartociudad, de forma que cuando el Centro de Emergencias recibe una llamada pueda localizar la ubicación del incidente de forma inmediata”.⁸¹

1.6. Identificación de elementos relacionales.

1.6.1. Diseño metodológico.

Como metodología base se utilizará el enfoque de sistemas (ES) el cual es aplicado debido a que mediante un análisis estructurado de detección de necesidades se facilita una visión Inter, Multi y Transdisciplinaria la cual ayudará a analizar y desarrollar la implementación de manera integral permitiendo

⁸¹ <http://pruebas.channelpartner.es/Casos/200807170011/-SEIB-112-respuesta-rapida-y-localizada.aspx>

identificar y comprender con mayor claridad y profundidad los problemas que presente sus múltiples causas y consecuencias. Al efecto se efectuarán las siguientes etapas:

- Desarrollar una descripción de las actividades de los Organismos de Cuenca, que bajo su revisión se implementará un modelo de tareas primarias. Dependiendo de la escala del estudio, puede ser necesario derivar varios modelos de actividad a distintos niveles de resolución, con objeto de describir completamente las necesidades de información.
- Derivar categorías de información requeridas para soportar las actividades de los modelos y las actividades particulares de las cuales la información puede ser obtenida.
- Para una estructura de organización en particular, definir las reglas de dirección en términos de la responsabilidad de toma de decisiones sobre las actividades que un área operativa existente tenga asignadas⁸².
- Definir las necesidades de un sistema de información que solucione los requisitos de desempeño de las actividades que soportan a cada sistema, tal que una red distribuida coherente pueda ser implementada haciendo uso eficiente de los recursos de computación.
- Finalmente se deberá definir el patrón de flujo mínimo de información, esto es, quien es responsable de abastecer cierta información y para quien⁸³. Esta etapa define al conjunto de procedimientos a la información que representan un uso eficiente de los recursos. Esta es la etapa más larga e incluye el proceso de diseño.

Es improbable que un estudio de esta clase sea emprendido en un contexto de “campo fértil”; por lo tanto se requerirá que algunos entes involucrados desplieguen los procedimientos de acceso a la información ya en existencia, junto con sus interacciones.

De esta forma, viendo a la CONAGUA como un ente integrado, conformado por partes que se interrelacionan entre sí a través de una estructura que se desenvuelve en un entorno determinado, se estará en capacidad de poder detectar con la amplitud requerida tanto la problemática, como los procesos de cambio que de manera integral.

⁸² Si la estructura de la organización no es una restricción

⁸³ El propósito habrá sido aportado por la etapa 2.

La consecuencia de esta perspectiva sistémica es hacer posible ver a la CONAGUA ya no como que tiene un fin predeterminado, sino que puede tener diversos fines en función de la forma cómo los involucrados en su destino la vean, surgiendo así la variedad interpretativa. Estas visiones estarán condicionadas por los intereses y valores que posean dichos involucrados, existiendo solamente un interés común.

1.6.2. Medición

Medir adecuadamente es la herramienta de gerencia que permite administrar objetivamente basándonos en datos cuantificables y verificables. El conocimiento profundo de un proceso, parte de admitir, conocer su variabilidad y sus causas, los procesos son imposibles de conocerse sin medición. Este proceso permitirá:

1. Conocer hacia dónde va un proceso, es decir, conocer sus tendencias y como varia.
2. Crea una cultura de estudiar continuamente los procesos y de medir permanente; este estudio y medición nos sirve para detectar las potencialidades y debilidades de un proceso.
3. Nos abre el camino para planificar y tomar mejores decisiones; gerenciar y controlar mejor la empresa.
4. Aumentar sus funciones laborales de los empleados y ocupan todo su tiempo
5. Nos permite evitar las crisis los errores y corregirlos si se presentan.
6. Le permite a la gerencia buscar culpables de los errores y fracasos.

La medición no es sólo recoger datos sino insertarlos adecuadamente en el proceso de toma de decisiones y en el mejoramiento de la empresa. La gerencia debe contar con teorías que den herramientas para recopilar los datos y para ayudar a interpretarlos. Éstos deben apoyar un estilo de dirección definido y una teoría administrativa adoptada por la organización. Además la recolección de datos debe ser una labor continua y permanente; analizando los datos para que se conviertan en información para el mejoramiento.

Indicadores

- Tipo y características de conexiones existentes.

- Acceso a medios de comunicación entre OC y Unidad Central de CONAGUA.
- Relaciones entre los elementos en estudio, bases de datos, incluyendo la evaluación de los conflictos potenciales y reales de intereses, entre personas, así como las expectativas de cada uno de ellos.
- Medición de escenarios de operación. Determinación mínima de recursos que utilizan las máquinas físicas realmente (RAM, microprocesador, disco duro y tarjeta de red). Para medir este consumo se utilizará una herramienta de medición de performance de Microsoft llamada Microsoft Assessment and Planning Toolkit (MAP 5.5)⁸⁴ así como el Monitor de Rendimiento. Para determinar el desempeño del equipo virtual se utilizará el programa VbenchMark 1.0.1.VMX.
- Índices
 - Unidad de análisis: Subdirección General Técnica (SGT) y Unidades en OC vinculadas con funciones de la SGT.
 - Existencia de datos.
 - Situaciones aleatorias.

1.6.3. Justificación técnica.

La elección de un paradigma para la ingeniería del software, en el desarrollo de los sistemas que se implementan en la CONAGUA, se lleva a cabo de acuerdo con la naturaleza del proyecto y de la aplicación, los métodos, controles, así como las herramientas a usar, los periodos de entrega requeridos, etc.

El modelo seleccionado para el desarrollo del procedimiento de creación de las geobases de datos distribuidas sobre servidores virtualizados es el modelo en espiral, debido a que actualmente es el enfoque más viable, dada la naturaleza propia del sistema de información el cual es a escala nacional. Este ente utiliza un enfoque evolutivo que permite tanto al desarrollador como al usuario, entender y reaccionar a los riesgos en cada nivel evolutivo. Este modelo, utiliza la creación de prototipos como un mecanismo de reducción de riesgo pero, lo que es más importante, permite al desarrollador aplicar el enfoque de creación de prototipos en cualquier etapa de la evolución del producto. Mantiene el enfoque sistémico que corresponde a los pasos sugeridos por el ciclo de vida clásico, pero incorporándolo dentro de un modelo de trabajo interactivo que

⁸⁴ <http://technet.microsoft.com/en-us/library/bb977556.aspx>

refleja de forma más realista la evolución de datos manejados en la CONAGUA. El modelo en espiral demanda una consideración directa de riesgos técnicos en todas sus etapas del proyecto y, si se aplica adecuadamente, debe reducir los riesgos antes de que se conviertan en problemas.

En la Figura 16 se muestra el método de diseño propuesto en donde con cada iteración alrededor de la espiral (comenzando del centro y siguiendo hacia el exterior), se construyen sucesivas versiones del proceso de interoperabilidad, cada vez más completas. Durante la primera vuelta alrededor de la espiral se definen los objetivos, las alternativas, las restricciones y se analizan e identifican los riesgos. Si el análisis de riesgo indica que hay una incertidumbre en los requisitos, se puede usar la creación de prototipos en el cuadrante de planificación de fases siguientes para dar asistencia tanto al encargado del desarrollo como al cliente. Se pueden usar simulaciones y otros modelos para definir más el problema y refinar los requisitos.

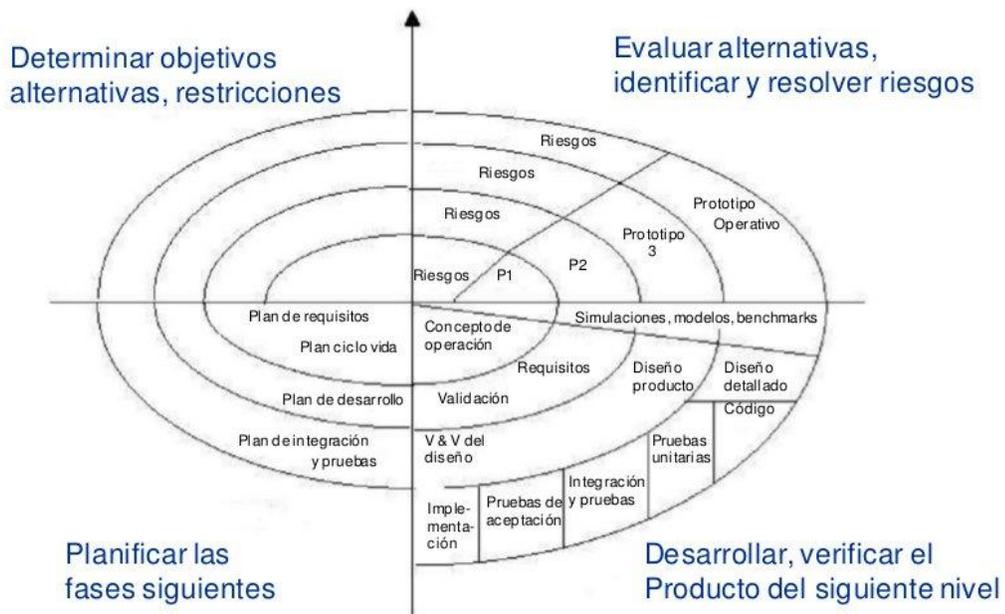


Figura 16. Modelo de diseño del sistema propuesto en espiral

Los usuarios del sistema de vinculación evalúan el trabajo de ingeniería (cuadrante de evaluación de alternativas) y sugieren modificaciones. Con base en los comentarios de los usuarios se produce la siguiente fase de planificación y de análisis de riesgo. En cada bucle alrededor de la espiral, la culminación del

análisis de riesgo resulta en una decisión de seguir o no seguir. Si los riesgos son demasiado grandes, se puede dar por terminado el proyecto.

Sin embargo, en la mayoría de los casos, se sigue avanzando alrededor del camino de la espiral ese camino lleva a los desarrolladores hacia fuera, hacia un modelo más completo y, al final, al propio sistema operacional. Cada vuelta alrededor de la espiral requiere ingeniería (cuadrante inferior derecho), que se puede llevar a cabo mediante el enfoque del ciclo de vida clásico o de la creación de prototipos. Debe tenerse en cuenta que el número de actividades de desarrollo que ocurre en el cuadrante inferior derecho aumenta al alejarse del centro de la espiral.

1.6.4. Limitantes contextuales.

El resultado del proyecto estará dirigido al personal que labora en la CONAGUA y que requiere consultar información cartográfica para la realización de sus actividades. Estas actividades están directamente relacionadas con la forma de operar de la SIGA, la cual está basada en una infraestructura de geobase de datos distribuida, de SIG tipo cliente-servidor, como se indica en la Figura 17. Muchos de los factores implicados en la consecución de la implementación de una geobase de datos distribuida sobre un ambiente virtualizado a menudo se enfocan tan sólo a subconjuntos de los verdaderos requisitos totales. En este trabajo, se contempla desde un contexto amplio, incluyendo:

- Virtualización entre proveedores de tecnología
- Estándares de virtualización y geobases de datos distribuidas.

La solución informática a desarrollar cubrirá únicamente las tareas de consulta y visualización de cartografía digital, mediante el uso de una interfaz orientada a un usuario final con pocos conocimientos en el manejo de este tipo de información

- Identificar, caracterizar y ser capaz de controlar los procesos críticos de negocios y sus componentes. Conocer la interrelación que tienen sus procesos con los de otros servicios.
- Incorporar tecnología efectivamente en sus procesos.
- Facilitar los procedimientos administrativos asociados a los trámites de modo de no solicitar a las personas la información que ya se posee en el Estado.
- Transparentar ante el ciudadano, la información que el Estado posee,

considerando el cumplimiento de los requisitos legales en materia de privacidad.

La Geobase de Datos Institucional del Agua (GeoAgua) se actualiza a partir de información obtenida de las áreas generadoras de CONAGUA en los tiempos que cada una considera adecuados.

Así, los tiempos de actualización varían. Por poner un ejemplo, la Subgerencia de Evaluación y Ordenamiento de Acuíferos, genera cambios en la cobertura de acuíferos cada que se publica la información respectiva en el Diario Oficial de la Federación (2006, 2009, julio 2010 y enero 2011). Así, se tienen periodos de actualización muy diversos. Estos valores están asentados en los datos presentados en los metadatos de cada una de las capa.



Figura 17. Forma de operación de la SIGA

Capítulo 2. ANTECEDENTES, FUNDAMENTOS LEGALES Y ETAPAS DEL PROYECTO

Resumen del capítulo

El objetivo de este capítulo es proveer al lector con una referencia de los fundamentos legales y los antecedentes sobre los cuales se sustenta la creación del Grupo, interdisciplinario⁸⁵, Temático del Agua y la forma en como en las sesiones de trabajo de éste grupo surge, en forma natural, la necesidad de contar con un vínculo de datos entre dependencias gubernamentales a fin de permitir la consulta y facilitar el acceso a los bancos de datos de información geográfica y metadatos con los que cuentan las áreas de información geográfica en la Secretaría del Medio Ambiente y Recursos Naturales (SEMARNAT) y la CONAGUA. También se señalan los principales objetivos perseguidos por este conjunto de Instituciones Gubernamentales al integrarse en un grupo interdisciplinario haciendo énfasis en la creación de un vínculo basado en la interoperabilidad de sistemas y bancos de información geográfica.

Un objetivo adicional, de éste capítulo, es mostrar un estudio interdisciplinario del problema de interoperabilidad, antes referido, así como el planteamiento de métodos de trabajo a utilizar.

Finalmente se presentan algunas recomendaciones obtenidas como resultado de este trabajo en conjunto entre dependencias del sector medio ambiente.

2.1. Antecedentes para la creación del Comité Técnico Estadístico y de Información Geográfica del Sector Medio Ambiente y Recursos Naturales.

La información estadística y geográfica oficial de México es el resultado de las acciones de generación, integración y difusión de datos que diversas instituciones de la administración pública del país realizan cotidianamente en esas materias.

⁸⁵ Tipo de estudio que requiere un sistema complejo., a través del estudio a través de una metodología adecuada, es decir que sirve como instrumento de análisis de los procesos que tienen lugar en el sistema complejo, que explican su evolución como totalidad organizada. Este estudio integrado es el resultado de un equipo con marcos epistemológicos, conceptuales y metodológicos compartidos. El concepto de sistemas complejos se relaciona directamente con el planteo interdisciplinario que se necesita para abordar los problemas que derivan de su funcionamiento como sistema, incorporando la necesidad de articular las disciplinas involucradas con la problemática o problema y respetando la especificidad del dominio de cada disciplina (c.f. ref. 2).

Hasta ahora esas acciones se habían llevado a cabo sin una coordinación que permita evitar duplicidades, que impida producir información que no tiene demanda y que atienda eficientemente las solicitudes de los diversos sectores de la sociedad. Estas acciones, así como las formas de vinculación para el trabajo entre las instituciones, están determinadas en la Ley de Información Estadística y Geográfica (LIEG).

La LIEG establece que todo el trabajo estadístico y geográfico de carácter oficial, así como las modalidades de vinculación, se organice de acuerdo al conjunto de disposiciones ahí contenidas, así como en su Reglamento y en otros ordenamientos legales y administrativos en los tres órdenes de gobierno y en los Poderes de la Unión.

En este contexto, surgen los Sistemas Nacionales Estadístico y de Información Geográfica (SNEIG).

La formación de estos Comités obedeció al mandato establecido en la Ley de Información Estadística y Geográfica⁸⁶ y de su Reglamento, constituyéndolos como un órgano colegiado de participación permanente con competencias tanto para elaborar y vigilar la ejecución del programa sectorial⁸⁷, como para ser el conducto mediante el cual se transmita y vigile el cumplimiento de las normas y disposiciones de carácter general que se expidan por el INEGI o el sector, para la captación, procesamiento y presentación de la información que se produzca en estas materias. El Comité Técnico Estadístico y de Información Geográfica del Sector Medio Ambiente y Recursos Naturales (CTEIGSMARN) es presidido por el titular de la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT) siendo sus Secretarios Técnicos de Normas y Pro Secretario Ejecutivo los titulares del INEGI y de la Subsecretaría de Planeación y Política Ambiental de la SEMARNAT respectivamente. Los vocales del Comité son los titulares de la Subsecretaría de Gestión para la Protección Ambiental de SEMARNAT, la Subsecretaría de Fomento y Normatividad Ambiental de SEMARNAT, la Dirección General de la Comisión Nacional del Agua, la Presidencia del Instituto Nacional de Ecología, la Procuraduría Federal de Protección al Ambiente, la Presidencia de la Comisión Nacional de Áreas Naturales Protegidas, la Dirección General del Instituto Mexicano de Tecnología del Agua, la Dirección General de la Comisión Nacional Forestal y la Secretaría

⁸⁶ Art. 13 y 14 de la Ley de Información Estadística y Geográfica, publicada en el DOF el 30 de noviembre de 1980.

⁸⁷ http://www2.inegi.gob.mx/sneig/contenidos/espanol/pronadeig/resumen_semarnat.aspx?c=2480

Ejecutiva de la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad.

A fin de cumplir con los ordenamientos legales establecidos en la LIEG y establecer redes con productores de información sobre el agua a nivel nacional, la CONAGUA se incorporó al CTEIGSMARN donde participan la SEMARNAT y el INEGI [ref. 6]. Así, la CONAGUA asume su corresponsabilidad para contribuir en el desarrollo de los objetivos de corto, mediano y largo plazos encaminados al fortalecimiento y consolidación de los Servicios y los Sistemas Nacionales de Estadística y de Información Geográfica; al mismo tiempo que se atiende, con una visión Institucional y de largo plazo, a los requerimientos de confiabilidad, pertinencia, accesibilidad y oportunidad que demandan los usuarios de la información de este sector para apoyar sus procesos de planeación y de toma de decisiones.

Es importante señalar que mediante estos Comités Técnicos Sectoriales el INEGI establece una red de socios para el intercambio de información estadística y geográfica en el ámbito de cada sector con la elaboración de los expertos en cada tema que compete al Sector. El Comité está configurado conforme a lo presentado en la Figura 18.



Figura 18. Estructura organizacional del Comité Técnico Estadístico y de Información Geográfica del Sector Medio Ambiente y Recursos Naturales.

En el seno de este comité se instaló el Grupo Temático del Agua, el cual establece, transmite y vigila el cumplimiento de normas y disposiciones de

carácter general establecidas por el INEGI, para la captación, procesamiento, análisis y difusión de información del sector hídrico.

2.2. Fundamentos legales, modificaciones y reformas a la Ley de Aguas Nacionales (LAN).

Las modificaciones a la LAN incorporan nuevos conceptos a tomar en cuenta en la programación y política hídrica:

- Política hídrica nacional y regional.
- Programas y subprogramas hídricos específicos, regionales, de cuenca, acuíferos y sectoriales con un enfoque de gestión integrada de los recursos hídricos (GIRH).
- Sistema Nacional y Regional de Información.
- Sistema Financiero del Agua.
- Planeación participativa entre tres niveles de gobierno y organizaciones de la sociedad civil.
- Evaluación socioeconómica de proyectos.
- Programas multianuales de inversión.
- Valoración económica y financiera del agua.
- Publicación de las disponibilidades del agua.

También se toman en cuenta nuevos conceptos en materia de recursos naturales y medio ambiente:

(3-LIV). Definición del uso para conservación ecológica

(7-VIII). La variable ambiental considerada de interés público

(14 Bis 5-II). La Gestión Integral de Recursos Hídricos (GIRH) como base de la política hídrica nacional.

(14 Bis 5-X). GIRH sustentada en las relaciones entre recursos hídricos con el aire, suelo, flora, fauna, otros recursos naturales, biodiversidad y los ecosistemas vitales.

(15-X). Cuota natural de renovación de las aguas en la programación hídrica.

(29 BIS 5-III). Restricción de uso del agua cuando afecte el caudal mínimo ecológico

(41-III). Declaratoria de reserva de agua para garantizar flujos mínimos ecológicos y conservación o restauración de ecosistemas vitales

(86-III). Programas integrales de protección de recursos hídricos y relaciones de uso de suelo con el agua

(86BIS 1-II). Preservación de humedales mediante reserva ecológica del agua

(87). Declaratoria de clasificación de cuerpos de agua.

En el Artículo 9, de la LAN se presentan las atribuciones de la CONAGUA, en dónde destacan:

Fracción XLVI. Mejorar y difundir permanentemente en el ámbito nacional el conocimiento sobre la ocurrencia del agua, la oferta y demanda, los inventarios, suelo, usos y usuarios y la información pertinente vinculada con el agua y su gestión.

Fracción XLVII. Integrar el Sistema Nacional de Información sobre cantidad, calidad, usos y conservación del agua (SINA), con la participación de los Organismos de Cuenca, en coordinación con los gobiernos de los estados y del Distrito Federal y con los Consejos de Cuenca, en concordancia con la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental (LFTAIPG).

Artículo 15: La planificación hídrica es de carácter obligatorio para la GIRH.

Fracción X. La planificación y programación hídrica nacional y de las cuencas se sustentará en una red integrada por el SINA y los Sistemas Regionales de Información sobre el agua (SIRA).

El artículo 15 Bis 6 de la LAN establece que:

“Uno de los principales instrumentos de la política hídrica nacional es el Sistema Nacional de Información sobre cantidad, calidad, usos y conservación del agua”

Artículo 14 bis 6.

“El Sistema Nacional de Información sobre cantidad, calidad, usos y conservación del agua es un instrumento básico para la planeación hídrica nacional”

2.2.1. El Grupo Temático del Agua

La idea de la implementación de una de estas herramientas de análisis geoespacial, dentro de la CONAGUA, se hizo realidad en la SIGA⁸⁸. Dadas las atribuciones que le confiere el Reglamento Interior de CONAGUA⁸⁹, en su artículo 81 inciso V y lo señalado en el artículo 82 inciso XI, XII y XVIII; la SIGA está encargado de la administración y operación del “Sistema de Información Geográfica del Agua, y solicitar su publicación y actualización en el sitio de Internet de la Comisión”⁹⁰,. “Aportar información, con base en los estudios de que disponga y los estudios que realice, directamente o a través de terceros, para apoyar la ejecución de acciones para la prevención y atención de emergencias generadas por sequías, eventos de contaminación repentina, presencia de elementos tóxicos, fenómenos hidrometeorológicos extremos, emergencias hidroecológicas y contingencias ambientales en materia de las aguas nacionales y sus bienes públicos inherentes, y vigilar y evaluar la evolución de fenómenos hidroclimáticos severos que puedan influir en las cuencas, cauces y almacenamientos de presas, así como realizar acciones tendientes a mitigar sus efectos negativos”⁹¹ y “Instrumentar, operar y actualizar el Sistema Regional de Información sobre cantidad, calidad, usos y conservación del agua; revisar y validar la información sobre escurrimientos, sedimentos y vasos que proporcionen los sectores e instancias públicas o privadas, el Sistema de Información Hidroclimatológica y de Banco de Datos a tiempo real o histórico, la red de monitoreo sistemático y permanente de la calidad de las aguas nacionales, continentales y costeras y, en su caso, la red regional de medición de la calidad del agua, de monitoreo y de laboratorios de análisis; proporcionar información con base en dichos sistemas para la integración del Sistema Nacional de Información, y capacitar en materia de dicho sistema y su operación para mejorar el conocimiento sobre la ocurrencia de las aguas nacionales y “Realizar delimitaciones geográficas y homologación de regiones hidrológicas, cuencas, subcuencas y microcuencas en que se divida y subdivida la región hidrológico-administrativa de su circunscripción.”⁹²

Partiendo de los objetivos e iniciativas plasmadas en la Agenda del Agua 2030⁹³, de los objetivos planteados en la LAN, en particular en sus artículos 7 bis inciso

⁸⁸ Vista como subgerencia no como sistema.

⁸⁹ <http://www.conagua.gob.mx/CONAGUA07/Noticias/ReglamentoInteriorCONAGUA.pdf>

⁹⁰ Reglamento interno de CONAGUA artículo 82 inciso XI

⁹¹ Ibid. artículo 82 inciso XII

⁹² Ibid. artículo 82 inciso XVIII

⁹³ <http://www.conagua.gob.mx/CONAGUA07/Temas/AgendadelAgua2030.pdf>

II y IV, 9 inciso XLVII, 12 bis 1, bis 4 y bis 6 inciso XXIX así como apoyo a los referido en el párrafo primero del artículo 14 bis; para su cabal cumplimiento, se enuncian las siguientes iniciativas las cuales sustentan la elaboración del presente trabajo:

En la Agenda del Agua 2030:

Para el logro de cuencas y acuíferos en equilibrio:

- 1) Robustecer las funciones de gobierno de CONAGUA y su organización, en particular las asociadas a los Organismos de Cuenca.
- 2) Iniciativas de carácter general:
 - a) Aplicar la evaluación orientada a resultados a todos los programas públicos que incidan o afecten el logro de la sustentabilidad hídrica.
 - b) Desarrollar sistemas regionales de información para reforzar la gestión del agua por cuenca y acuífero.

En la Ley de Aguas Nacionales (LAN):

- 1) Se declaran de interés público:
 - a) La descentralización y mejoramiento de la gestión de los recursos hídricos por cuenca hidrológica, a través de Organismos de Cuenca de índole gubernamental.
 - b) El mejoramiento permanente del conocimiento sobre la ocurrencia del agua en el ciclo hidrológico, en su explotación, uso o aprovechamiento y en su conservación en el territorio nacional, y en los conceptos y parámetros fundamentales para alcanzar la gestión integrada de los recursos hídricos, así como la realización periódica de inventarios de usos y usuarios, cuerpos de agua, infraestructura hidráulica y equipamiento diverso necesario para la gestión integrada de los recursos hídricos.
- 2) Atribución de la CONAGUA, en su Nivel Nacional:
 - a) Integrar el Sistema Nacional de Información sobre cantidad, calidad, usos y conservación del agua, con la participación de los Organismos de Cuenca, en coordinación con los gobiernos de los estados, el Distrito Federal y con los Consejos de Cuenca en concordancia con la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental.

En particular la iniciativa 4.c) implica contar con una infraestructura de TI⁹⁴ la cual permita el manejo integrado del agua, desde el ámbito de los OC hacia unidad central de CONAGUA como a la inversa. Para ello es necesario el trabajo coordinado, en equipo, entre éstos y las gerencias que se ubican en oficinas centrales para que, con base en un análisis detallado de la información que poseen, se implementen líneas de acción así como estrategias enfocadas a contribuir al bienestar social y al desarrollo sustentable del recurso hídrico sin descuidar la preservación del medio ambiente. Para el efecto es necesario el intercambio de información entre los OC y oficinas centrales de una forma automática, expedita, segura, confiable y de calidad. Para avanzar en este sentido es que surge la necesidad de generar una red que vincule a los principales bancos de información, tanto alfanuméricos⁹⁵ como geográficos⁹⁶, de cada una de las áreas que conforman a la CONAGUA.

Esta implementación permitirá distribuir la información referente al recurso hídrico a través de medios de TI para presentar un conjunto único de datos relativos a la cantidad, calidad, usos y conservación del agua dentro del ámbito nacional. Esto facilitará el fortalecimiento de los vínculos de trabajo entre los OC y oficinas centrales de la CONAGUA; propiciando la confianza de la sociedad en esta última al contar con la seguridad de que la información consultada es única, oficial y confiable; trascendiendo límites políticos y geográficos permitiendo con ello cumplir con compromisos de desarrollo, mejorando la cooperación, colaboración y el compromiso de la gestión integrada de agua.

b) Programa de Actividades

El programa de actividades asume la conformación del equipo interdisciplinario y la definición en una primera instancia de los marcos epistémicos, conceptuales y metodológicos a compartir por todas las disciplinas involucradas. Como lo expresa García, en un estudio interdisciplinario, *“el objetivo es llegar a una interpretación sistémica de la problemática original que presenta el objeto de estudio. A partir de allí, será posible lograr un diagnóstico integrado, que provea las bases para proponer acciones concretas y políticas generales alternativas que permitan influir sobre la evolución del sistema”*.

⁹⁴ Tecnologías de Información.

⁹⁵ Datos en forma de texto o números organizados en forma tabular.

⁹⁶ Gráficos de un mapa en formato digital.

Las etapas propuestas a desarrollar en el estudio interdisciplinario serían las siguientes:

ETAPA I: ANÁLISIS-DIAGNÓSTICO

Definición del objeto de estudio.

1. Reconocimiento del problema. Formulación de preguntas con base en las necesidades identificadas en conjunto con personal de OC.
2. Análisis de estudios anteriores realizados sobre aspectos diversos del problema en cuestión.
3. Identificación de elementos y relaciones que caractericen el sistema donde se involucra lo abordado en 1 y 2, con sus condiciones de contorno. Reuniones de trabajo en donde se profundice sobre la caracterización del problema.
4. Identificación de la relación del problema en cuestión a investigar con cada subsistema, para verificar o refutar hipótesis sobre sus funciones dentro del sistema. Relación del problema con otros problemas del grupo de trabajo en su conjunto, definiendo la escala y el alcance de la red de problemas involucrados.
5. Investigaciones disciplinarias del problema, en el contexto de cada disciplina y de los subsistemas establecidos.
6. Redefinición del sistema en función de la integración de los resultados obtenidos en 4 y reformulación de las preguntas iniciales.
7. Repetición de las etapas 4 y 5 en relación con la nueva definición del sistema.

Segunda integración de resultados y nueva definición del sistema.

Repetición sucesiva de las etapas 6 y 7 tantas veces como sea necesario hasta la explicación coherente del sistema en función de las respuestas a cuestionamientos surgidos en el proceso.

ETAPA II: PROPUESTAS ALTERNATIVAS

Los estudios de propuestas alternativas implican prever nuevos procesos que se pondrían en marcha cuando se introduzcan transformaciones en el sistema. El estudio específico de cada propuesta consistiría en dos instancias:

Comprensión de las modificaciones a introducir, poniendo en relieve los objetivos y evaluando los recursos que requerirá su puesta en marcha y sostenimiento. En esta etapa del proceso se utilizará la metodología de le Cruz de Malta (Montaño, 2009).

Análisis sistémico de cada propuesta: interacciones entre subsistemas según modificaciones a realizar, impactos, características de la nueva estructura que adoptaría el sistema, etc. Esta instancia involucra la evaluación de todas aquellas acciones de carácter estructural y no estructural.

En función de la extensión del presente trabajo, sólo se incorporaron brevemente algunas de las acciones a estudiar: **alternativas de utilización de nuevas tecnologías**, propuestas de uso más eficiente de la interoperabilidad, propuestas que promuevan la reorientación y desarrollo de un anillo digital que permita enlazar todas las dependencias gubernamentales a través de una red privada virtual.

2.3. Vinculación entre bases de datos geográficas.

Objetivo.

Vincular, entre sí, las geobases de datos instaladas en al menos 3 Organismos de cuenca y unidad central de la CONAGUA a fin de permitir el acceso a la información geoespacial de forma eficaz a los usuarios de cada Organismo de cuenca y servir a apoyo a varios proyectos Institucionales que requieran del uso de este tipo de geobases de datos.

Proyecto macro.

El proyecto macro consiste en desarrollar una descripción de las actividades de SIGA y su vinculación con sus áreas equivalentes en OC, esto es, desarrollar un modelo de tareas primarias. Dependiendo de la escala del estudio, puede ser necesario:

- 1) Derivar varios modelos de actividad a distintos niveles de resolución, con objeto de describir completamente a las necesidades de información.
- 2) Derivar categorías de información requeridas para soportar las actividades de los modelos y las actividades particulares de las cuales la información puede ser obtenida.
- 3) Para una estructura grupo en particular, definir las reglas de dirección en términos de la responsabilidad de toma de decisiones sobre las actividades que un área operativa existente tenga asignadas⁹⁷.
- 4) Usar estas definiciones de reglas para convertir los flujos de información de actividad a actividad en los flujos de información de “regla a regla”, esto es,

⁹⁷ Si la estructura de la organización no es una restricción

definir las necesidades de información particulares de un área operativa basándose en el análisis de las actividades de que es responsable.

- 5) Definir las necesidades de un sistema de información que solucione los requisitos de desempeño de las actividades que soportan a cada sistema, tal que una red coherente pueda ser desarrollada haciendo uso eficiente de los recursos de computación.
- 6) Finalmente se deberá definir el patrón de flujo mínimo de información, esto es, quien es responsable de abastecer cierta información y para quien⁹⁸. Esta etapa define al conjunto de procedimientos a la información que representan un uso eficiente de los recursos.

2.4. Proyecto piloto

El proyecto piloto fue implementado con el apoyo de personal de los Organismos de Cuenca Península de Yucatán, Frontera Sur y Golfo Centro. El prototipo está constituido por 1 instalación física distribuida de la Geobase de Datos, en el Organismo de Cuenca Península de Yucatán y 2 virtualizadas, una implementada y administrada por personal de Golfo Centro y otra por personal de Frontera Sur, conforme se muestra en la Figura 19. De esta forma se efectuarán operaciones de replicación y sincronización de geobases de datos. Para el efecto los recursos requeridos con los que ya cuenta la SIGA para el desarrollo de la implementación informática son los siguientes:

- 1 Servidor PowerEdge M610, virtualizado, con enclosure PowerEdge M1000e de Dell con procesadores Intel[®] Xeon[®] 5600 de cuatro núcleos opción 95 W TDP, 2 Gb en RAM como mínimo, 136 Gb disponibles en disco duro al menos.
- Unidad NAS-SAN Marca NetApp Modelo Fas 2050 con espacio de almacenamiento de 50 Tb.
- Microsoft Windows Server 2008, con Hyper-V.
- RDBMS Microsoft MS-SQL 2005 superior o equivalente.
- Licencias de ArcSDE 9.1
- Conexión MAN o WAN sobre TCP/IP
- Cartografía digital creada bajo el concepto de geoformato, es decir, que posea referencia geográfica.

⁹⁸ El propósito habrá sido aportado por la etapa 2.

Los recursos adicionales requeridos en cada Organismo de Cuenca son los siguientes, para servers virtuales:

- 1 servidor virtual con procesador de dos núcleos opción 2 Gb en RAM como mínimo, 200 Gb disponibles en disco duro al menos.
- Microsoft Windows Server 2008.
- RDBMS Microsoft MS-SQL 2005.
- Licencia de ArcSDE 9.1 y/o ArcGIS 9.1
- Conexión con NIC virtualizada WAN sobre TCP/IP
- Cartografía digital creada bajo el concepto de geoformato, es decir, que posea referencia geográfica.

Los recursos adicionales requeridos en el Organismo de Cuenca Península de Yucatán son los siguientes, para server real:

- 1 servidor físico con procesador de dos núcleos opción 2 Gb en RAM como mínimo, 200 Gb disponibles en disco duro al menos.
- Microsoft Windows Server 2008.
- RDBMS Microsoft MS-SQL 2005.
- Licencia de ArcSDE 9.1 y/o ArcGIS 9.1
- Conexión WAN sobre TCP/IP
- Cartografía digital creada bajo el concepto de geoformato, es decir, que posea referencia geográfica.

Descripción de los tres prototipos de ambiente virtualizado con los que se trabajará.

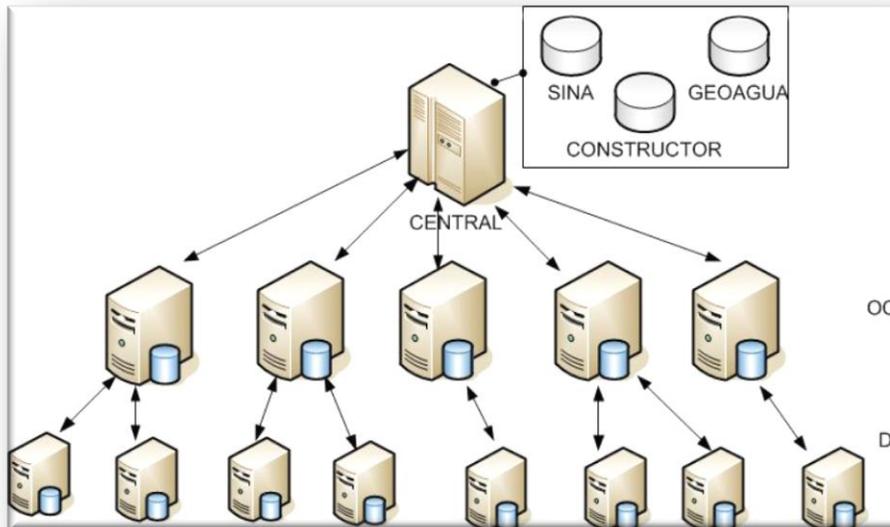


Figura 19. Diagrama de configuración global de la Geobase de Datos Distribuida

Primer prototipo.

El primer prototipo consiste en la virtualización de servidores blade, mostrados en la Figura 20. Se contará, por lo menos, con 1 servidor de este tipo para permitir la implementación de 2 servidores virtuales en este servidor físico. 2 servidores, de los respectivos Organismos de Cuenca estarán virtualizados sobre VMWare en los respectivos equipos desktop. El tercer OC, Península de Yucatán, contará con un servidor físico, debido a que en estos OC ya se cuenta tanto con personal capacitado como con los elementos de TI necesarios. Cabe mencionar que además de la virtualización de servidores también se utilizará la virtualización de aplicaciones, conforme se muestra en la Figura 21; estas serán entregadas a los clientes bajo demanda, reduciendo los problemas de compatibilidad existentes entre diversas plataformas de SO. Cada uno de los servidores tendrá su propio RDBMS, operando sobre MSSQL 2008, que administrará la Geobase de Datos respectiva sobre plataforma ArcSDE empresarial. Estas geobases de datos estarán relacionadas al servidor central

mediante vinculación de servidores y operación de servidores remotos⁹⁹

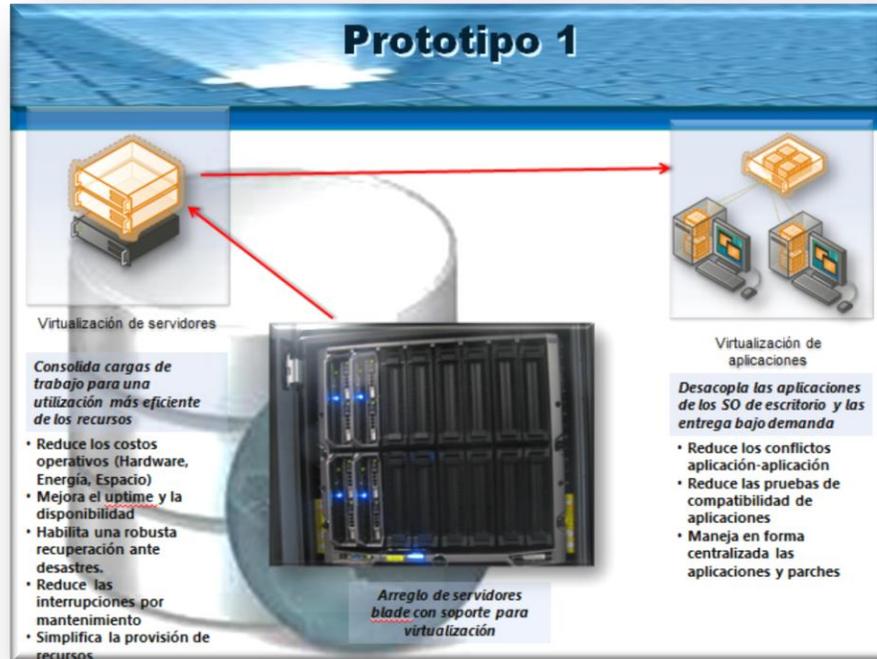


Figura 20. Hardware para el primer prototipo de operación

99 Para mayor referencia ver la documentación respectiva de MSSQL. [http://msdn.microsoft.com/es-es/library/aa560998\(v=bts.10\).aspx](http://msdn.microsoft.com/es-es/library/aa560998(v=bts.10).aspx), [http://msdn.microsoft.com/es-es/library/aa560998\(v=bts.70\).aspx](http://msdn.microsoft.com/es-es/library/aa560998(v=bts.70).aspx), <http://msdn.microsoft.com/en-us/library/ms190187.aspx>, http://www.guillesql.es/Articulos/SQLServerFAQ_Servidor_Vinculado_OPENQUERY_Notacion_4_Part.espx

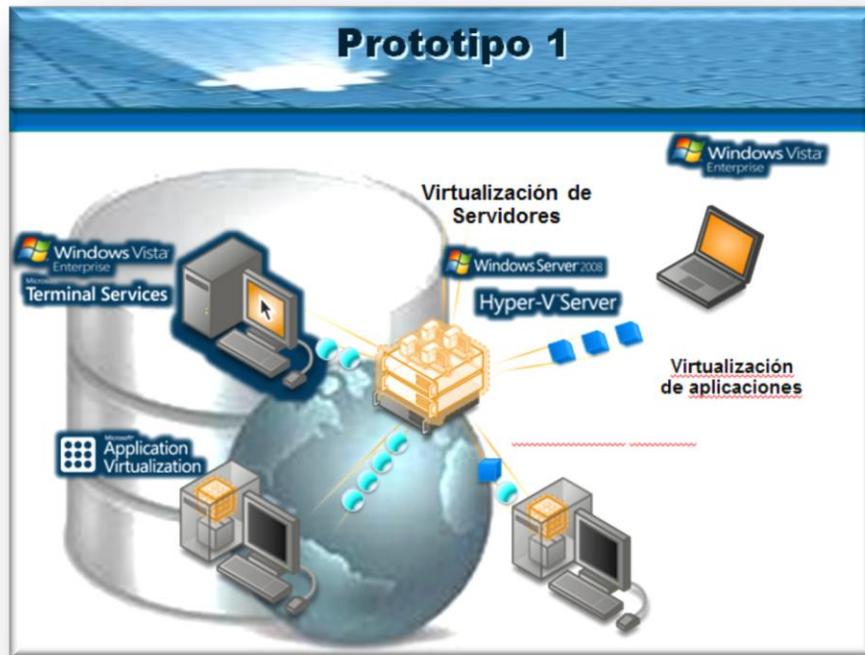


Figura 21. Software para el primer prototipo

Segundo prototipo.

El segundo prototipo consiste en la utilización de un esquema de operación virtualizado igual al primero, solo que en este caso se añadirá un soporte para almacenamiento de información sobre una unidad NAS-SAN FAS2550 de NetApp con soporte para LUNS¹⁰⁰ desde y hacia servidores MSSQL, según se muestra en la Figura 22. En cuanto al software, en la capa de virtualización de aplicaciones, se sustentará a partir de una infraestructura de hardware soportada por servidores virtuales instalados sobre plataforma Microsoft Windows 2008 R2 con Hyper-V lo cual permitirá generar modelos de aplicaciones específicas que servirán de base a la implantación de la plataforma operativa del SIG, esto se detalla en la Figura 23.

¹⁰⁰ <http://media.netapp.com/documents/tr-3428.pdf>. Pág. 12

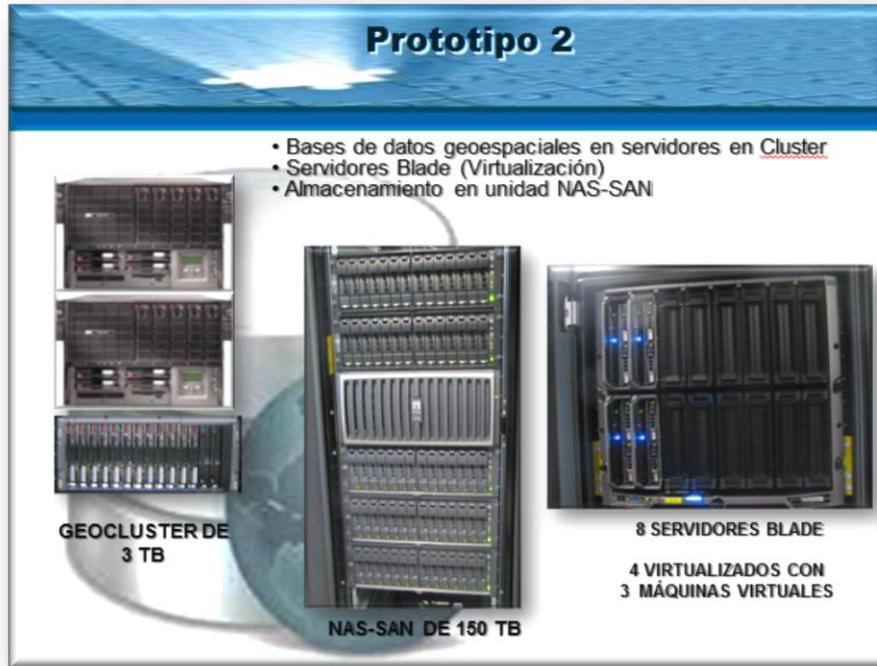


Figura 22. Hardware para el segundo prototipo

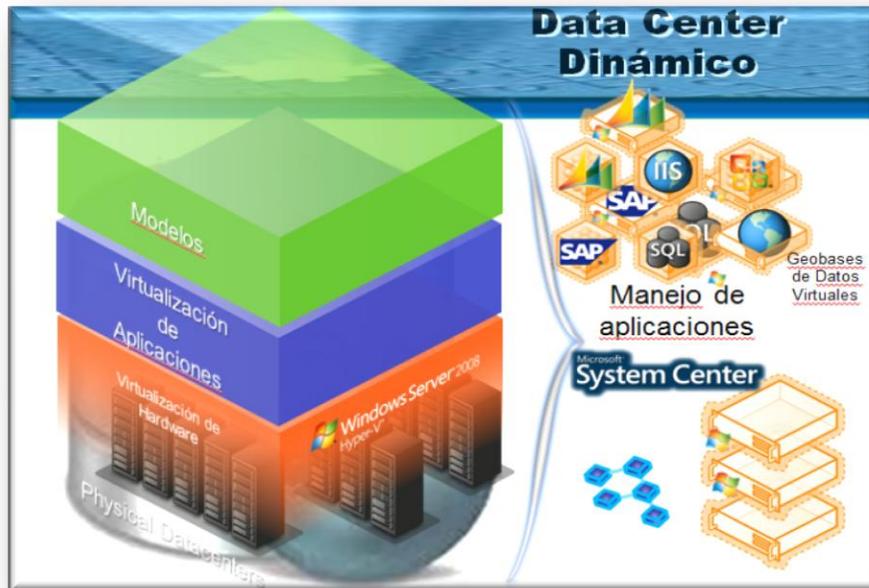


Figura 23. Diagrama de software para el segundo prototipo

Tercer prototipo.

El tercer prototipo consiste en un arreglo complejo que constará de una granja de servidores en sitio o, de preferencia, en cómputo de nube enfocado a un modelo SOA¹⁰¹ lo cual redundaría en ahorros tanto en hardware, energía eléctrica así como en los requerimientos de personal, según se indica en la Figura 24. En la Figura 25 se muestra la infraestructura de software la cual tiene un nuevo enfoque fundamentalmente sustentado con una nueva infraestructura de elementos tales como Database-as-a-Service, Informatio-as-a-Service, Process-a-a-Service, Application-as-a-Service, Platform-as-a-Service, Integration-as-a-Service, Security-as-a-Service, Management/Guvernance-as-a-Service, Testing-as-a-Service e Infraestructure-as-a-Service, .



Figura 24. Hardware para el tercer prototipo

¹⁰¹ c.f. Glosario de términos

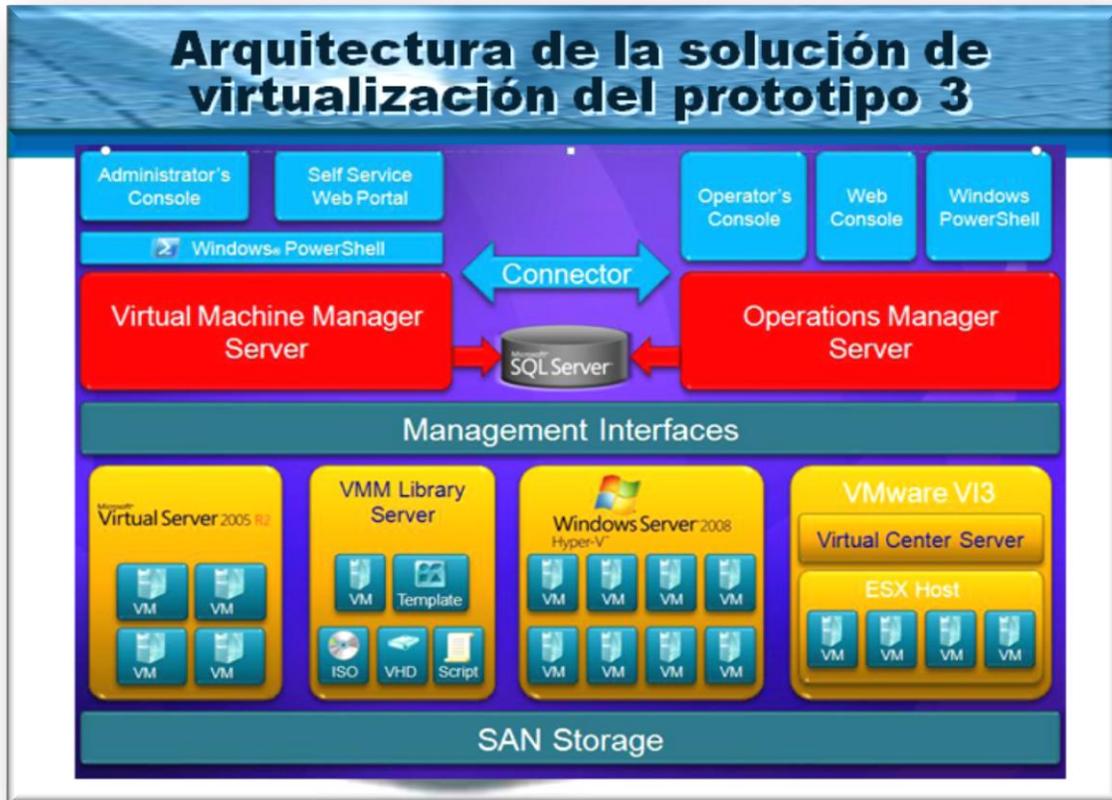


Figura 25. Diagrama de software para el tercer prototipo

Capítulo 3. IMPLEMENTACIÓN DE LA GEOBASE DE DATOS INSTITUCIONAL DEL AGUA EN UN AMBIENTE DISTRIBUIDO SOBRE SERVIDORES VIRTUALIZADOS.

Resumen.

La finalidad ésta sección es presentar al lector la metodología utilizada para la implementación de la Geobase de Datos Institucional del Agua (GeoAgua) en un ambiente distribuido sobre servidores virtualizados. Lo anterior pretende sentar las bases para una futura integración de todas las geobases de datos de cada uno de los Organismos de Cuenca en un ambiente distribuido así como las bases sobre las cuales se sustenta esta solución tales como la fragmentación, plataforma de virtualización, funcionalidad, particionamiento, replicación y asignación.

3.1. Método a implementar.

El método¹⁰² en el que se basa este trabajo está sustentado en el diseño de bases de datos distribuidas de abajo hacia arriba (bottom-up). En la primera parte se considerarán los dos procesos fundamentales: la fragmentación y la asignación; los cuales pueden ser abordados en forma simultánea. En una segunda parte se realizará, primeramente, la partición para luego asignar los fragmentos generados.

Pasos a seguir.

1. Recopilación de requerimientos. Determinación de requisitos para obtener tanto datos como las necesidades de procesamiento de los usuarios, los requisitos del sistema, los objetivos del rendimiento, seguridad, disponibilidad y flexibilidad
2. Análisis de requerimientos.
3. Diseño de vistas. En esta etapa se definen las interfaces del usuario así como las aplicaciones que usarán las bases de datos así como datos estadísticos o estimaciones de las mismas sobre frecuencia de acceso de cada aplicación a cada tabla, que permita optimizar el diseño conceptual.
4. Diseño y modelo conceptual. Integración de las vistas del usuario. En esta

¹⁰² Método es el procedimiento para alcanzar los objetivos o medio utilizado para llegar a un fin y la metodología es el estudio del método.

etapa se obtienen los esquemas conceptuales globales, información de acceso y esquemas externos.

5. Diseño de la distribución
 - a. Fragmentación
 - b. Asignación.
6. Diagramas de enlaces.
7. Implementación de la i-ésima propuesta de solución.
8. Diseño físico.
9. Mediciones de desempeño de la distribución, funcionalidad de la alternativa de solución y ajustes.

3.2. Implementación del método.

3.2.1. Recopilación de requerimientos.

Para implementar la Geobase de Datos Distribuida se requiere saber si los enlaces digitales con los que se cuentan en la unidad central de CONAGUA con los Organismos de Cuenca cuentan con velocidades de transmisión adecuados esto facilitará el proceso de sincronización de las geobases de datos. Así mismo, se debe determinar la seguridad con la que cuenta cada servidor a fin de garantizar el acceso protegido hacia los bancos de datos.

La plataforma de TI relativa a SIG con la que opera el SIGA es ArcGIS según lo expuesto en el capítulo 1. Así mismo, la plataforma base para la gestión de SIG con la que cuentan los 13 Organismos de Cuenca también es ArcGIS.

Los enlaces digitales con los que cuenta la CONAGUA están sustentados por una red de cableado estructurado nivel 6, una sala de telecomunicaciones así como enlaces a PABX y tramas digitales. También cuenta con enlaces de tecnología de acceso en la última milla de tipo alámbrico, satelital o microondas. Las subredes empleadas tienen capacidad de 254 host cada una, operando en la red de área local. Los tipos y capacidad del enlace a Internet cuentan con tecnología de acceso ADSL, satelital, celular y dial-up; permitiendo acceder a la intranet a través de la red de telecomunicaciones vía VPDN en la localidad. Todo lo anterior está soportado con switches de red que se encuentran instalados en los cuartos de comunicaciones, MDF¹⁰³ o IDF's¹⁰⁴. Estos dispositivos cuentan

¹⁰³ Instalación principal de distribución principal. Recinto de comunicación primaria de un edificio. El Punto central de una topología de networking en estrella donde están ubicados los paneles

con puertos de fibra óptica que permiten dar servicio a los enlaces de fibra, considerando dos puertos por cada par de fibras sin incluir a los pares de redundancia. Adicionalmente la COANGUA cuenta en oficinas centrales con puntos de acceso inalámbrico (access point), empleados para brindar conectividad LAN inalámbrica a usuarios móviles en salas de juntas o de capacitación.

Los enlace de telecomunicaciones (ancho de banda) con el que cuenta la CONAGUA tienen una capacidad del enlace dedicado de 29297 Mbps¹⁰⁵. Esto indica una alta velocidad según se muestra en la Figura 26 en donde se puede observar una prueba de velocidad ofrecida por la compañía McAfee¹⁰⁶.

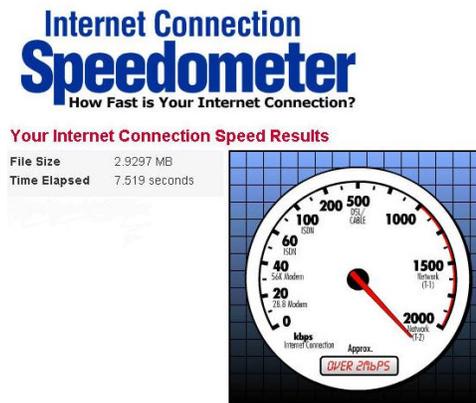


Figura 26. Velocidad de conexión a Internet CONAGUA

En cuanto a la seguridad, los equipos que integrarán la Geobase de Datos distribuida operarán sobre MSSQL 2005¹⁰⁷ que cuenta con alta seguridad. Por otro lado, en Unidad Central de CONAGUA se cuenta con firewall por hardware y servidores Proxy. Para garantizar la seguridad de las aplicaciones los servidores de publicación web se encuentran dentro de una zona de acceso desmilitarizada (DMZ¹⁰⁸).

de conexión, el hub y el router.

¹⁰⁴ Instalación de distribución intermedia. Recinto de comunicación secundaria para un edificio que usa una topología de red en estrella. El IDF depende del MDF.

¹⁰⁵ Ancho de banda del enlace expresado en megabits por segundo (Mbps)

¹⁰⁶ http://promos.mcafee.com/speedometer/test_3000.asp

¹⁰⁷ <http://www.microsoft.com/spain/sql/technologies/security/default.msp>

¹⁰⁸ Una **zona desmilitarizada** (DMZ, demilitarized zone) o "**red perimetral**" es una red local que se ubica entre la red interna de una organización y una red externa, generalmente Internet. El objetivo de una DMZ es que las conexiones desde la red interna y la externa a la DMZ estén permitidas, mientras que las conexiones desde la DMZ sólo se permitan a la red externa -- los equipos en la DMZ no pueden conectar con la red interna. Esto permite que los equipos de la DMZ puedan dar servicios a la red externa a la vez que protegen la red interna en el caso de que intrusos comprometan la seguridad de los equipos situados en la zona desmilitarizada.

La forma de descomponer las relaciones de la base de datos en pequeños fragmentos, tiene varias razones. En principio, el problema de la fragmentación radica en encontrar la unidad apropiada de distribución. Una relación no es una adecuada unidad; primero, las vistas de la aplicación normalmente son subconjuntos de relaciones. Además, la localidad de los accesos de las aplicaciones no está definida sobre relaciones enteras pero sí sobre subconjuntos de las mismas. Por ello, sería adecuado considerar como unidad de distribución a estos subconjuntos de relaciones si las aplicaciones tienen vistas definidas sobre una determinada relación que se encuentra en varios sitios de la red. En este punto se tienen dos opciones. Primero, la relación no estará replicada y se almacenará en un sitio único, o bien, existe réplica en todos o algunos de los sitios en los cuales reside la aplicación. Las consecuencias de esta estrategia es la generación de un conjunto de accesos remotos innecesario. Además, se pueden realizar réplicas innecesarias que ocasionen problemas en la ejecución de las actualizaciones y este hecho no es deseable debido a que el espacio de almacenamiento está limitado.

La descomposición de una relación en fragmentos, tratados cada uno como unidad de distribución, hace posible el proceso concurrente de las transacciones. Además, la vinculación de estas relaciones, provocará la ejecución paralela de una consulta al dividirla en una serie de subconsultas que operarán sobre los fragmentos.

Un aspecto importante en el diseño de la distribución es la cantidad de factores que contribuyen a un diseño adecuado. La organización lógica de la base de datos, la localización de las aplicaciones, las características de acceso de las aplicaciones a la base de datos y las características del sistema en cada sitio, tienen un valor fundamental en la distribución. La información necesaria para el diseño de la distribución puede dividirse en cuatro categorías:

1. Banco de datos.
2. La aplicación.
3. La red de computadoras
4. Computadoras en sí mismas.

Las dos últimas son de carácter cuantitativo y servirán para desarrollar el proceso de asignación.

3.2.2. Análisis de requerimientos.

Existen 12 reglas que deben cumplir las bases de datos distribuidas para su adecuada operación (Date, 2004). Estas son:

1. Autonomía local – Ningún sitio debe depender de otro para desempeñar sus funciones.
2. Independencia del sitio central – Un SBDD no necesita depender más del sitio central que de cualquier otro. Todos los sitios deben ser igualmente remotos. Cada sitio debe tener su propio diccionario de datos y seguridad.
3. Operación continua – La realización de cualquier función no debe afectar la base de datos distribuida y debe operar 7 x 24 con hot backup.
4. Transparencia local – Los usuarios deben percibir como si la geobase de datos en su conjunto estuviese almacenada en su servidor local sin diferencia entre lugares específicos.
5. Fragmentación transparente – Los usuarios deben percibir como si estuviesen utilizando una sola geobase de datos centralizada.
6. Replicación transparente – Los usuarios no deben preocuparse de ninguna replicación de datos. Los snapshots deben contener todos los datos.
7. Procesamiento de consultas distribuidas – Un SBDD debe procesar consultas tan rápido como sea posible aunque los datos se encuentren distribuidos.
8. Manejo de transacciones distribuidas – Un SBDD debe manejar con eficacia las de las transacciones múltiples sitios.
9. Independencia del hardware – Un SBDD debe ser capaz de operar en diferentes tipos de hardware.
10. Independencia del sistema operativo – Un SBDD debe ser capaz de operar en diferentes sistemas operativos.
11. Independencia de la red – Un SBDD debe ser capaz de operar en diferentes tipos de redes si ser afectadas por el protocolo utilizado.
12. Independencia del RDBMS – Un SBDD debe ser heterogéneo.

La implementación de la solución de distribución de geobase de datos será de tipo replicado en un ambiente fragmentado con un enfoque de diseño de abajo hacia arriba (bottom-up).

Es necesario justificar el enfoque de diseño en relación a la razón por la cual se deben replicar los datos. La respuesta es contar con partes semejantes a la geobase de datos con la que se cuenta a nivel central en cada Organismo de

Cuenca pero acotada por el entorno espacial que confina a cada una de estas unidades regionales. Lo anterior se basa en varias premisas las cuales son:

- Desempeño y balance de cargas
- Reducción de carga de red
- Particionamiento de datos geoespaciales
- Cómputo desconectado
- Uso empresarial de la geobase de datos
- Consolidación de servidores (reorganización)
- Arquitectura de alta disponibilidad
- Recuperación ante desastres

Desempeño y balance de cargas.

Para acortar las distancias que dividen artificialmente la realidad geoespacial se necesita un marco que extienda la capacidad de procesamiento entre varios servidores distribuidos geoespacialmente. Sin embargo las necesidades de análisis de información varían de un área geográfica a otra esto incluye: la traducción de formatos de los datos a una estructura de datos uniforme pero transitoria; sistemas de coordenadas consistentes, proyecciones cartográficas y representaciones de datos dependientes de la plataforma así como la recuperación de atributos asociados y metadatos. Esto requiere de la creación de instancias separadas, pero duplicadas a fin de apoyar los flujos de trabajo tanto local como regional. Así, se consigue la creación de un gran sistema transaccional administrando una gran cantidad de procesos de edición y manejo de datos (OLTP¹⁰⁹) lo cual permite mantener un alto nivel de desempeño.

Reducción de carga de red.

Debido que los OC están geográficamente dispersos, la replicación permite la distribución de los datos de los servidores ArcSDE de ArcGIS los cuales están más cerca de los usuarios de los datos geográficos. Esta configuración reduce el tráfico que soportan los servicios de telecomunicación entre los usuarios interactivos y la Geobase de Datos respectiva, para la impresión de mapas los cuales requieren la movilización de grandes cantidades de datos del nodo a la impresora o plotter. Adicionalmente proporciona alta disponibilidad a los usuarios remotos quienes, de otra forma, tienen que depender de las conexiones WAN;

¹⁰⁹ Online Transactional Processing

es claro que este hecho no es gratuito debido a que el hecho de generación de réplicas debe ser considerado.

Particionamiento de datos geospaciales.

La Geobase de Datos es una aplicación tecnológica sustentada por datos geospaciales. La nueva generación de TI para administración de bases de datos geospaciales es todavía “dato-céntrica”; es decir, a medida que incrementa la utilización de los SIG en la toma de decisiones, la calidad y la actualidad de los datos adquieren mayor importancia. Se hace necesario incluir mecanismos para verificar la calidad de los datos que han de incluirse en los diseños de las bases de datos. Esto se logra implementando un nivel de interoperabilidad, mediante el uso de parámetros de geodistribución, que permitan un acceso eficaz a la información geoespacial, conforme se indica en la Tabla 6.

Tabla 6. Parámetros de geodistribución.

Nivel de interoperabilidad	Prerrequisito para la interoperabilidad	Estado
Institucional	Deseo de interoperar	Variado y sin especificar
Modelos de información	Formalización de los descriptores de datos	Etapas primarias de desarrollo
Esquemas de datos	Adopción de estándares de bases de datos	Varía según sector
Intercambio de datos	Herramientas e interfaces (APIs) ¹¹⁰ estándares de la industria	Disponibles y en expansión
Redes	Protocolos estándares de redes	Bien establecidos

El seguimiento de los datos y de las transformaciones asociadas a los mismos se hace más difícil en la medida que los datos se distribuyen más ampliamente y se mantienen descentralizadamente. La división o fragmentación horizontal mostrada en forma lógica en la Figura 27 y en forma física en Figura 28, como se explicó en el primer capítulo, trabaja sobre las tuplas dividiendo la relación en subrelaciones que contienen un subconjunto de las tuplas que alberga la primera. La fragmentación vertical, en cambio, se basa en los atributos de la relación para efectuar la división. Estos dos tipos de partición podrían

¹¹⁰ c.f. Glosario

considerarse los básicos. Sin embargo, existen otras alternativas. Fundamentalmente, se tiene una fragmentación mixta cuando el proceso de partición hace uso de los tipos anteriores. La fragmentación mixta puede llevarse a cabo de tres formas distintas:

1. Desarrollando primero la fragmentación vertical y, posteriormente, aplicando la partición horizontal sobre los fragmentos verticales¹¹¹.
2. Aplicando primero una división horizontal para luego, sobre los fragmentos generados, desarrollar una fragmentación vertical¹¹².
3. De forma directa considerando la semántica de las transacciones.

Cuando se va a fragmentar una base de datos se debe valorar el grado de fragmentación que va a alcanzar, ya que éste será un factor que influirá notablemente en la ejecución de las consultas. El grado de fragmentación puede variar desde una ausencia de la división, considerando a las relaciones unidades de fragmentación; o bien, dividir a un nivel cada tupla o atributo tal que forme un fragmento. Ante estos dos casos extremos, se debe buscar un caso intermedio, el cual debe establecerse sobre las características de las aplicaciones que hacen uso de la base de datos.

Cuando una serie de datos se replican, las razones para ello se encuentran en torno a la seguridad y a la eficiencia de las consultas. Si existen muchas reproducciones de un elemento, en caso de fallo en el sistema, se podría acceder a esos datos ubicados en sitios distintos. Además, las consultas que acceden a los mismos datos pueden ejecutarse en paralelo, ya que habría copias en diferentes reservorios. Por otra parte, la ejecución de consultas de actualización y de escritura, implicaría la actualización de todas las copias que existan en la red, cuyo proceso puede resultar complicado. Por lo tanto, un parámetro a valorar la cantidad de consultas de lectura que se efectuarán, así como el número de consultas de escritura que se llevarán a cabo.

¹¹¹ Denominada partición VH

¹¹² Denominada partición HV

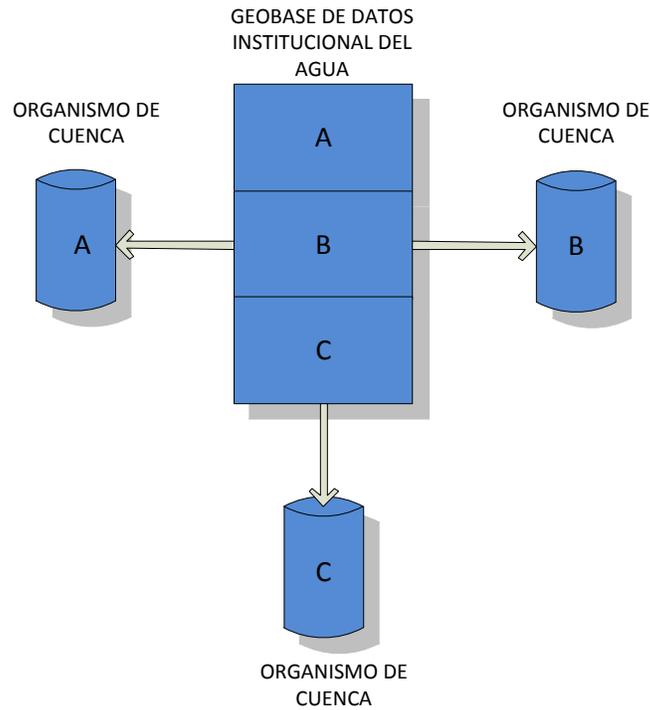


Figura 27. Geobase de datos distribuida particionada



Figura 28. Partición geoespacial de la Geobase de Datos

Cómputo desconectado.

Esta información implica al esquema conceptual global. Es importante señalar cómo las relaciones de la base de datos se conectan con otras. En la Figura 29, se presenta la forma de conexión de relaciones. Normalmente se denomina relación propietaria a la que se encuentra situada en la cola del enlace, mientras que a la ubicada en la cabecera del vínculo se refiere como relación miembro. Dicho de otra forma se puede pensar en relaciones de origen cuando se hace referencia a las propietarias y destino cuando se refiere a las miembro. Definiremos dos funciones: propietaria y miembro, las cuales proyectarán un conjunto de enlaces sobre un conjunto de relaciones. Dado un enlace, devolverán el miembro y el propietario de la relación, respectivamente. La información cuantitativa necesaria gira en torno a la cardinalidad de cada relación, notada como $\text{card}^{\text{®}}$ ¹¹³.

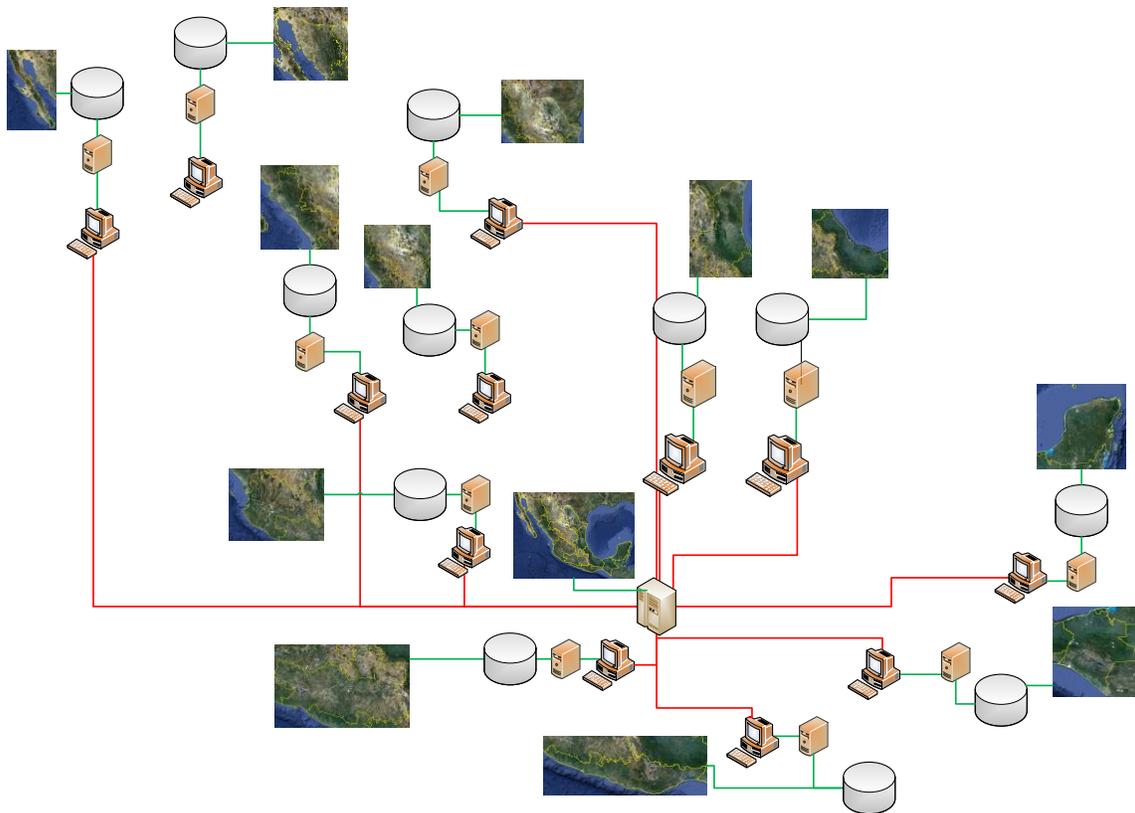


Figura 29. Diagrama de fragmentación de la información geoespacial incluida en la Geobase distribuida del Agua.

¹¹³ Definida de la forma clásica como una herramienta para comparar conjuntos finitos la cual debe mostrar que dos conjuntos finitos tienen la misma cardinalidad si existe una relación biyectiva entre sus elementos.

3.2.3. Uso empresarial de la Geobase de Datos.

La nueva forma de operación de la Geobase de Datos Institucional del Agua involucra la creación y envío de copias de datos, o los datos originales, entre dos o más geobases de datos a fin de contar en un solo reservorio con datos de alto interés geoespacial para la Institución. Lo anterior quiere decir que, además de contar con réplicas en cada uno de los Organismos de Cuenca, también se debe contar ya sea con acceso directo a otras bases de datos o, bien, con copias de información que requiera ser representada geográficamente. Para mantener estas copias en sincronía, la distribución de datos también incluye medios para aplicar ediciones hechas a cada copia a lo largo de todas éstas. En la Figura 30 se muestra la forma de interacción de las bases de datos.

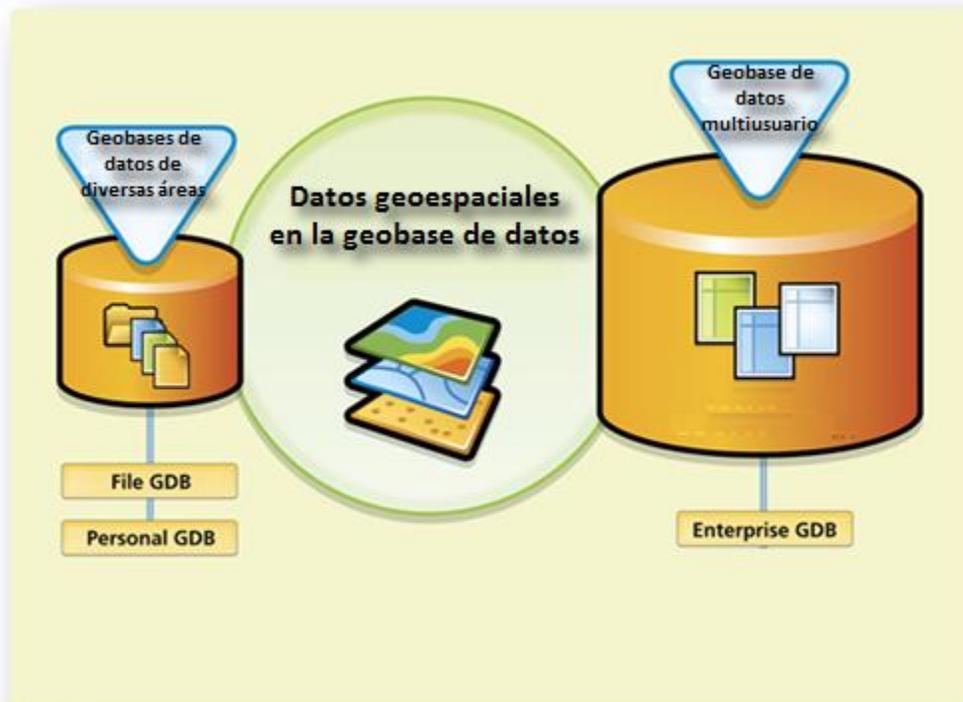


Figura 30. Integración de información geoespacial en el ámbito empresarial

La distribución de datos puede servir como medio para mejorar la disponibilidad y desempeño en el acceso a los datos. La disponibilidad de datos se refiere a la facilidad de accederlos. Una disponibilidad baja puede ser el resultado en redes lentas o poco fiables. Por ejemplo, empleados que trabajen en campo pueden tener problemas al acceder datos en la oficina dependiendo en donde se

encuentren. También es posible que ocurra este caso cuando, debido a requerimientos de seguridad o a un fallo en la infraestructura, los sistemas se desconecten. La distribución de datos permite hacer copias de fragmentos de la geobase de datos de forma que la disponibilidad es optimizada.

Las cuestiones de desempeño pueden surgir cuando varios usuarios accedan la misma geobase de datos a la vez. Por ejemplo, el sistema puede tener muchos editores así como muchos lectores consultando los datos al mismo tiempo. Mediante la distribución de datos es posible disminuir la carga de determinados segmentos de la red.

Enseguida se describen algunos casos en los cuales se aplicará una distribución de datos en la geobase de datos.

Usuarios móviles: Al efectuar captura de geocoordenadas en campo es necesario acceder la geobase de datos tanto en oficina central como en Organismos de Cuenca. Cada usuario móvil, conforme se muestra en la Figura 31, trabajará con una copia local de algunos datos provenientes de estas bases de datos. Cuando estos usuarios regresan a la oficina, envían sus cambios y reciben actualizaciones de las geobases de datos central o las ubicadas en cada OC.

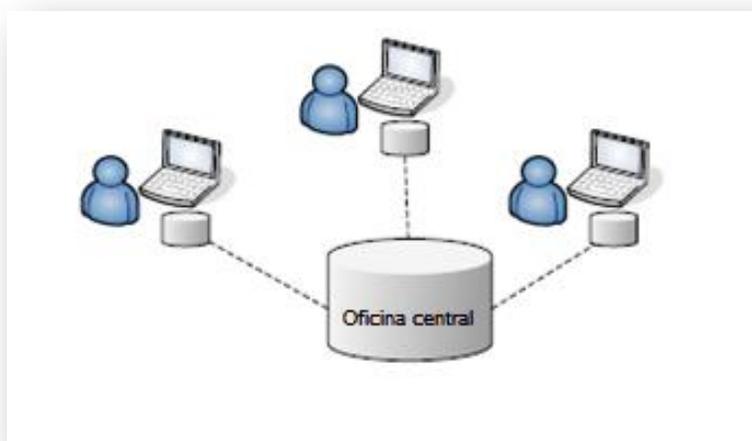


Figura 31. Acceso a segmentos de la Geobase de Datos Distribuida en forma remota

Geobase de datos multigrupo (CONAGUA Central): La estructura operativa de la CONAGUA y la cantidad de personal con el que cuenta hace necesario el manejo de datos dividido a través de diferentes grupos. Por ejemplo, la Gerencia de Aguas Superficiales e Ingeniería de Ríos (GASIR) puede encargarse del manejo de las utilerías de redes hidrográficas mientras que, por otro lado, la Gerencia de Aguas Subterráneas (GAS) está encargada de la administración del agua subterránea de la misma área geográfica. Con estas geobases de datos,

algunos de los datos podrían requerir autorización de acceso donde fuese necesario, esto se muestra en la Figura 32. La distribución de datos puede ser usada para proveer a cada grupo con acceso a la información que solamente requiere.

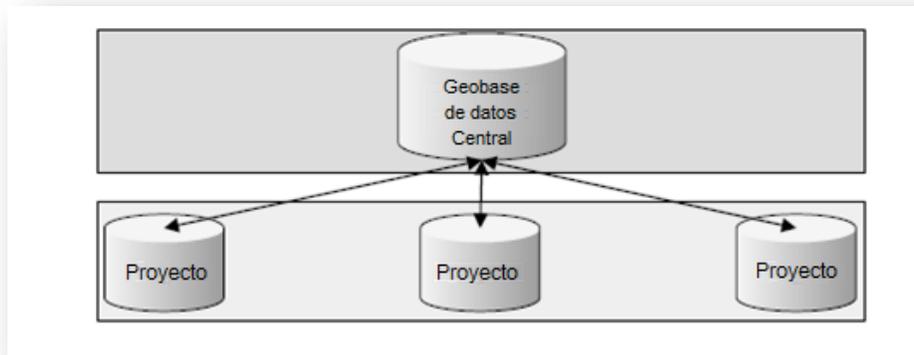


Figura 32. Diagrama de vinculación y consulta de datos en formato multigrupo

Organismos de Cuenca: La CONAGUA cuenta con varias oficinas localizadas en diversas regiones del país, cada una requiere acceso a los datos geográficos de esa región. La distribución de datos puede ser usada de forma tal que cada OC tenga su propia copia de la información para su uso cotidiano, según se indica en la Figura 33. Un proceso puede ser ejecutado periódicamente para sincronizar cambios de forma que cada OC tenga su información lo más actualizado posible y acorde con la información central.

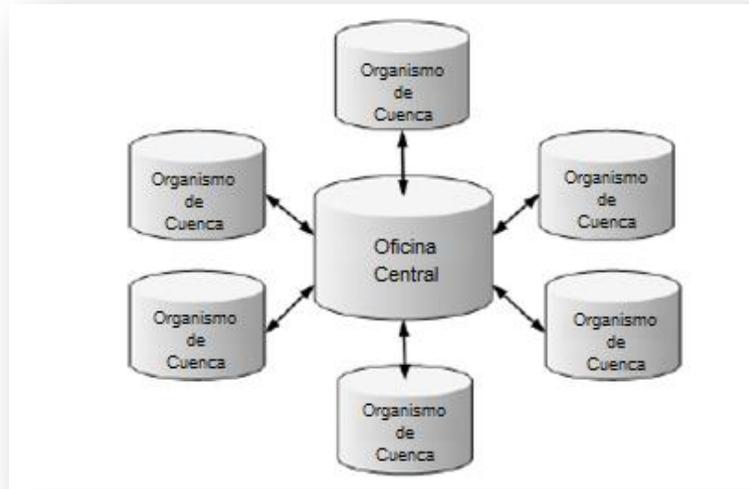


Figura 33. Distribución de la Geobase de Datos Institucional del Agua

Organización multinivel (Organismos de Cuenca-Direcciones Locales): La CONAGUA tiene una estructura organizacional de niveles múltiples. Cuenta con oficinas a nivel regional en Organismos de Cuenca, estatal en Direcciones Locales y en la ciudad de México a nivel central, como se indica en la Figura 34. La distribución de datos le permitirá a cada oficina manejar localmente sus datos y al mismo tiempo compartirlos con el nivel inmediato superior e inferior.

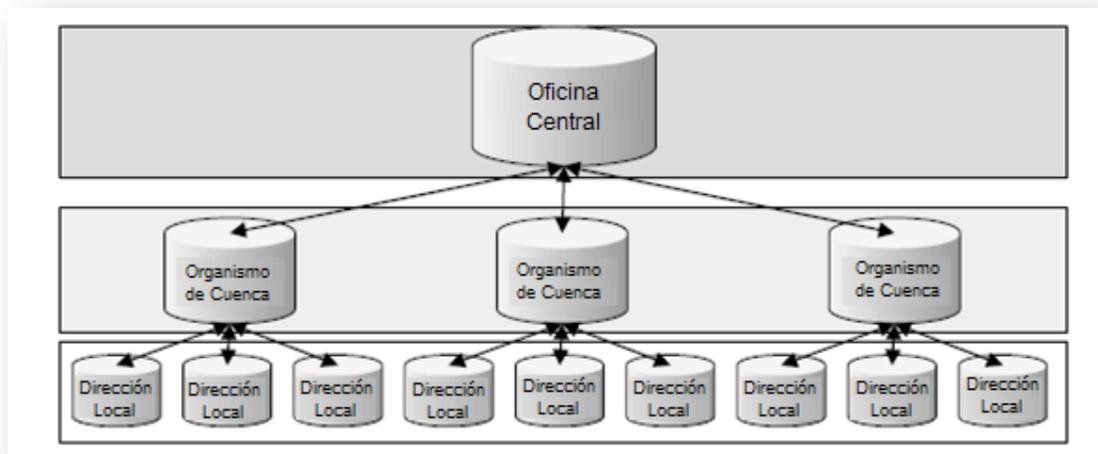


Figura 34. Geobase de Datos Institucional del Agua distribuida en forma multinivel

Trabajo con contratistas: La CONAGUA, en muchas ocasiones, requiere trabajar con personal externo a la Institución quienes son contratados para

desarrollar alguna aplicación y quienes, a su vez, necesitan acceder a la Geobase de datos Institucional. Con una infraestructura basada en un sistema de bases de datos distribuida, es posible compartir una copia de una porción de la geobase de datos con el contratista sin necesidad de interrumpir las actividades cotidianas. El contratista podrá entonces efectuar actualizaciones periódicas las cuales estarán incorporadas a la geobase de datos Institucional, en caso de ser necesario, conforme a lo mostrado en la Figura 35.

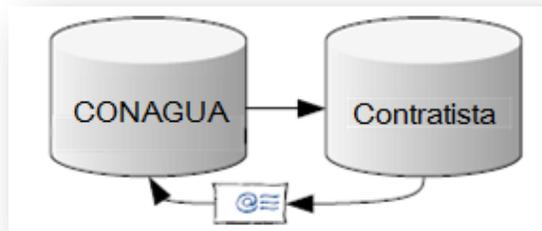


Figura 35. Acceso por parte de personal externo a segmentos requeridos sin afectar la operación del resto de los equipos

Arquitectura de alta disponibilidad.

En la Figura 36 se muestra el modelo del manejo transaccional de datos. Este modelo, permite pensar en una base de datos que cuente con varias computadoras, llamadas también nodos, que están intercomunicados mediante una red de telecomunicaciones para la transferencia de datos y órdenes entre sitios. Estos sitios están distribuidos geográficamente a lo largo y ancho del territorio nacional Mexicano a través de redes a larga distancia. El tipo de topología de red tiene un impacto importante en el rendimiento, sobre las estrategias para el procesamiento de las consultas distribuidas y el diseño de la base de datos distribuida; aunque en cuanto al aspecto arquitectónico de alto nivel esto no debe ser un problema que afecte demasiado.

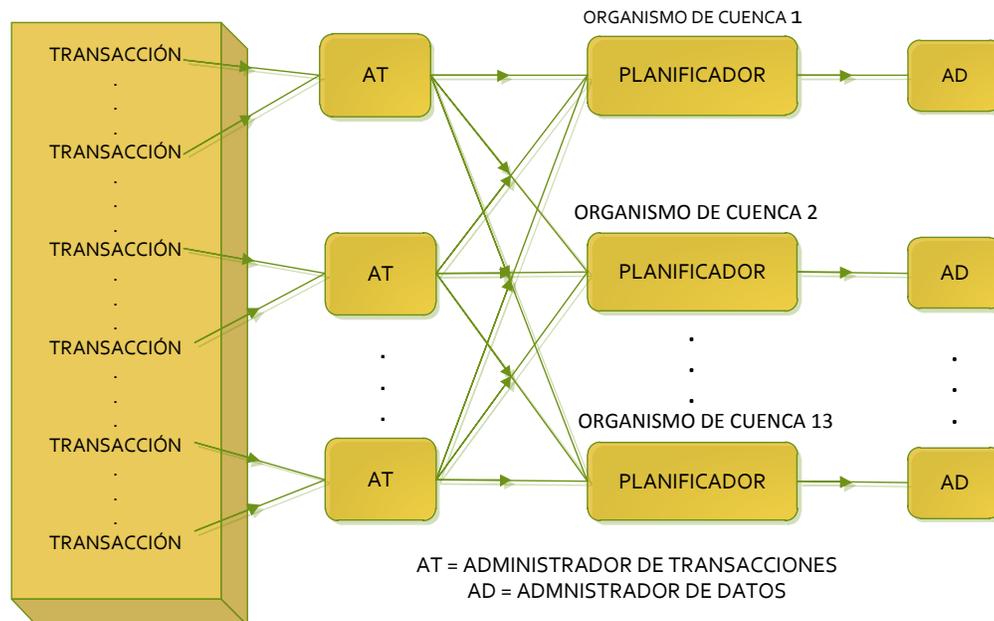


Figura 36. Modelo general del SBDD

Las transacciones son administradas a través de los mecanismos de control de concurrencia los cuales deben asegurar la consistencia de los objetos y cada transacción atómica será completada en un tiempo finito. Estos métodos de control deben ser serializables¹¹⁴; es decir, debe existir una secuencia equivalente en que las operaciones de cada transacción aparecen antes o después de otra transacción pero no entremezcladas. Una ejecución serial de transacciones es siempre correcta.

Extendiendo los argumentos para la serializabilidad y los algoritmos de control de concurrencia para la ejecución en ambientes distribuidos, se deben sincronizar las transacciones concurrentes de los usuarios.

Uno de los problemas de la concurrencia, en las bases de datos distribuidas, es la consistencia de copia múltiple, las cuales se producen cuando un mismo dato está en varias ubicaciones al mismo tiempo. Si cada planificador de ejecución local es serial y las órdenes locales serializadas son idénticas; lo cual significa

¹¹⁴ La serialización (o *marshalling* en inglés) consiste en un proceso de codificación de un objeto en un medio de almacenamiento (como puede ser un archivo, o un buffer de memoria) con el fin de transmitirlo a través de una conexión en red como una serie de bytes o en un formato humanamente más legible como XML o JSON, entre otros. La serie de bytes o el formato pueden ser usados para crear un nuevo objeto que es idéntico en todo al original, incluido su estado interno (por tanto, el nuevo objeto es un clon del original). La serialización es un mecanismo ampliamente usado para transportar objetos a través de una red, para hacer persistente un objeto en un archivo o base de datos, o para distribuir objetos idénticos a varias aplicaciones o localizaciones.

que respetan el orden de secuencia, entonces el planificador global es serializable.

Para el control de concurrencia, en una base de datos distribuida se tienen dos alternativas:

1. El bloqueo (lock) que garantiza la ejecución concurrente. Se debe tener cuidado de que no se produzcan interbloqueos.
2. Las marcas de tiempo las cuales garantizan la ejecución concurrente según el orden fijado en estas marcas. Esto asegura la serialización global.

La función principal de un procesador de consultas relacionales es transformar una consulta en una especificación de alto nivel, normalmente en cálculo relacional, a una consulta equivalente en una especificación de bajo nivel, normalmente alguna variación del álgebra relacional¹¹⁵.

La transformación es correcta si la consulta de bajo nivel tiene la misma semántica que la consulta original, es decir, si ambas consultas producen el mismo resultado. Para verificar si es correcta la transformación se hace un mapeo bien definido entre el cálculo relacional y el álgebra relacional.

Recuperación ante desastres.

Pasos para el diseño e implementación de la solución.

Para implementar un adecuado esquema de replicación es necesario realizar los siguientes pasos:

1. Diseñar el ambiente.
2. Creación de los links entre servidores.
3. Creación de los snapshots.
4. Creación de grupos.

La iniciativa esencial para la consolidación de la geoconexión consiste en la creación de geobase de datos del agua. La geobase constituye un conjunto continuo y plenamente integrado de datos geoespaciales a escala nacional que brinda información de contexto y de referencia. La geobase está fundada sobre la necesidad de coleccionar los datos una sola vez, lo más próximo a la fuente generadora, y de crear valor a partir de estos datos mediante el desarrollo de varias aplicaciones.

¹¹⁵ <http://hp.fciencias.unam.mx/~alg/bd/alg.pdf>

La Geobase de Datos Institucional del Agua constituye una iniciativa nacional supervisado por la Subgerencia de Información Geográfica del Agua (SIGA).

La iniciativa geobase constituye una innovación considerable puesto que ningún interesado por sí solo ni posee ni controla la base de datos en cuestión. Los interesados crean, proveen, mantienen y distribuyen datos geográficos provenientes de fuentes diferentes. La iniciativa de la Geobase de Datos Institucional del Agua permite la integración vertical de datos provenientes de fuentes federales, estatales y municipales. La geobase constituye entonces un instrumento primordial para el desarrollo sostenible de recursos, para la seguridad pública, para la protección del medio ambiente.

En lo particular la experiencia del Grupo Temático del Agua en este ámbito constituye un ejemplo de óptimas prácticas en materia de acceso a los bancos de información. Hemos aprendido a romper barreras institucionales y a consolidar las bases de datos de nuestros asociados para crear una variedad más enriquecida de información para los usuarios del sistema.

Se ha descubierto que cuando se trabaja en colaboración se logra más. De esta manera se favorece el acervo de datos geoespaciales al que los usuarios tienen acceso lo que resulta en más gente haciendo uso creativo de los datos

Una vez que se dispone de los datos básicos de la arquitectura del sistema y de los datos complementarios provenientes de los diversos interesados, se precisa de una manera para acceder a lo que este así disponible.

Cuando dos consultas provenientes de diferentes integrantes de la red tratan de actualizar el mismo elemento de datos o si el sistema falla durante la ejecución de una consulta; intuitivamente se puede pensar que el concepto principal que debe manejar la base de datos es la de una ejecución consistente de consultas. Por eso es necesario hacer uso de una transacción la cual es entendida como una unidad básica de cómputo consistente y confiable.

Así, se espera tener, por un lado, una transparencia adecuada de las acciones concurrentes a una base de datos y por otro, tener transparencia en el manejo de las fallas que se pueden presentar en la base de datos. De esta forma las transacciones proporcionan una ejecución atómica y confiable en presencia de fallas, una ejecución correcta en presencia de accesos de usuario múltiples y un manejo correcto de réplicas.

EL SBDD de la GeoAgua debe contemplar el hecho de que participen varios Organismos de Cuenca en la ejecución de una transacción por lo que es más

difícil garantizar la propiedad de atomicidad. El fallo de uno de estos Organismos de Cuenca o el fallo de la línea de comunicación entre ellos pueden llevar a un resultado erróneo. Es por esto que existe el gestor de transacciones cuya función es asegurar la ejecución atómica de las transacciones gestionando la ejecución de aquellas (locales o globales) que acceden a datos almacenados en su localidad y el coordinador de transacciones que es el encargado de coordinar la ejecución de varias transacciones iniciadas en su localidad.

A más bajo nivel nos encontramos con la necesidad de transferir datos entre el almacenamiento en disco y la memoria principal. De esto se encarga el gestor de búferes.

En el SGBDD de la GeoAgua, todos estos módulos se encuentran tanto a nivel local en cada equipo como a nivel de nodo virtualizado. Estos últimos son los denominados gestores globales de transacciones.

El procedimiento a seguir cuando se ejecuta una transacción global en un nodo N_1 es la siguiente:

- El gestor global de transacciones del nodo N_1 , divide la transacción en una secuencia de subtransacciones, siguiendo la información guardada en el catálogo global del sistema.
- El encargado de la comunicación de datos del nodo N_1 envía dichas subtransacciones a los nodos adecuados, por ejemplo N_2 y N_3 .
- Los gestores globales de transacciones de los nodos que reciben los datos, se encargan de gestionarlos y los resultados de las secuencias de instrucciones SQL se devuelven a través del encargado de la comunicación de datos al primer nodo N_1 .

3.3. Implementación de la primera propuesta de solución.

3.3.1. Modelo conceptual.

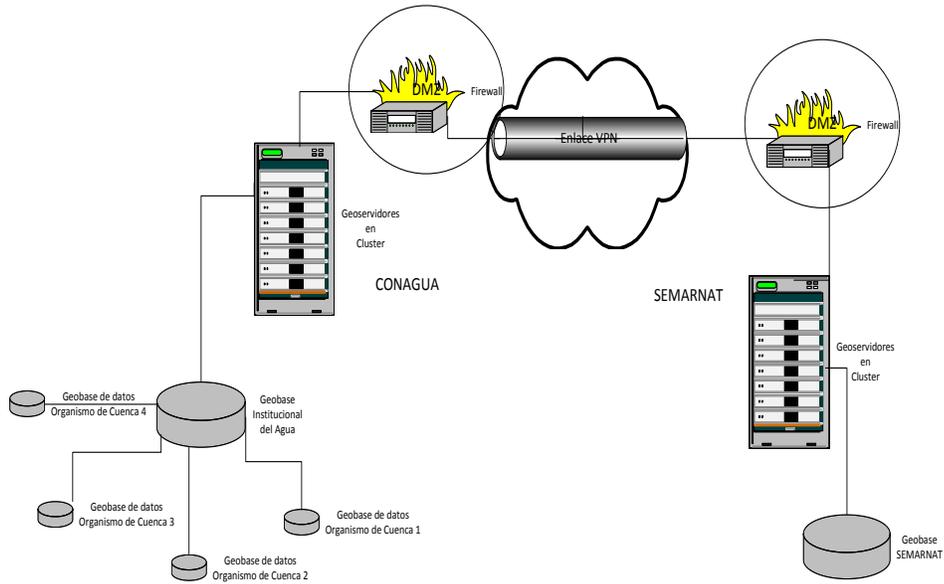


Figura 37. Diagrama del modelo conceptual de operación de geobase de datos distribuida e interoperabilidad técnica

Diagrama de colaboración.

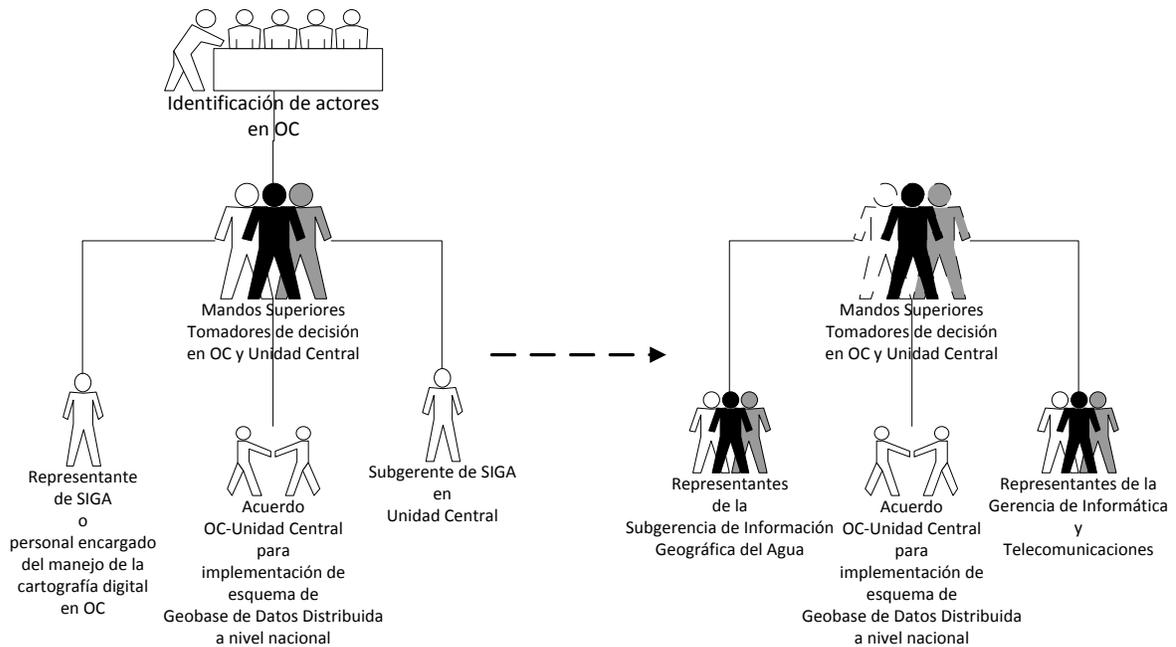


Figura 38. Diagrama de colaboración

La conexión a la Geobase de Datos del Agua se puede efectuar con cualquier programa que cuente con servicios de conexión a bases de datos espaciales; entre estas herramientas se encuentran ArcGIS (ArcMap y ArcCatalog), ArcExplorer y ArcExplorer Java for Education. La interconexión entre las geobases de datos de CONAGUA y otras dependencias del sector se realizará a través de elementos que garanticen la seguridad de acceso a los datos. En este sentido se conectarán los servicios a través de los firewalls institucionales, conforme se indica en la Figura 37.

Por otro lado será necesario establecer relaciones de colaboración y confianza entre los entes actores que administran las geobases de datos, en este sentido será necesario generar una estructura de colaboración entre estos, como se indica en la Figura 38, en la cual se puede observar el trabajo requerido para involucrar a los tomadores de decisión en el desarrollo del proyecto.

3.4. Replicación de geobases de datos entre Unidad Central y Organismos de Cuenca.

Las técnicas orientadas a objetos y las extensiones asociadas a los sistemas de gestión de bases de datos relacionales (RDBMS) tratan otros problemas técnicos y, en cierta medida, problemas asociados con esquemas de datos¹¹⁶. Sin embargo, estos esfuerzos únicamente proporcionan la estructura tecnológica para el intercambio de datos y no aseguran que los datos sean conocidos o estén disponibles ni que las organizaciones vayan a adoptar la tecnología que permita su uso.

La arquitectura del sistema distribuido se basa en un formato en el cual cada procedimiento tiene su propio gestor de transacciones locales, cuya función es asegurar las propiedades ACID¹¹⁷, este gestor ayuda para la ejecución de las transacciones globales, de forma que cada procedimiento del sistema contiene en sí dos subsistemas:

- a. El gestor de transacciones, en sí mismo el cual gestiona las transacciones que tienen el total acceso a los datos guardados en la geobase local, pero tomando en cuenta que las transacciones de este tipo solo son locales.
- b. El coordinador de transacciones: coordina la ejecución de las diferentes transacciones tanto locales como globales.

¹¹⁶ "Spelling out the Spatial Data Soup", GeoWorld, marzo de 2000, pag. 38

¹¹⁷ c.f. Capítulo 1. Transacciones en base de datos distribuidas

El seguimiento de los datos y de las transformaciones asociadas a los mismos se hace más difícil en la medida que los datos se distribuyen más ampliamente y se mantienen descentralizadamente. Esto ha llevado a incrementar la importancia de tecnologías como los almacenes de datos, los servicios de metadatos y al aumento de la utilización de herramientas de modelado de la información.

3.5. Medición de la funcionalidad de la alternativa de solución.

El análisis de viabilidad de un sistema de información se centra en cuatro áreas de interés básico: técnica, económica, operativa y programática. Para este análisis se emplearán los siguientes indicadores:

- Tipo y características de conexiones existentes.
- Grupos de trabajo entre dependencias existentes.
- Acceso a métodos de comunicación entre dependencias.
- Actores involucrados.
- Identificar y definir las características de los participantes clave.
- Entender las relaciones entre los elementos en estudio, individuos, incluyendo la evaluación de los conflictos potenciales y reales de intereses así como las expectativas de cada uno de ellos.
- Nivel de responsabilidad.

Funcionalidad.

Medición de la distribución de la geobase de datos: El valor de un sistema interoperable puede ser medido como función del número de transacciones conectadas por segundo. La función puede ser representada por:

$$\text{Interoperabilidad} \approx \sum_{i=1}^m \lambda^i \left(t! \frac{1}{(t-n)!n!} \right)$$

Dónde:

t es el número total de transacciones a interoperar.

N es número de transacciones que necesitan ser combinadas para completar un proceso.

M es el número de subsistemas

λ es el factor de correlación

Para medir estos parámetros se utilizó el medidor de rendimiento del sistema de Windows, en el servidor de Geobases de Datos, a fin de medir las transacciones por segundo, las transacciones activas y las pendientes de duplicación durante un día laborable normal.

Los valores promedio reportados son de 6 transacciones/seg véase Figura 39 y Figura 40.

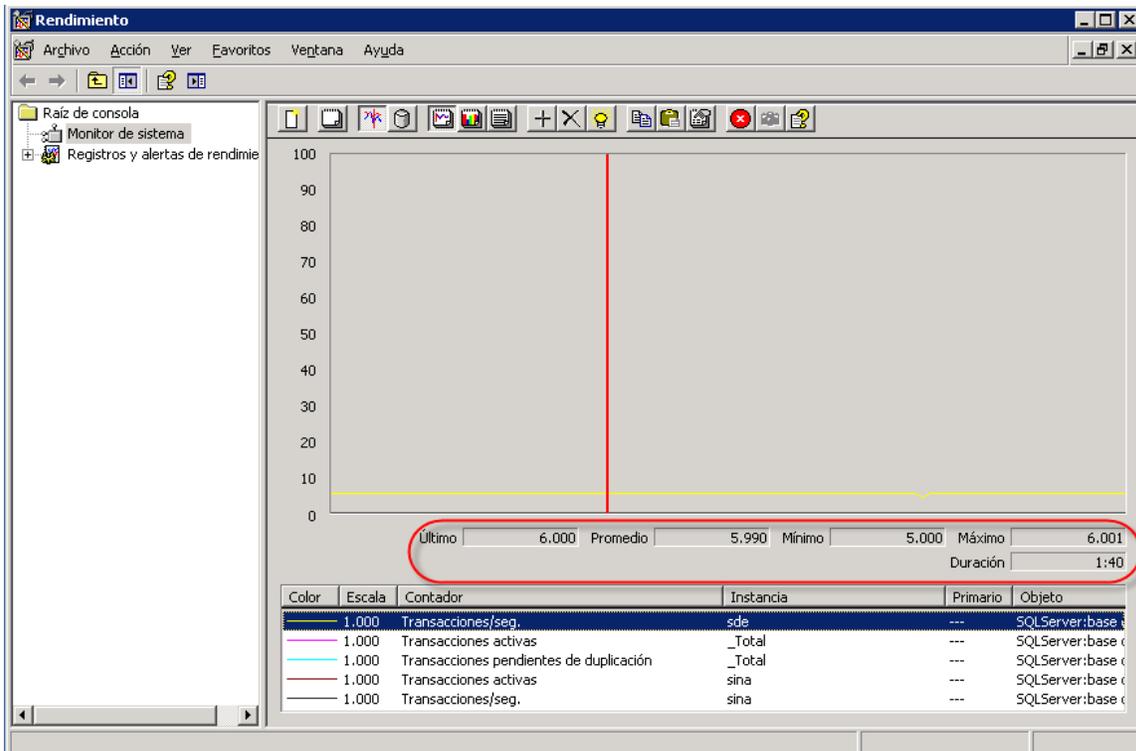


Figura 39. Proyecto de consulta de información geográfica a partir de las capas de información contenidas en la GEOAGUA

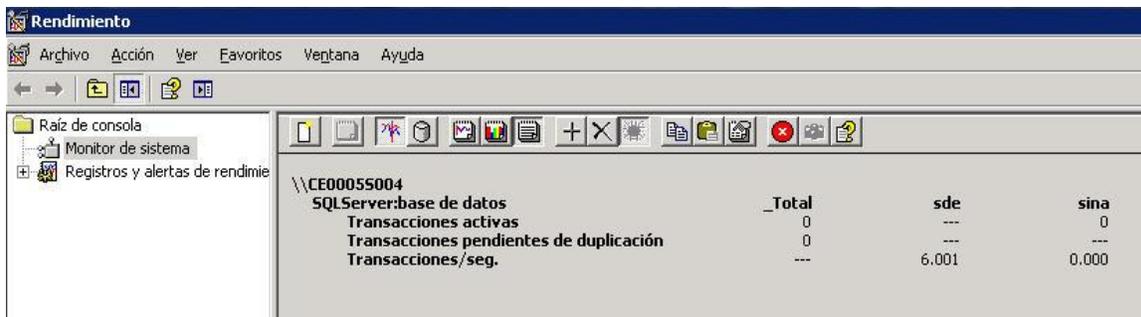


Figura 40. Reporte de transacciones por segundo reportadas para la GeoAgua

Por lo tanto el valor del parámetro t es de 6. Es decir se tiene 1 transacción cada 0.16 seg.

El número de transacciones que necesitan ser combinadas para completar un proceso es de 2.

Usuario	Base de datos	Estado	Transacciones abiertas ▾
CE00055004\Administrador	master	runnable	2
sde	constructor	sleeping	1
sde	sina	sleeping	1
sde	sde	sleeping	1

El subsistema es único por lo que m tiene un valor de 1.

El factor de correlación es 1. De lo anterior obtenemos un valor de interoperabilidad de 15.

Capítulo 4. LINEAMIENTOS, APORTACIONES, MEJORAS Y ACCIONES A FUTURO

Los métodos señalados en el capítulo 3, permitieron obtener resultados que garantizan la implementación de un modelo de vinculación de geobases de datos totalmente interoperable ente la CONAGUA y SEMARNAT. Para tal efecto fue necesario establecer un modelo de vinculación de bases de datos que solucione el problema y permita garantizar una solución viable. Este último capítulo se divide en tres partes. La primera hace referencia al cambio que se requiere para la implementación de la propuesta de interoperabilidad. La segunda presenta las políticas de interoperabilidad necesarias para la viabilidad del proyecto a futuro así como la implementación de los estándares a los que deberá sujetarse el proyecto para garantizar su óptimo funcionamiento; es decir, el compartir información entre organizaciones implica la optimización de los recursos, la reducción de los costos de producción y mantenimiento tanto de la información misma como de aplicaciones así como desarrollo de acuerdos y políticas para la homogeneización de la información geográfica (GSDI, INSPIRE, IDEE) además de la adopción de estándares y el retorno de la inversión efectuada. Finalmente, en la tercera sección sugiere una iniciativa relativa a la implementación de conformar un esquema denominado Geoinformación Gubernamental Sostenible. Adicionalmente se presentan otras dos iniciativas; la creación de: 1) una red denominada Sistema de Gobierno Electrónico Digital Interoperable de Recursos Hídricos (SGEDIRH) y 2) un esquema Nacional de Interoperabilidad Conjunto de criterios y recomendaciones en materia de seguridad, conservación y normalización de la información, de los formatos y de las aplicaciones que deberán ser tenidos en cuenta que garanticen la interoperabilidad.

4.1. El cambio que se requiere para el trabajo de geobases de datos distribuidas.

La interoperabilidad representa un “medio de traspaso” de información entre los organismos gubernamentales. Para implementarla en el ámbito de gobierno es necesario un cambio cultural en los servicios públicos y sus funcionarios. Para el efecto es fundamental poner al ciudadano como centro de atención y objetivo esencial del trabajo realizado a la par de colaborar con otras Instituciones gubernamentales a través de la generación de nuevas capacidades y una de las más importantes es la interoperabilidad. Esto lleva aparejado el hecho irrefutable

de tomar conciencia y hacerse consciente de que las labores y desempeño de un área o departamento e inclusive el trabajo de toda una dependencia gubernamental necesariamente afectan a otras. Esta implicación hace imperativo el identificar, caracterizar y ser capaz de controlar los procesos críticos de negocios realizados en la vinculación de bases de información y sus componentes: conocer la interrelación que tienen sus procesos con los de las demás dependencias de gobierno lo cual, a su vez, está asociado a los servicios que presta a la ciudadanía.

Operar con estándares de calidad definidos y controlables así como la incorporación de tecnologías efectivas en sus procesos de obtención y procesamiento de datos para la obtención de información confiable, oportuna, actualizada, accesible y altamente disponible conllevaría a mejorar la calidad de los servicios a las personas y facilitaría el acceso a información que los otros Organismos Públicos requieren. Lo anterior permitiría coadyuvar a la simplificación de los procedimientos administrativos asociados a los trámites de modo que no sería necesario solicitar a las personas la información que ya se posee en el Estado.

Transparentar ante el ciudadano, la información que el Estado posee considerando el cumplimiento de los requisitos legales en materia de privacidad. Este objetivo es satisfecho mediante la implementación de sistemas interoperables los cuales faciliten que la información entregada al ciudadano cumpla con los requisitos legales de confiabilidad de los datos y, en el futuro, se ajuste a una nueva normatividad para el acceso¹¹⁸ de datos entre Organismos Públicos de Gobierno. Por ello es preponderante establece los términos mínimos legales y normativos necesarios para permitir y regular el intercambio de información entre los Organismos del Estado esto facilitará soportar, por la red interoperable antes propuesta, los diferentes grados de madurez tecnológica de los Organismos del Estado mediante la comunicación entre los distintos tipos de tecnología presentes en los estos. Cabe aclarar que para lograrlo es fundamental garantizar altos niveles de seguridad tanto en el acceso a la información como en las transacciones lo cual se alcanzará si se permite la interoperabilidad con base en condiciones y estándares predefinidos.

¹¹⁸ En este punto se hace referencia al acceso a los bancos de información con los que cada dependencia gubernamental cuenta en lugar de referirse al "intercambio" de información como tal, debido a que la capacidad tanto de procesamiento como de almacenamiento de información es finito y, por ende, es necesario la implementación de una estrategia que facilite la consulta al todo el acervo de datos con los que cuanta a fin de evitar duplicidad y pérdida de vigencia en la información almacenada.

El conformar una red interoperable entre todas las dependencias del sector Medio Ambiente, y mejor aún considerar todo el Gobierno Federal en su conjunto, traería grandes beneficios globales entre los que destacan:

- Disminución de tiempos de acceso a los datos tanto en los organismos gubernamentales como en los trámites de consulta de los ciudadanos.
- Simplificación del proceso de trámites.
- Proporciona una base escalable de interoperabilidad para los servicios públicos.
- Mayor eficacia y eficiencia de los organismos del Estado.
- Posibilidad de contar con información oportuna y de calidad para la toma de decisiones por parte de las organizaciones públicas.
- Facilita la auditoría de las transacciones efectuadas y su propósito, protegiendo el derecho a la privacidad de las personas.
- Permite un uso más eficiente y racional de las plataformas tecnológicas, ya existentes, de los Servicios Públicos.
- Posibilidad de articular y dar respuesta a solicitudes complejas de los ciudadanos o los servicios públicos a través de una plataforma inteoperada.
- Evita posibles discrecionalidades en la provisión de información
- Estándares de interoperabilidad publicados, disponibles y regulados.
- Monitoreo continuo de los niveles de servicio y reglas de negocio establecidas.
- Apoyo al diseño o de nuevos servicios de información y articulación de solicitudes de información.
- Permite la trazabilidad de los requerimientos de información y de las solicitudes de nuevos servicios.
- Relevamiento, en su caso, y nivelación de madurez digital en Instituciones Gubernamentales.

4.2. Políticas de geobases de datos distribuidas, estándares, retorno de inversión y geoinnovaciones. Iniciativas a futuro.

La determinación de la implementación de una red¹¹⁹ de interoperación y la construcción de nodos e interoperaciones entre todas las instituciones del sector medio ambiente y del Gobierno Mexicano en su conjunto con base en una descripción formal de sus componentes, tanto geoespaciales como de tipo documental y su interrelación, además de requerir de políticas generales a nivel nacional es necesario contar con un modelo de madurez para el Gobierno Electrónico y un modelo normativo-descriptivo que permita medir esta madurez para implementar buenas prácticas de e-gobierno y determinar “cuánto le falta” a una institución para incluirse en esta red interoperable. El aspirar a implementar este esquema en el ámbito nacional requerirá de personal bien capacitado en el manejo y administración de bases de datos tanto espaciales como relacionales distribuidos así como en el desarrollo de esquemas y metadatos lo cual permitirá consensuar y publicar la estructura de los documentos compartidos de gobierno.

Dado lo anterior se sugiere la creación de una red¹²⁰ denominada Sistema de Gobierno Electrónico Digital Interoperable de Recursos Hídricos (SGEDIRH) el cual permita implementarla y dar un seguimiento a los proyectos desarrollados por las Instituciones de gobierno que conformes esta red.

4.2.1. Estándares en Geobases de Datos Distribuidas.

Las transacciones por medio electrónico entre dependencias de gobierno con ciudadanos u otras dependencias gubernamentales se facilitan, a través de la interoperabilidad, el intercambio de información. No obstante, ello depende de poder compartir la información de forma universal, independientemente de la tecnología que soporte su almacenamiento, procesamiento o distribución. Así la interoperabilidad ha acompañado a la evolución del hombre y al desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC) desde sus propios inicios. Pensemos por un momento en la creación de la escritura y el propio alfabeto, pasando por los sistemas de numeración o de medición del tiempo, hasta la invención de la imprenta para difundir esos conocimientos o los protocolos que actualmente hacen posible la comunicación en Internet. El objetivo ha sido y será siempre el mismo: comunicar, dialogar y entenderse.

¹¹⁹ También conocida como framework de interoperabilidad o interoperable.

¹²⁰ Ibid.

En su acepción más simple, según se mencionó en el capítulo 1, la interoperabilidad trata de asegurar que los sistemas trabajen juntos. La interoperabilidad en el software se puede lograr de muchas maneras alternativas o complementarias, pudiendo basarse en estándares, especificaciones, interfaces abiertos, documentación técnica, etc. Esto contrasta con el mundo del hardware, en el que sólo se alcanza basándose en estándares.

Los estándares consisten en definiciones, formatos o procesos que han sido aprobados por unas determinadas organizaciones de estandarización o aceptados “de facto” como tales por la industria. Los estándares, como activadores de la capacidad de comunicación, pueden ser de dos tipos: los estándares abiertos y los estándares propietarios, ambos con capacidad de ser ampliamente aceptados.

Los **estándares abiertos**¹²¹ son especificaciones técnicas disponibles públicamente, sin restricciones de acceso o implementación por parte de ningún fabricante y desarrolladas mediante un proceso de consenso. En la Figura 41 se muestran algunas de las principales compañías que desarrollan software en estándares abiertos contrarias a las filosofías de desarrollo y ventas utilizadas por las compañías propietarias de diseño de software que se muestran en la Figura 42. Hay muchas definiciones, que los abordan desde diferentes perspectivas, pero en todas ellas hay ciertos aspectos comunes: alrededor del desarrollo, de la aprobación y del mantenimiento del estándar debe haber un proceso abierto y basado en el consenso, el proceso debe ser transparente, está sujeto a los términos de propiedad intelectual RAND/FRAND (*reasonable and nondiscriminatory*) que no obligan pero permiten, según el criterio del poseedor de la propiedad intelectual a licenciarlo sin compensación económica alguna y la documentación debe ser de carácter público y accesible bajo términos razonables (gratuitamente o mediante una cuota determinada)¹²².



Figura 41. Logos de estándares abiertos en Geomática y SIG

¹²¹ <http://www.opengeospatial.org/>, <http://www.gsdi.org/>, http://www.idee.es/show.do?to=pideep_pidee.ES, <http://mapserver.inegi.gob.mx/geografia/espanol/cartcat/convencion/menu/5116.pdf>

¹²² Algunos ejemplos de éstos son HTTP, HTML, TCP/IP, 802.11, XML, SOAP, el lenguaje SQL, etc.

Por otro lado, los **estándares propietarios** son especificaciones técnicas desarrolladas y mantenidas por una sola entidad, sea un grupo privado o un consorcio, disponibles comercialmente bajo acuerdos de licencias sujetos a propiedad intelectual, que regulan el alcance de uso de los mismos¹²³.



Figura 42. Logos de estándares propietarios en Geomática y SIG

La ventaja de los estándares abiertos es que permiten interoperar en un mercado de libre competencia entre múltiples implementaciones de hardware y software, si bien la interoperabilidad no solo se apoya en estándares abiertos: los estándares propietarios son la consecuencia de aportaciones importantes en I+D financiadas por compañías que perciben ventajas en sus inversiones, de tal forma que sus definiciones, formatos o procesos llegan en no pocas ocasiones a ser aceptados también mayoritariamente por la industria.

No debemos pasar por alto que en este ambiente de convivencia, siempre se debe evitar una sobreestandarización, para no frenar el desarrollo y asegurar la libertad de elección sobre la manera de intercambiar la información.

En la Administración Pública, la puesta en práctica de programas de interoperabilidad basada en estándares abiertos puede suponer aplicaciones y beneficios en el despliegue de e-servicios de gobierno-a-gobierno, empresa-a-gobierno y ciudadano-a-gobierno.

Los programas de interoperabilidad basados en estándares son capaces de aprovechar con éxito las inversiones existentes en la Administración Pública y facilitan el medio para que los sistemas existentes puedan participar en un más amplio entorno de sistemas de información. Mediante el aprovechamiento de las mejores prácticas de interoperabilidad, los sistemas de las Administraciones Públicas son capaces de mejorar:

Los servicios centrados en los ciudadanos. Asegurando el despliegue de servicios públicos para facilitarles transparencia, a través de información relevante, y a ser posible en cualquier momento y lugar.

La eficacia operacional. Facilitando la alineación de procesos empresariales y tecnológicos en la Administración Pública para trabajar con más eficacia,

¹²³ Como ejemplo de éstos podemos citar Java, el formato PDF de Adobe, las APIs Win32 o el formato Microsoft Office XML.

percibiéndose como una organización colectiva más que como un conjunto de almacenes de información separados.

4.2.2. El retorno de la inversión ROI¹²⁴

La interoperabilidad entre nuevos entornos y sistemas existentes permite cualquier migración gradual, eficiente y evolutiva a nuevas plataformas, aprovechando las inversiones existentes y asegurando la libertad de elección ante las futuras. Así pues, mediante soluciones de interoperabilidad basadas en estándares los sistemas y aplicaciones podrían formar parte de una red transparente de servicios para la propia administración, la sociedad y el ciudadano, que ayuden a resolver problemas del mundo real, de manera eficiente y rentable.

Las Administraciones Públicas buscan la manera de hacer más eficientes sus procesos y operaciones internas, y más enfocadas al ciudadano. En este sentido, las TI juegan un papel clave para permitirles organizarse y gestionarse de una manera más eficiente y rentable. Las decisiones de negocio verdaderamente informadas y la creación de políticas sólo son posibles cuanto mejor y con más eficacia se hacen disponibles los datos a través de todo el abanico de sistemas de la Administración Pública que contienen la información pertinente.

La interoperabilidad puede ayudar a resolver estos problemas y alcanzar estos objetivos, proporcionando la capacidad de intercambiar la información entre sistemas con más eficacia y haciendo un mejor uso de la misma. Creando esencialmente un mundo en el que el resultado es mayor que la suma de las partes.

Concretamente, los programas de interoperabilidad que alcanzan el éxito pueden dar soporte a importantes soluciones sociales y políticas, tales como la accesibilidad, la identificación de usuario, la privacidad y la seguridad; promueven la capacidad de elección, la competencia y la innovación; reducen gastos de adquisición y mantenimiento; fomentan el acceso abierto a la información y tienen en cuenta la compatibilidad hacia atrás; incrementan la

¹²⁴ Son las siglas en inglés de Return On Investment y es un porcentaje que se calcula en función de la inversión y los beneficios obtenidos, para obtener el ratio de retorno de inversión. Es un valor que mide el rendimiento de una inversión, para evaluar qué tan eficiente es el gasto que estamos haciendo o que planeamos realizar. Los valores de ROI cuanto más altos mejor. Si tenemos un ROI negativo es que estamos perdiendo dinero y si tenemos un ROI muy cercano a cero, también podemos pensar que la inversión no es muy atractiva.

eficacia, la flexibilidad y el valor de las inversiones existentes en sistemas; y aumentan la transparencia hacia los usuarios, proporcionándoles información de valor añadido, a ser posible en tiempo real y obtenida al reunir datos procedentes de diversos bancos de información.

La implementación de sistemas interoperables debe reconocer tanto la heterogeneidad de actores y relaciones, estructuras de datos, modelos de procesos, soluciones tecnológicas como las arquitecturas e implementaciones en cada dependencia gubernamental. Así este proceso de instauración debe ser progresivo basado no solamente en esquemas de interoperabilidad técnica sino en aspectos de interoperabilidad semántica y disociación *estándar–aplicación*.

Este proyecto implica algo más que el ahorro de coste y tiempo, implica el derecho de los participantes a los servicios de administración electrónica de la información los cuales se garantizan en la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental (LFAIPG)¹²⁵ lo cual implica no condicionar el acceso a la información, libertad de elección de la fuente de datos, vigencia en el tiempo así como la libertad de implementación.

Este escenario de dinámicas de participación y transparencia requiere tanto la Elevación de las especificaciones a organismos de normalización (INEGI) como la publicación de las especificaciones para su conversión en norma, dando como resultado el conocimiento y escrutinio público de las especificaciones, análisis y comparación de las especificaciones y un debate abierto y público.

Así, se sugiere un esquema Nacional de Interoperabilidad Conjunto de criterios y recomendaciones en materia de seguridad, conservación y normalización de la información, de los formatos y de las aplicaciones que deberán ser tenidos en cuenta que garanticen la interoperabilidad.

Dado lo anterior es importante considerar la utilización de estándares abiertos así como, en su caso y de forma complementaria, estándares que sean de uso generalizado. Con ello se puede elaborar con la participación de todas las Administraciones una propuesta de uso de estándares interoperables en la Administración Pública Sectorial del Medio Ambiente la cual se mantenga actualizada y de manera permanente. A través de la generación de un grupo que dimensione la interoperabilidad, la cadena de interoperabilidad en si misma así como los estándares aplicables: su clasificación, selección, ciclo de vida, fuerza normativa, conformidad se podrá garantizar los elementos esenciales para la

¹²⁵ [http://www.ordenjuridico.gob.mx/Federal/PE/PR/Leyes/11062002\(1\).pdf](http://www.ordenjuridico.gob.mx/Federal/PE/PR/Leyes/11062002(1).pdf)

implementación de una red interoperable, en particular, las infraestructuras de interoperabilidad en sí mismas.

4.2.3. Geoinnovaciones¹²⁶

Una tercera iniciativa se denomina Geoinnovaciones. Se sugiere que este programa se concentre sobre la capacidad del Gobierno Federal especializado en geomática para desarrollar tecnologías geomáticas avanzadas y aplicaciones en conjunto con el sector privado, a fin de acelerar la conexión de bases de datos del sector público con la infraestructura.

Una gran parte de la industria geomática Mexicana consiste de pequeñas y medianas empresas. Son empresas muy competitivas e innovadoras. Pero la mayor parte de empresas de nuevas tecnologías requiere más recursos para transformar los conceptos de investigación en productos y servicios aptos para el mercado el consumo del Gobierno Federal y en ocasiones no cumple con las necesidades de procesamiento y/ consulta.

Geoinnovaciones ayudaría a proporcionar financiamiento para tales proyectos. Por cada peso invertido por Geoinnovaciones las empresas y otros socios invertirían dos. Este financiamiento conjunto tendría un impacto considerable sobre el número, el tamaño y la sincronización de proyectos, y contribuiría a mejorar el nivel que tienen México en cuanto a la infraestructura de datos espaciales y de sistemas de información territorial con las que cuenta.

Los interesados ayudarían a crear la demanda que complementa el impulso de la oferta de las nuevas ideas y tecnologías provenientes del Sector Gubernamental.

Los instrumentos, las aplicaciones y los servicios que se generasen como resultado aumentarían y mejorarían los conjuntos de datos disponibles vía el Internet.

Adicionalmente se ha observado que la cantidad de usuarios de SIG en el sector gubernamental ha crecido rápidamente en los últimos años entonces este sector precisa contar con una fuente estable de trabajadores altamente capacitados. Para el efecto será necesario contar con estos profesionales de geomática altamente capacitados lo cual se podría lograr mediante la creación de un centro de intercambio de información y conocimientos basado en el Internet con el fin

¹²⁶ Coleman, David, "El papel de la cooperación en el desarrollo de la infraestructura Canadiense de datos espaciales", https://www.fig.net/pub/mexico/powerpoints/ts_03_coleman_span_ppt.pdf, Universidad de New Brunswick

de armonizar capacidades disponibles con las oportunidades de capacitación mediante cursillos y oportunidades de desarrollo profesional. El programa podría financiar a personal de varias Instituciones mediante la impartición de sus conocimientos a otros empleados que carezcan de estos conocimientos a fin de estandarizar y difundir en poco tiempo este conocimiento en geomática.

4.3. Iniciativa de Geoinformación Gubernamental Sostenible

Como otra iniciativa adicional se propone la creación de un programa que haga una gran diferencia positiva en las vidas de los mexicanos vinculada a las dependencias gubernamentales. Esta iniciativa que podría denominarse Geoinformación Gubernamental Sostenible haría disponibles tecnologías modernas de cartografía a los jefes de gobierno de comunidades rurales, costeras e indígenas, a través de redes interoperables para que estas comunidades puedan efectuar decisiones informadas afectando su desarrollo sostenible. Ello les permitiría a los líderes tomar decisiones y análisis de su entorno mediante:

1. La comprensión de sus propias necesidades y posibilidades de desarrollo;
2. La estimulación de la innovación local;
3. La toma de decisiones relacionadas con la planificación comunitaria y el ordenamiento territorial.

Mediante la entrega de datos, de un sistema de información geográfica y de actividades de capacitación la iniciativa de Geoinformación Gubernamental Sostenible proporcionaría a los líderes comunitarios los instrumentos para entender su situación local y para planificar acciones apropiadas.

De lo arriba presentado queda claro que existe una demanda latente considerable por las capacidades adquiridas en materia de geomática. Una vez que los gobiernos estatales y/o municipales se dan cuenta de que tipo de información está disponible querrán participar inmediatamente; eso, en sí mismo, genera beneficios indirectos más allá del mandato de la iniciativa. Esta iniciativa constituye un proyecto conjunto involucrando los gobiernos federal y estatal con el sector privado para fomentar la transferencia de tecnología del sector privado hacia el gobierno.

Capítulo 5. CARACTERÍSTICAS, ANÁLISIS DEL RENDIMIENTO, BENEFICIOS Y ESPECIFICACIONES DEL SISTEMA PROPUESTO

En este capítulo se efectúa un análisis del rendimiento esperado del sistema propuesto, mostrando las ventajas y desventajas al compararlo con un sistema similar desarrollado con técnicas habituales, mediante un cuadro comparativo.

En términos de los usuarios del sistema, típicamente el personal de los Organismos de Cuenca y las Direcciones Locales de la CONAGUA, se muestra un resumen de algunos de los beneficios que éstos obtendrán con la implementación de este sistema. Se explica la situación de estos antes y después de la existencia del sistema, señalando los beneficios y perjuicios en el desarrollo de sus tareas diarias, en cada caso.

Finalmente se describe el universo de la información y los métodos de revisión y validación requeridos para cada tipo de información implementada en las ETL (Extracción-Transformación-Carga) para evitar que la información se corrompa y se introduzca basura. Se indican cuáles serían los mejores indicadores de que la geobase está funcionando adecuadamente. Adicionalmente se describen los procedimientos de actualización y mantenimiento de la geobase para garantizar su actualización a lo largo del tiempo, indicando cuál es la inversión requerida y cuáles son las políticas de administración de usuarios, por ejemplo, por roles, tipos, asignación de herramientas.

5.1. Software de prueba utilizado para el análisis del rendimiento.

El programa utilizado para la realización de las pruebas de rendimiento fue el JMeter¹²⁷. Se trata de una aplicación de escritorio, de código abierto, tipo Java diseñada para determinar la funcionalidad, a través de pruebas de desempeño y carga, a fin de medir el rendimiento de alguna aplicación ya sea tipo web o de escritorio. Originalmente fue diseñado para probar las aplicaciones web, pero se ha expandido a otras funciones de prueba.

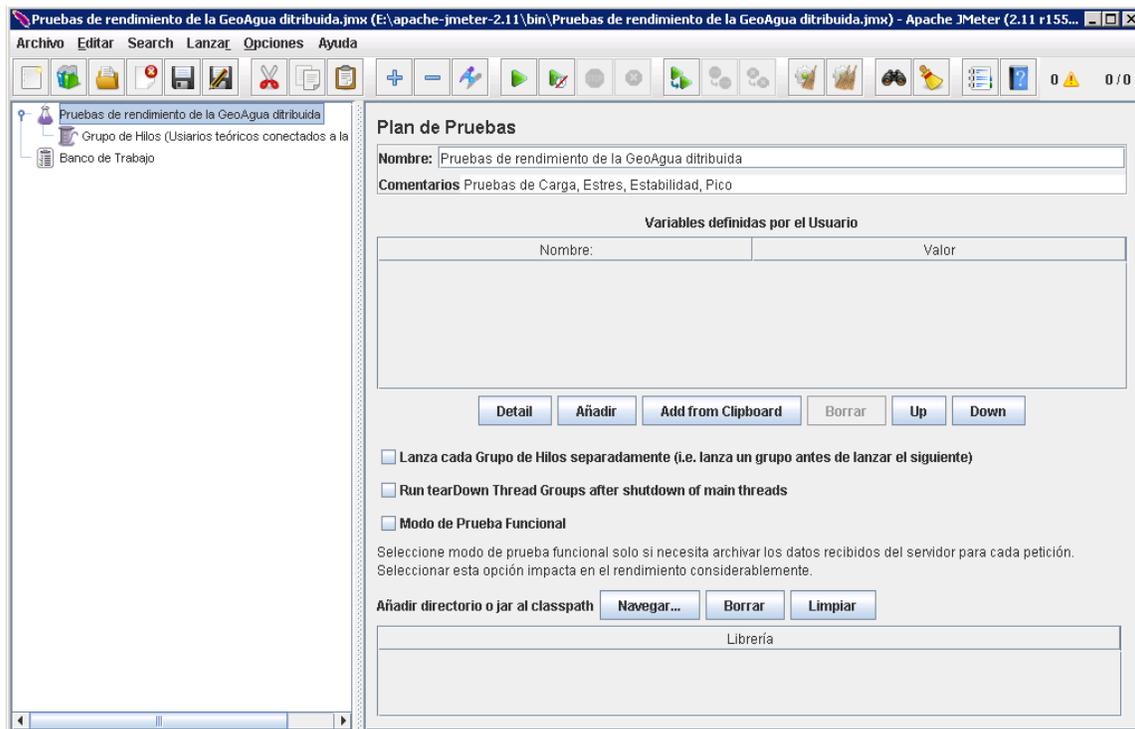
La aplicación JMeter puede ser utilizada para probar el rendimiento, tanto de recursos estáticos como dinámicos tales como archivos, lenguajes web dinámicos - PHP, Java, ASP.NET, objetos Java, bases de datos y consultas, servidores FTP y más. Se puede utilizar para simular una carga pesada en un servidor, grupo de servidores, la red o algún objeto, en sí mismo, para analizar el rendimiento general bajo diferentes tipos de carga. Se puede utilizar para hacer

¹²⁷ http://jmeter.apache.org/download_jmeter.cgi

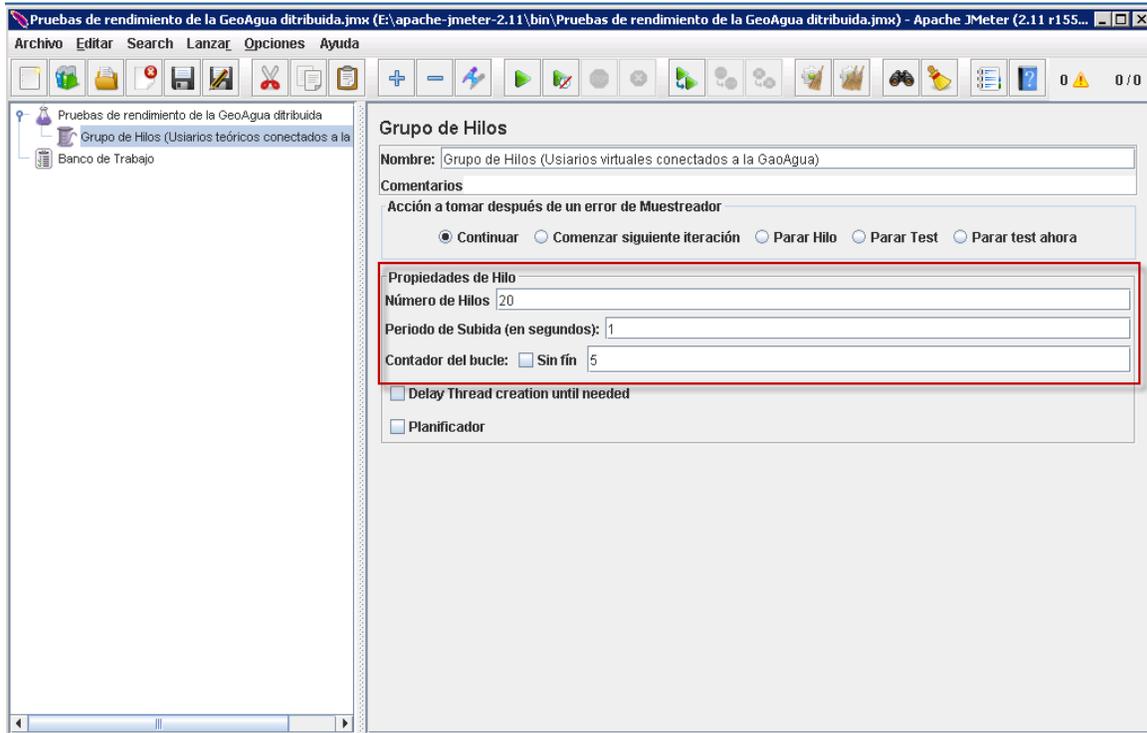
un análisis gráfico de rendimiento o para probar su comportamiento: objeto de servidor → script bajo carga pesada concurrente.

Configuración del software de prueba.

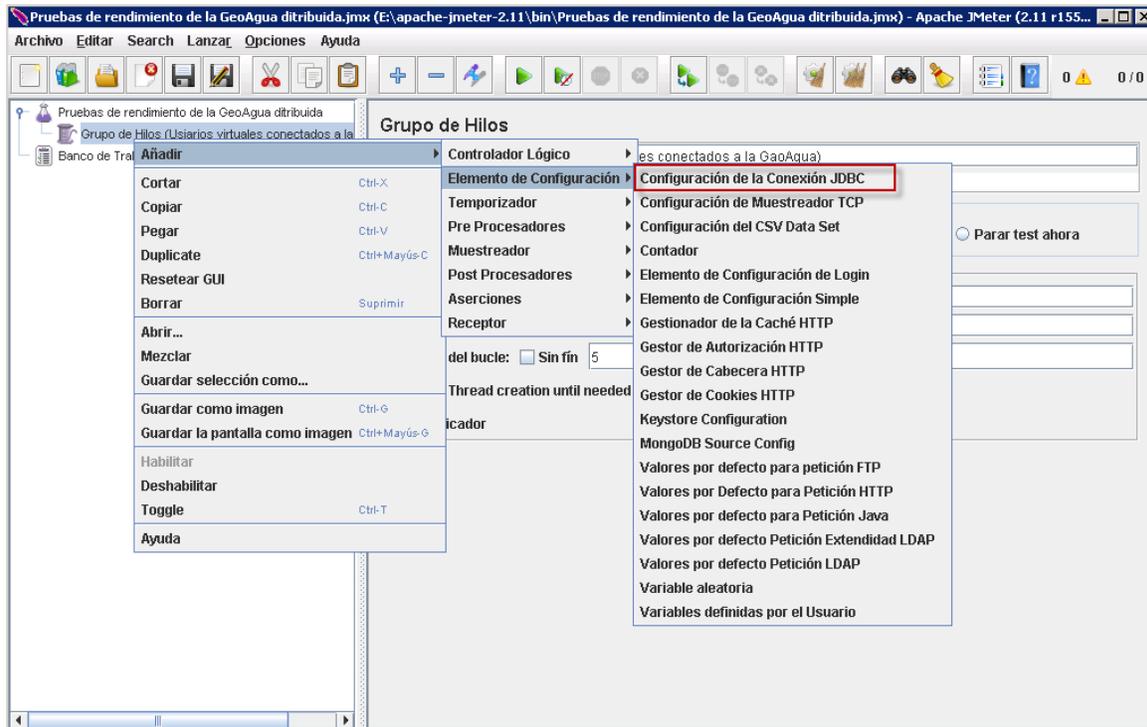
- a.- Creación de crea un plan de pruebas. Se asigna un nombre acorde al tipo del conjunto de pruebas a desarrollar y se guarda en una carpeta del disco duro local.

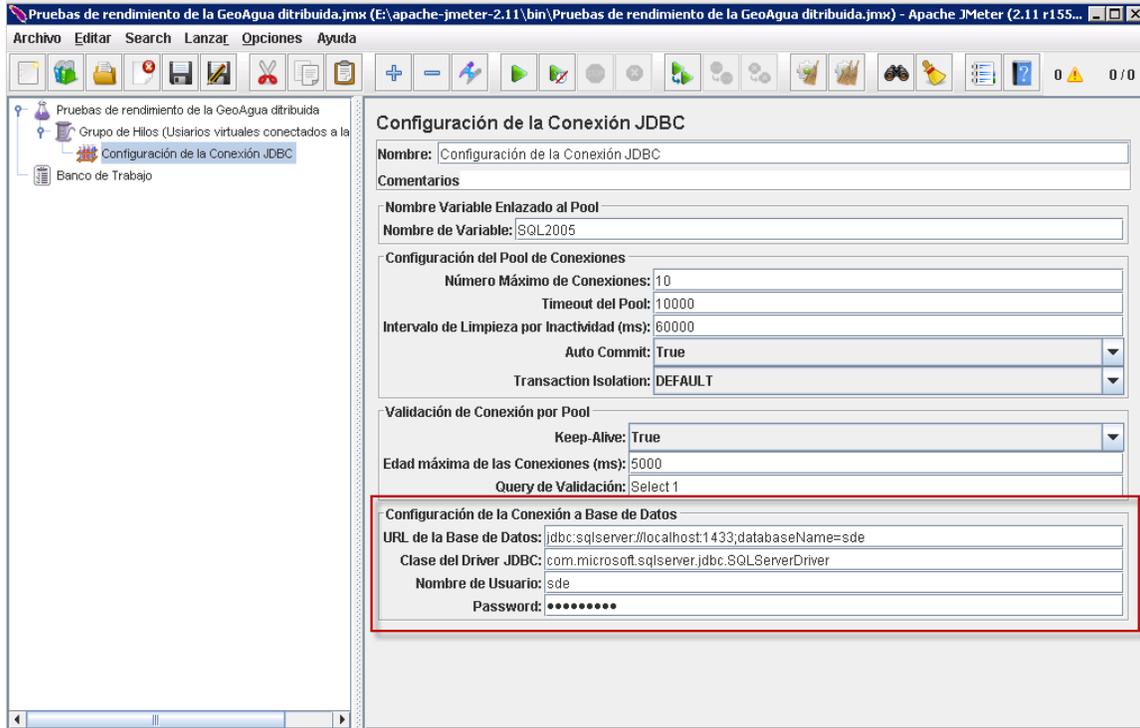


- b.- Creación de un Grupo de Hilos (usuarios virtuales asignados para las pruebas). Para este laboratorio se simulan 20 usuarios y 5 ciclos.

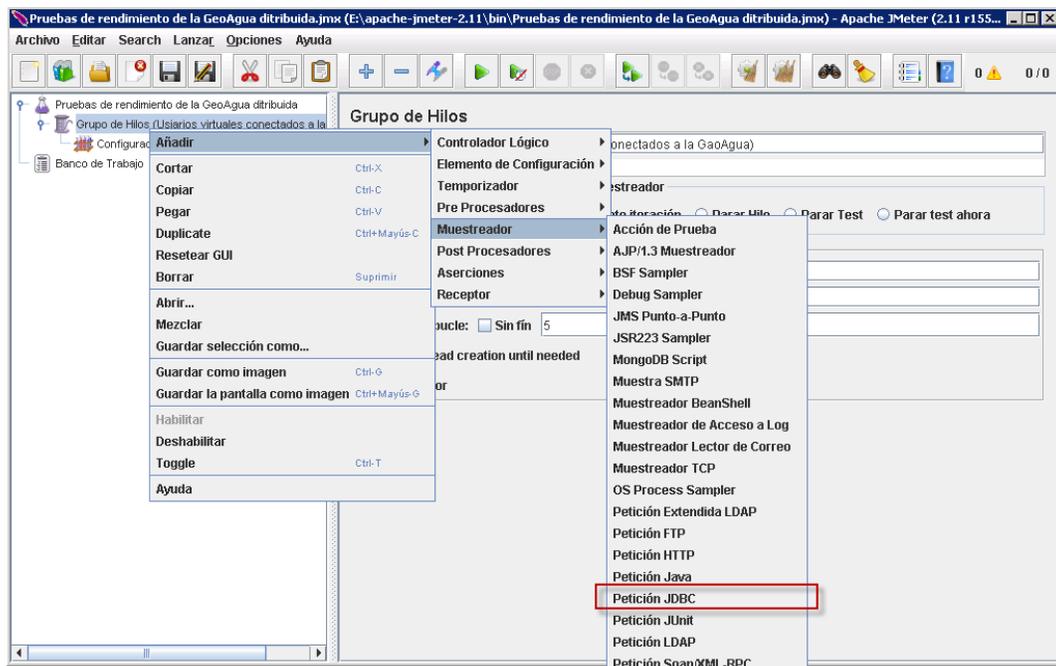


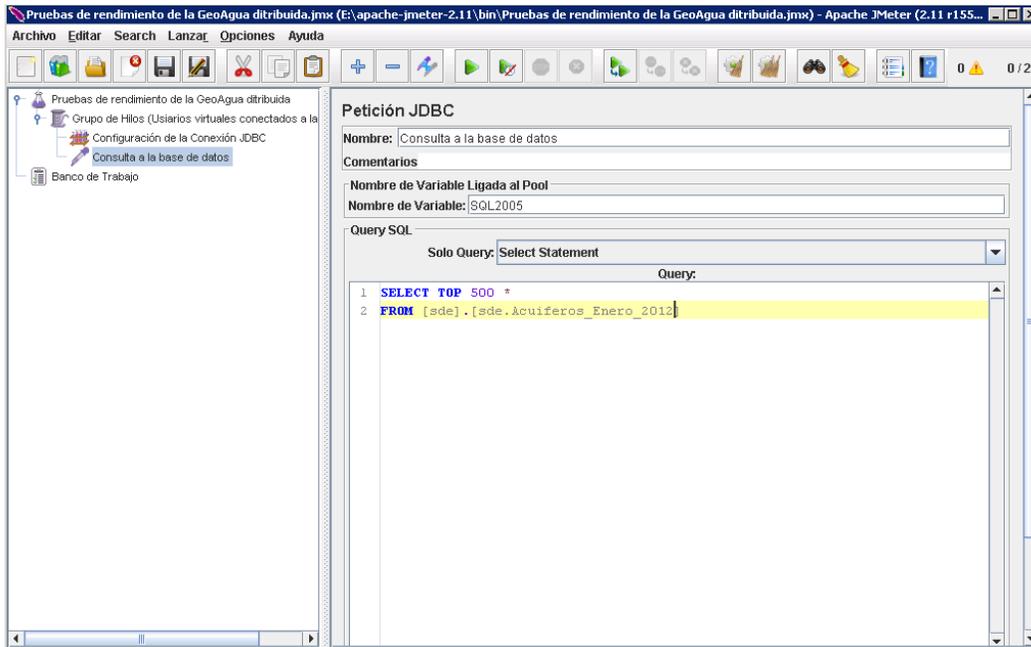
c.- Enseguida se procede a configurar la conexión JDBC, utilizando el puerto 1433, la base de datos GeoAgua, el driver SQL del fabricante Microsoft y las credenciales de DBA pertenecientes a la base de datos.



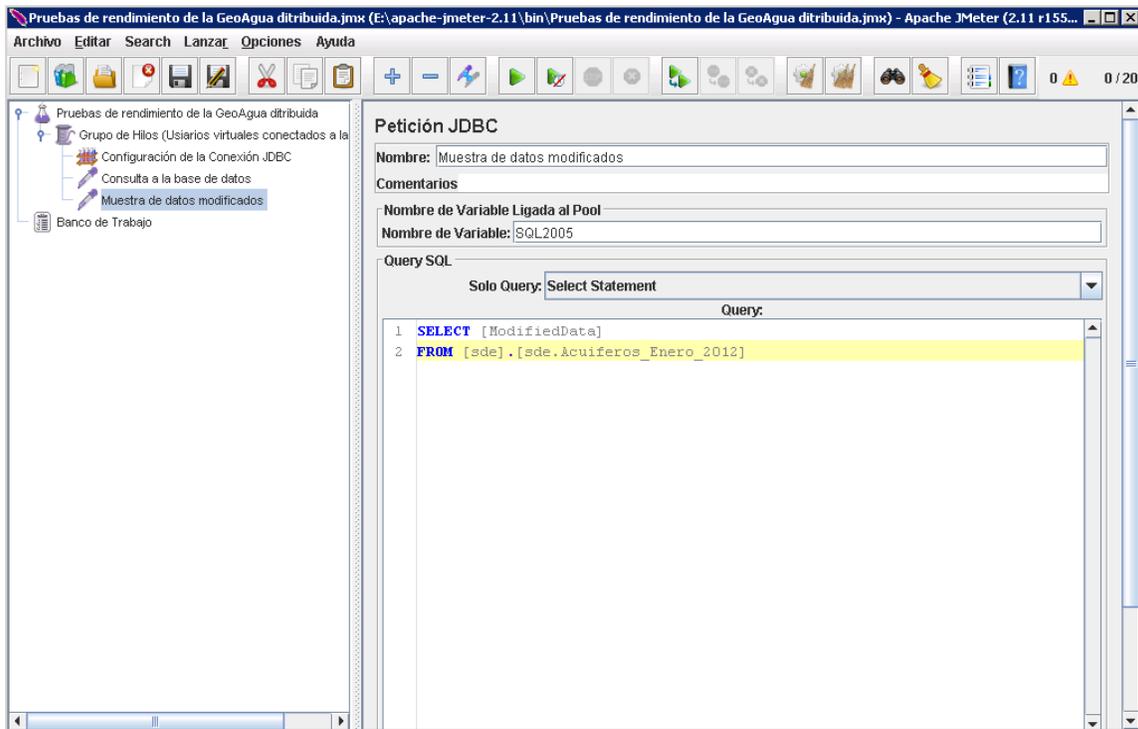


d.- Para continuar, es necesario hacer consultas a la base de datos GeoAgua. La primera consulta selecciona las primeras 500 filas de la tabla sde.Acuiferos:Enero_2012, como se ve a continuación:

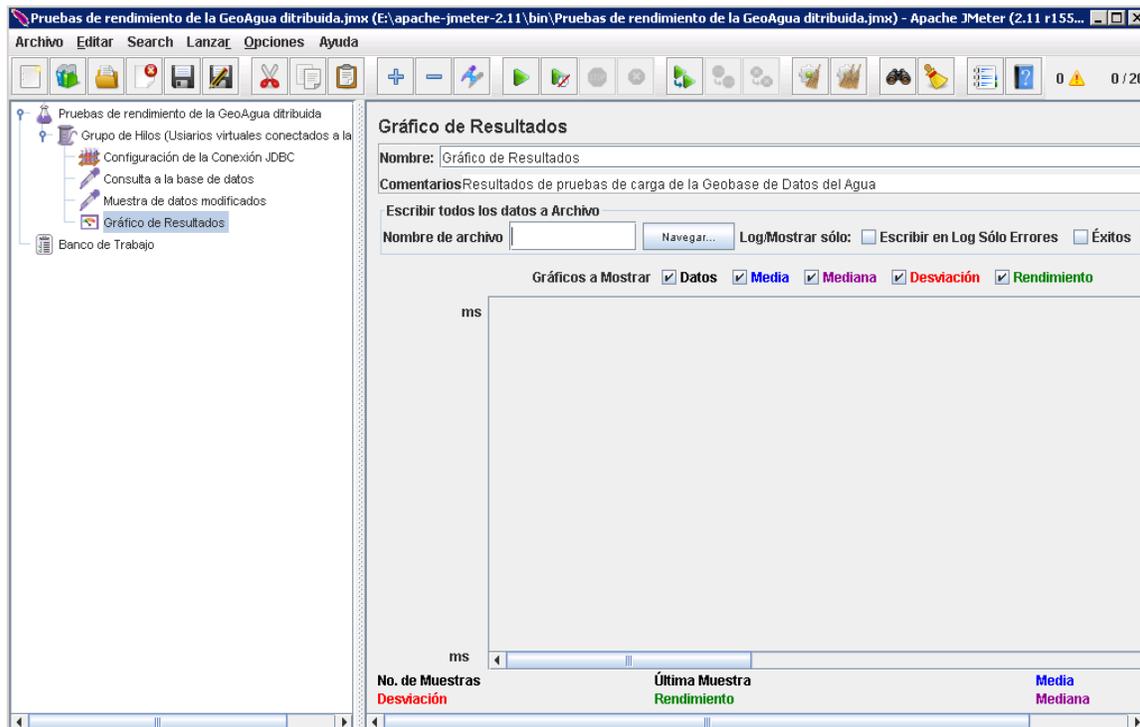
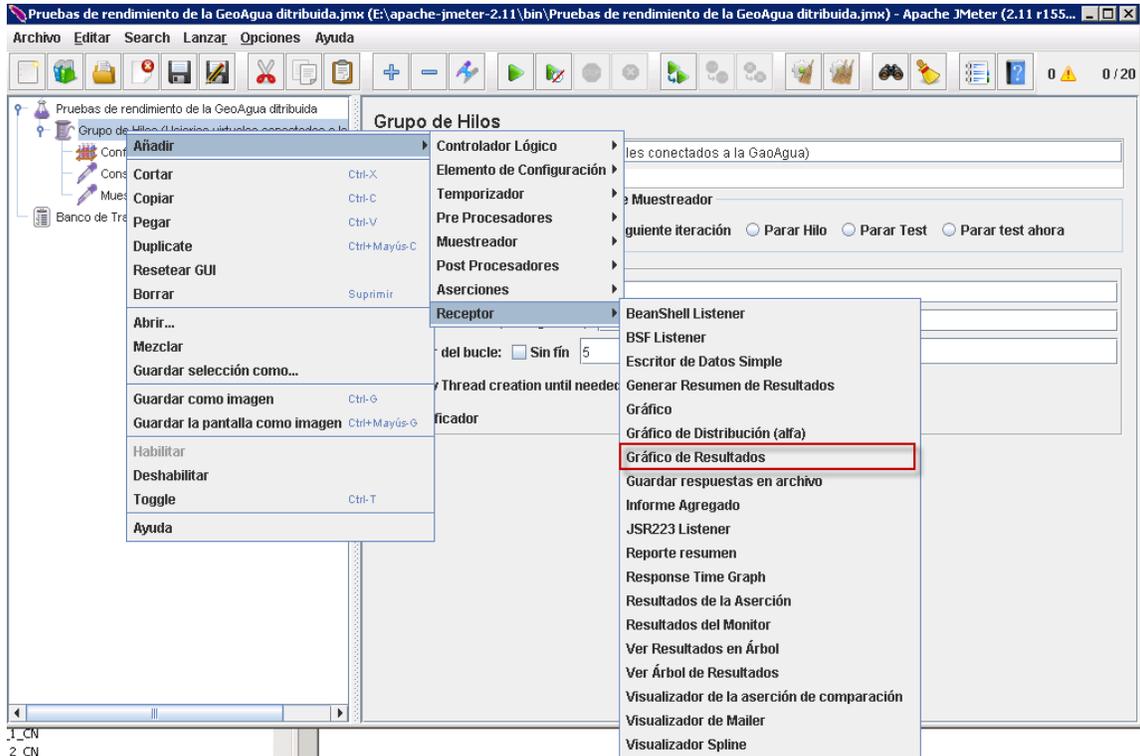




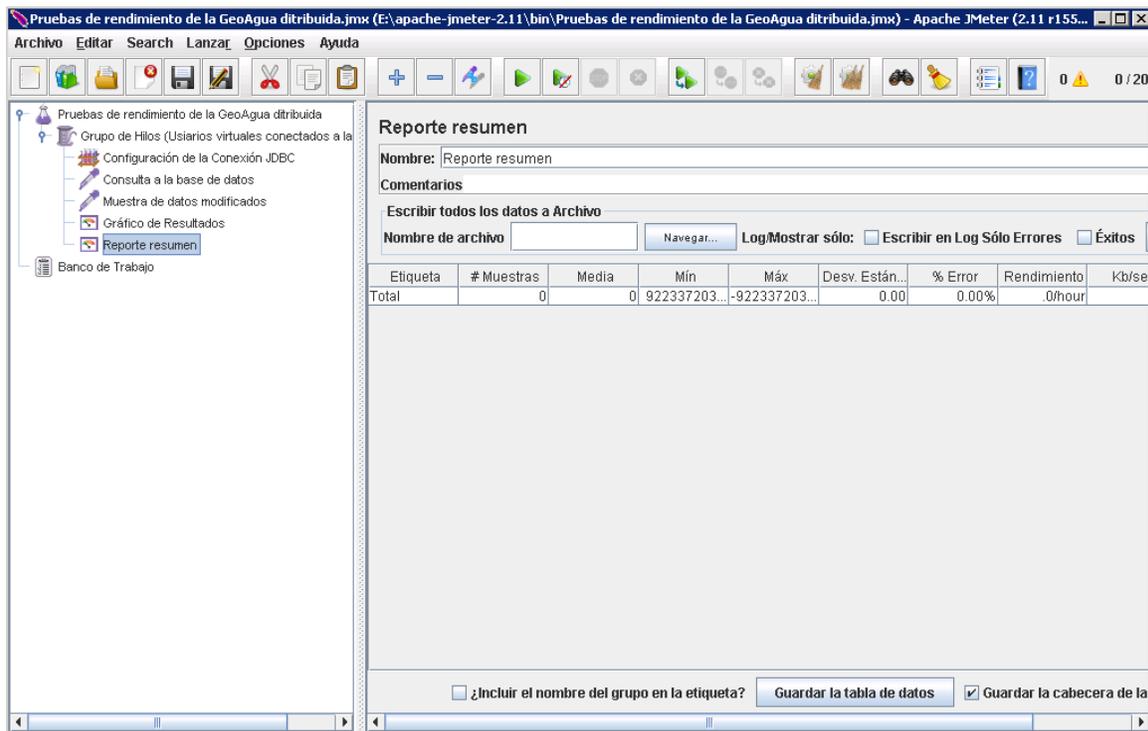
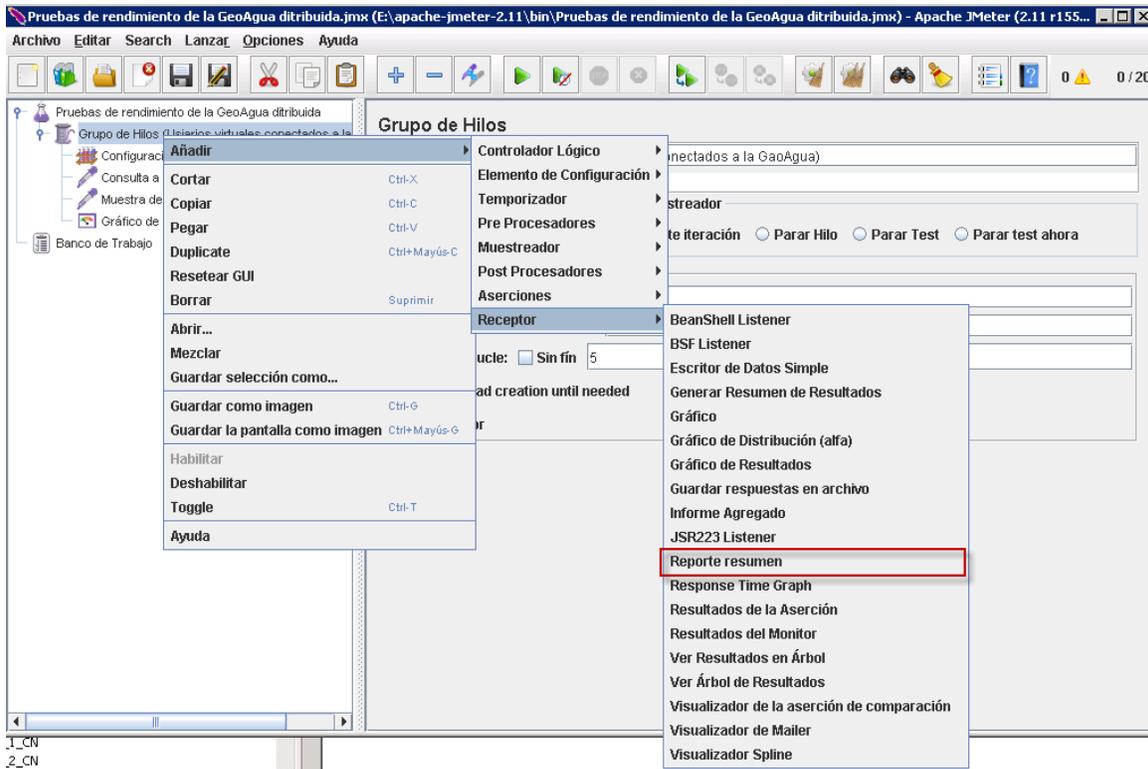
La segunda consulta del laboratorio muestra todas las filas de la columna ModifiedData de la tabla [sde].[sde.Acuiferos_Enero_2012].



e.- El siguiente paso es crear un gráfico con los resultados de la prueba de carga.



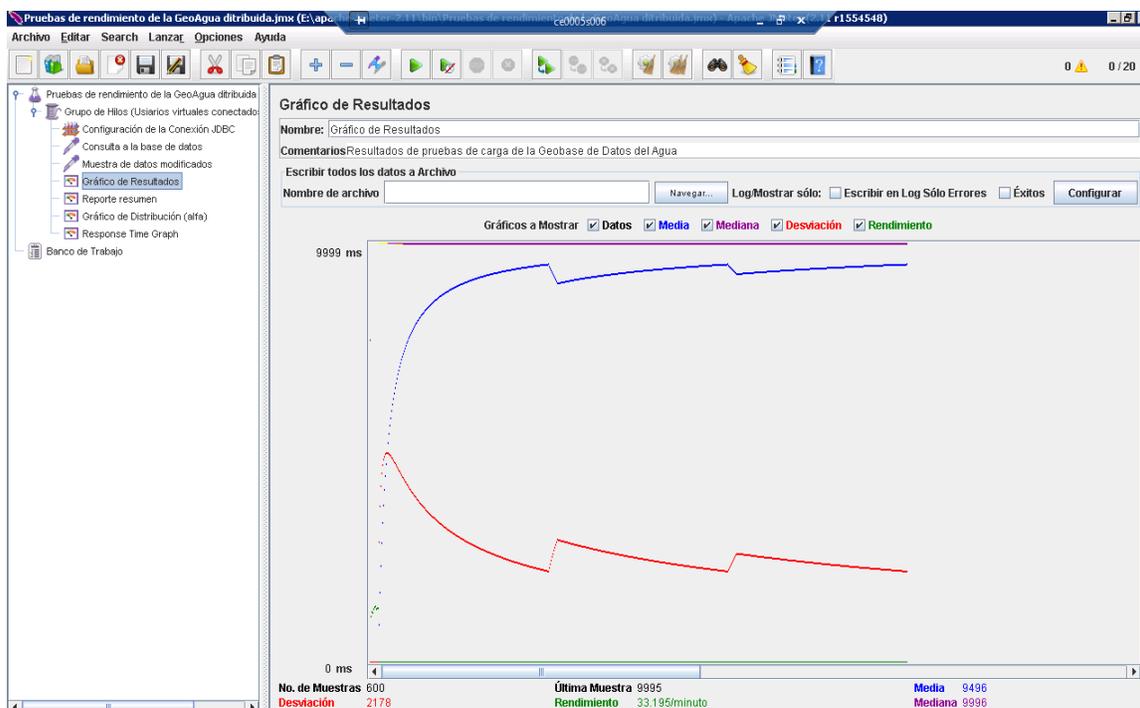
f.- Antes de ejecutar la consulta para visualizar los resultados se crea un reporte con el resumen del análisis.



5.1.1. Pruebas de carga

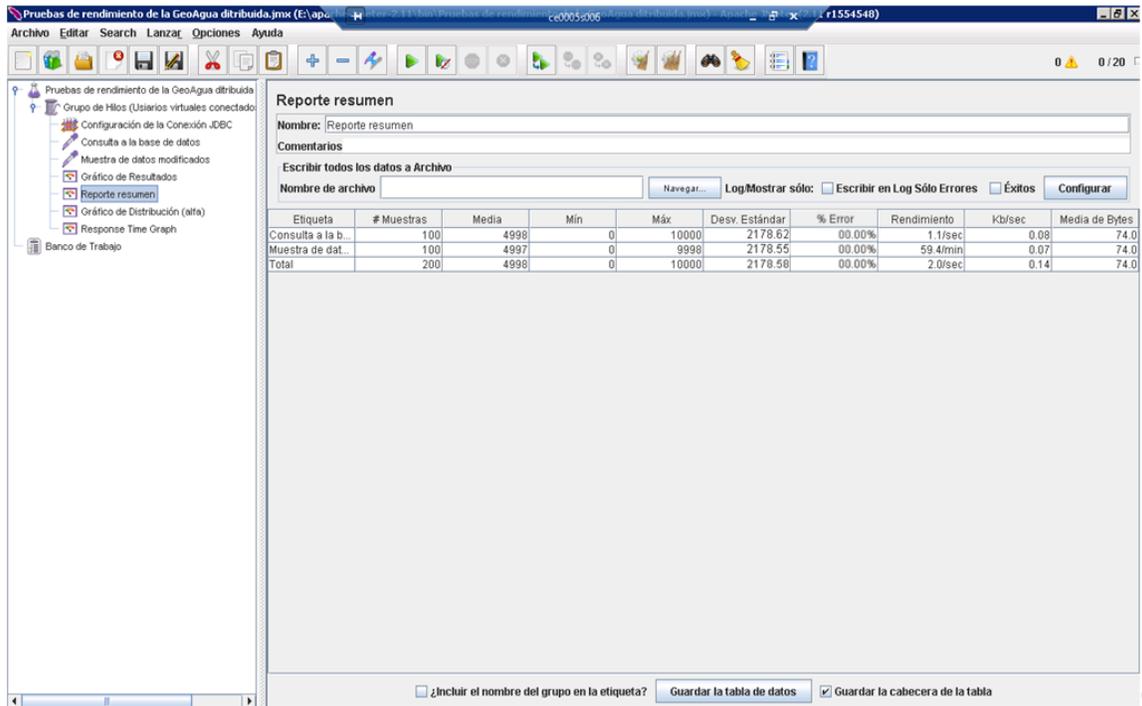
Subconjunto de pruebas de rendimiento. Se realiza mediante el incremento constantemente de la carga en la Geobase de Datos, comparándolo con un sistema similar desarrollado con técnicas habituales, hasta el momento en que alcanza el umbral límite. El objetivo principal de la prueba de carga es el de identificar el límite superior del sistema en términos de base de datos, y hardware de red, etcétera. Estas pruebas también se utilizan para calcular métricas como el tiempo medio entre fallos. En la Gráfica 1, se muestran los resultados de carga en la Geobase de Datos para una muestra de 9996 elementos simulando 20 threads o accesos de usuarios, con un número máximo de 10 conexiones y 5 ciclos. Obsérvese el tiempo de throughput (rendimiento) de 54.85/minuto. La Gráfica 2 muestra un resultado que crece de forma asintótica hacia el valor de 9999 ms. En donde se alcanza un punto de saturación o cuello de botella. Como puede verse, el tiempo promedio para acceder a una página es de 9,496 segundos, realizándose un total de 300 requerimientos al servidor. El tiempo total utilizado para los 20 threads, se puede calcular con la siguiente ecuación:

$$\text{Tiempo Total} = \# \text{Muestras} * \text{Media} = 20 * 9,496 = 189,920 \text{ milisegundos}$$



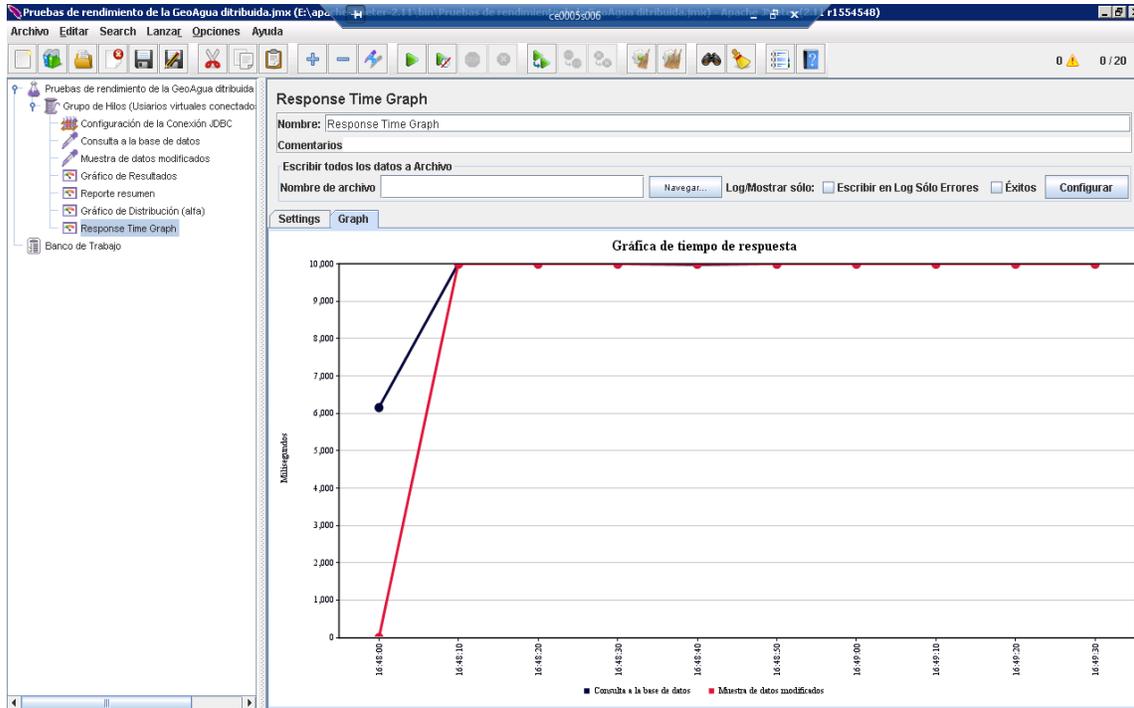
Gráfica 1. Resultados de carga en la Geobase de Datos para una muestra de 9996 elementos. Obsérvese el tiempo de throughput (rendimiento) de 54.85/minuto.

En la Gráfica 2 se presentan los valores de la última muestra, los cuales crean una fila por cada petición en el test. Obsérvese los valores del porcentaje de error cometido, del 0%, los cuales son una buena referencia respecto a la medición realizada.



Gráfica 2. Valores de la última muestra, crea una fila por cada petición en el test

En la Gráfica 3 se presentan los valores de los tiempos de respuesta. Obsérvese que tienden a estabilizarse hacia los 10 segundos. Lo anterior puede deberse a la forma de conexión del server en la NIC.



Gráfica 3. Tiempo de respuesta. Obsérvese que tiende a estabilizarse hacia los 10 segundos. Lo anterior puede deberse a la forma de conexión del server en la NIC.

Evaluación de datos generados por los procesos de prueba

En primer lugar se evaluó el resultado obtenido a través de un intervalo de confianza para una distribución Normal al 95% de confianza. La ecuación del mismo es la siguiente:

Ecuación 1 Definición de intervalo de confianza

$$\left[TP - \frac{(Z_{0.95})D}{\sqrt{N}}, TP + \frac{(Z_{0.95})D}{\sqrt{N}} \right]$$

Donde:

- Tiempo promedio (TP) = 4998
- Desviación (D) = 2178.62
- Tamaño de la muestra (n) = 300
- $Z_{0.95} = 1.96$

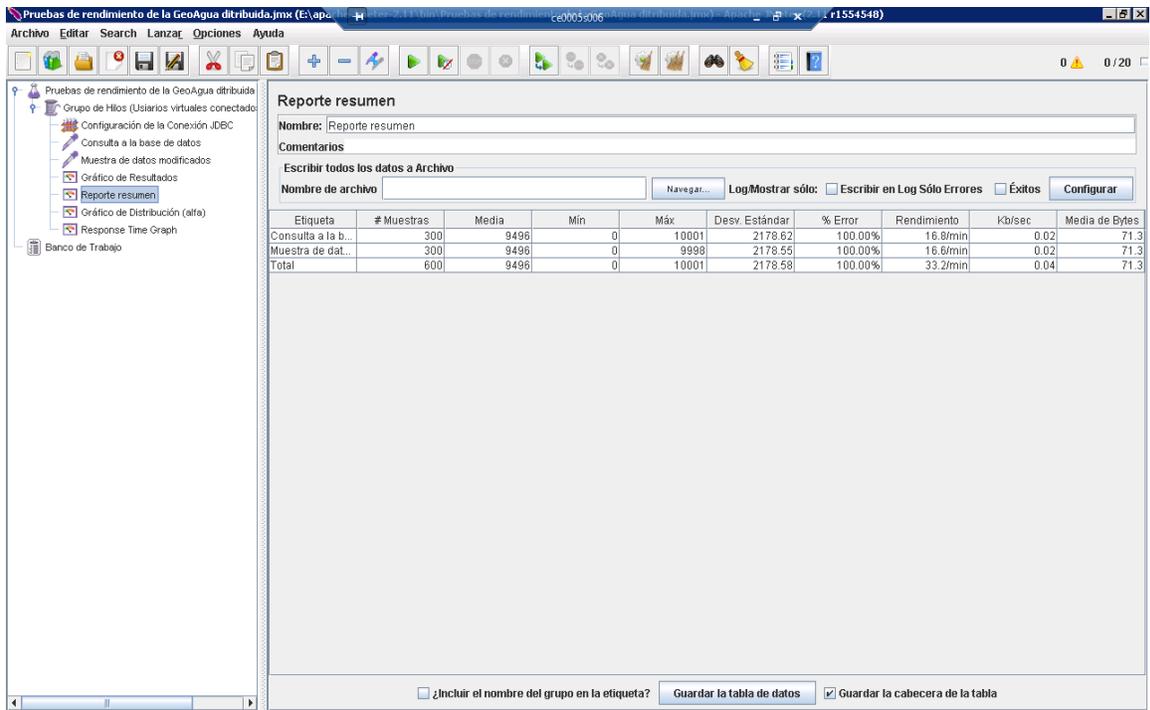
Así, el valor del intervalo es:

$$\left[4998 - \frac{(1.96)2178.62}{\sqrt{300}}, 4998 + \frac{(1.96)2178.62}{\sqrt{300}} \right] = [4751, 5425]$$

Por lo tanto, se puede esperar que el tiempo de respuesta promedio esté entre 4.7 y 5.4 segundos para una cantidad de 20 usuarios simultáneos realizando 300 solicitudes.

5.1.2. Pruebas de estrés

Subconjunto de pruebas de rendimiento utilizadas para evaluar el comportamiento de la Geobase de Datos más allá de las condiciones normales de carga o de pico, comparándolo con un sistema similar desarrollado con técnicas habituales. Se trata básicamente de probar la funcionalidad de la Geobase de Datos bajo altas cargas. Normalmente, esto está relacionado con problemas de sincronización o pérdidas de memoria. En nuestro caso se repiten las pruebas anteriores solo que ahora se incrementa el número de usuarios de 100 en 100 conforme a lo indicado en la Gráfica 4.



Reporte resumen

Nombre: Reporte resumen

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo: Navegar... Log/Mostrar sólo: Escribir en Log Sólo Errores Éxitos Configurar

Etiqueta	# Muestras	Media	Min	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Media de Bytes
Consulta a la b...	300	9496	0	10001	2178.62	100.00%	16.8/min	0.02	71.3
Muestra de dat...	300	9496	0	9998	2178.55	100.00%	16.6/min	0.02	71.3
Total	600	9496	0	10001	2178.58	100.00%	33.2/min	0.04	71.3

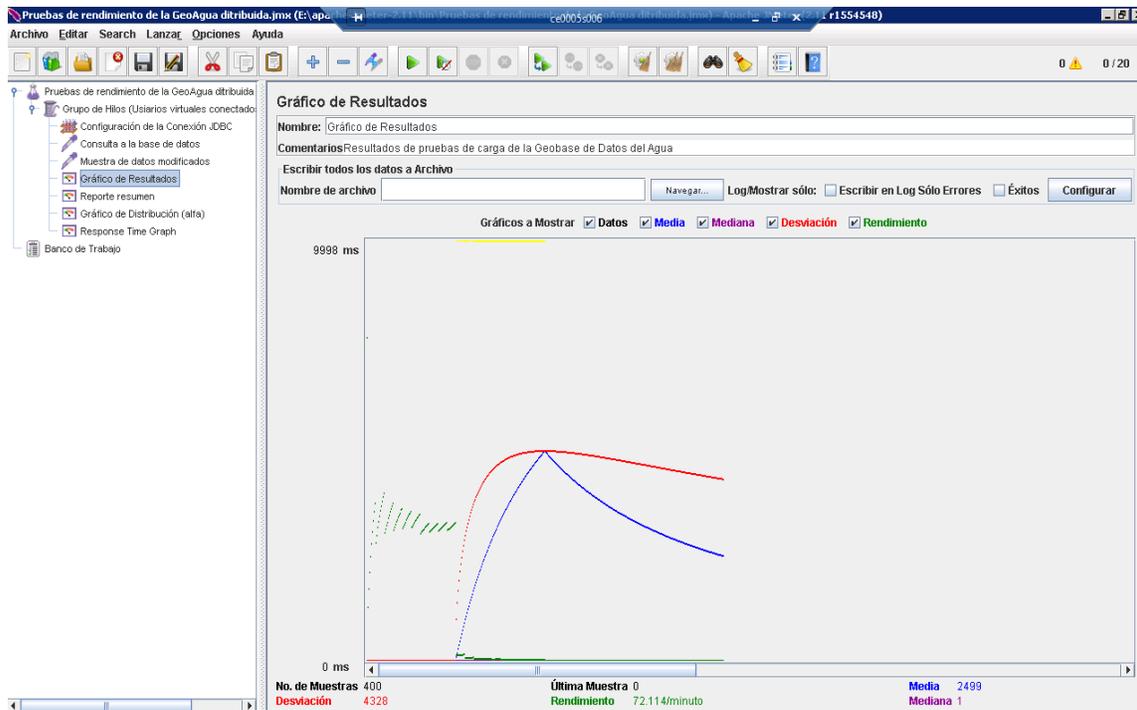
¿Incluir el nombre del grupo en la etiqueta? Guardar la cabecera de la tabla

Gráfica 4. Parámetros de conexión a la geobase de datos con un número máximo de 100 usuarios

En segundo lugar se evaluó el resultado obtenido a través de un intervalo de confianza para una distribución Normal al 95% de confianza para una prueba con grupos de 100 usuarios, conforme a los datos presentados en la Gráfica 6. La ecuación del mismo es la siguiente:

$$\left[9496 - \frac{(1.96)2178.62}{\sqrt{100}}, 9496 + \frac{(1.96)2178.62}{\sqrt{100}} \right] = [9068, 9923]$$

Por lo tanto, se puede esperar que el tiempo de respuesta promedio esté entre 9.0 y 9.9 segundos para una cantidad de 100 usuarios simultáneos realizando 100 solicitudes.

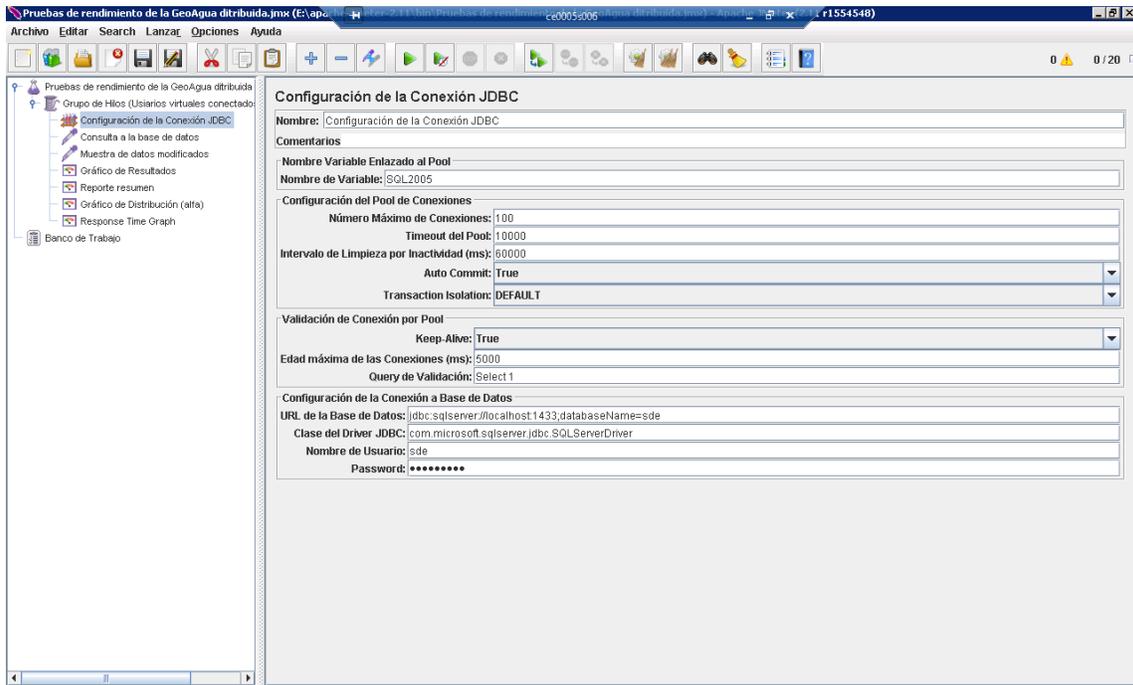


Gráfica 5. Resultados de rendimiento para 100 usuarios

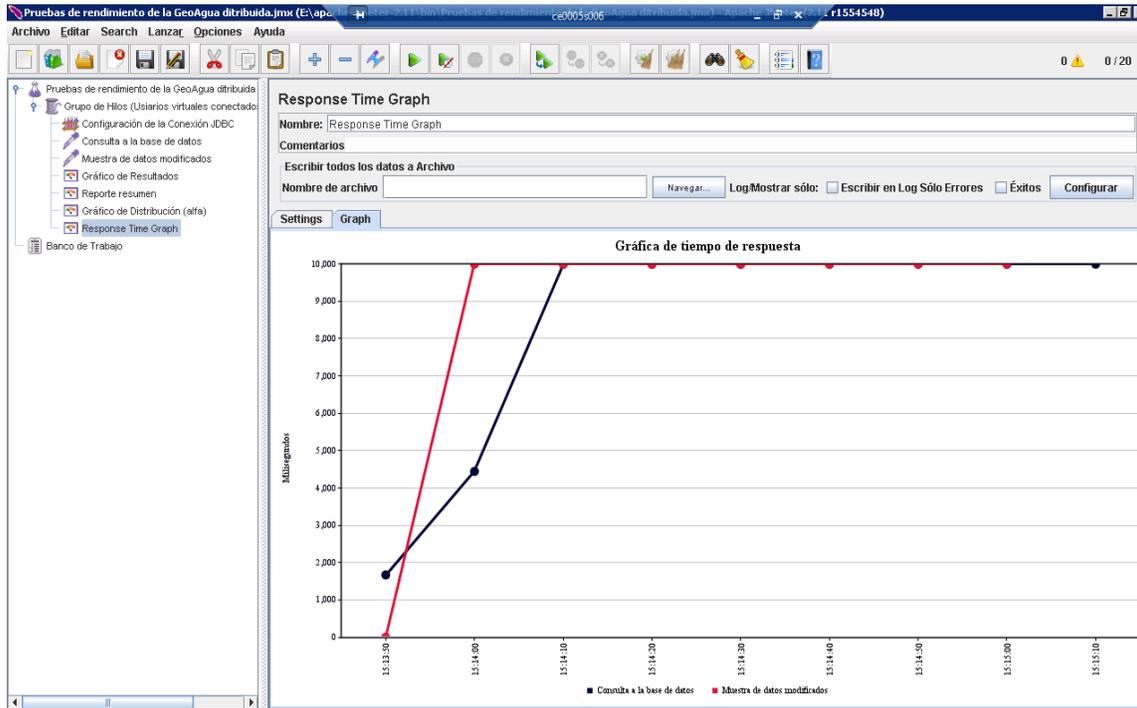
En la Gráfica 5 se presentan los resultados de rendimiento para 100 usuarios. Cabe señalar que conforme se observa el tiempo de respuesta promedio está entre 9.0 y 9.9 segundos para una cantidad de 100 usuarios simultáneos realizando 100 solicitudes.

Los valores de comparación de los tiempos de respuesta entre sistema estándar y el propuesto se muestran en la Gráfica 7, en donde se observa que el sistema propuesto presenta una mayor estabilidad con el tiempo y que, además, los tiempos de trabajo picos es factible superarlos ampliamente derivado de la distribución de cargas. Cabe mencionar que el tiempo de respuesta inicial del sistema propuesto es mayor; empero al evolucionar en el tiempo se logra un mejor desempeño. Es de mencionar que en cuanto a los resultados de carga en la Geobase de Datos para una muestra de 9996 elementos el tiempo de throughput (rendimiento) de 54.85/minuto para una carga de laboratorio con una

simulación de 50 usuarios y 5 ciclos. Se obtuvieron resultados estáticos para casos mayores a 100 usuarios con lo que no se presentaron mejoras. Esto puede deberse a la arquitectura de los propios servidores sobre los que está instalada la Geobase de Datos los cuales tienen, al menos, 10 años de uso lo cual induce una total obsolescencia.



Gráfica 6. Reporte resumen de número de muestras y pruebas realizadas



Gráfica 7. Tiempo de respuesta entre sistema estándar y el propuesto.

En la Tabla 7 se muestra el análisis del rendimiento esperado del sistema propuesto como un elemento preliminar a los resultados obtenidos mediante el uso de JMeter.

Tabla 7. Análisis del rendimiento esperado del sistema propuesto

Sistema utilizado	Ventajas	Desventajas
Sistema propuesto	<ol style="list-style-type: none"> Administración del acervo cartográfico digital más eficiente pues permite reconstruir la geobase de datos en su totalidad a partir de los fragmentos que se encuentran en Organismos de Cuenca (OC) y viceversa. Servicios de consulta y visualización de datos georreferenciados operando con mejor tiempo de repuesta debido a que al estar operando en un sistema distribuido los tiempos por retraso en las comunicaciones ya no se aplican. 	<ol style="list-style-type: none"> Falta de personal en los OC quienes deberán administrar, localmente su geobase de datos. Se contaría con una red “distribuida” de servidores geoespaciales vinculados a la Geobase de Datos central; sin embargo esto será posible hasta que se realice la migración a servidores físicos conforme se cuente con el personal necesario para su operación a fin de garantizar una alta

	<p>3. En caso de no poder operar en forma física, se pueden seguir trabajando en un entorno virtual desde la unidad central.</p> <p>4. Uso de estándares tanto propietarios como de Open Source.</p> <p>5. Mediante el aprovechamiento de las mejores prácticas de virtualización de geoservidores y geobases de datos distribuidas, la CONAGUA será capaz de alcanzar un intercambio de datos mejorado, así como un acceso integrado a sistemas y aplicaciones. El resultado, será el empleo más eficaz de la información geoespacial y alfanumérica dentro de la propia Institución.</p> <p>6. Reduce el tiempo de configuración¹²⁸ y puesta a punto de servidores nuevos.</p> <p>7. Reduce la necesidad de nuevos servidores así como el número de los servidores existentes mejorando el uso de estos últimos; es decir, permite la implementación de una estrategia de consolidación de servidores.</p> <p>8. Reduce los costos de propiedad tales como consumo de energía eléctrica, enfriamiento, espacio y mantenimiento; sólo por mencionar algunos.</p> <p>9. Prever el impacto de una aplicación sobre otra cuando se efectúa alguna actualización o cambio de plataforma, por ejemplo, permite correr diferentes versiones de ArcGIS sobre el mismo servidor físico.</p> <p>10. Incrementa la continuidad de la operación de los servicios a través de la reducción del tiempo</p>	<p>disponibilidad de servicios.</p> <p>3. Desarrollo de una descripción de las actividades de los Organismos de Cuenca. Dependiendo de la escala del estudio, puede ser necesario derivar varios modelos de actividad a distintos niveles de resolución, con objeto de describir completamente las necesidades de información de cada OC.</p> <p>4. Para una estructura de organización en particular, definir las reglas de dirección en términos de la responsabilidad de toma de decisiones sobre las actividades que un área operativa existente tenga asignadas.</p> <p>5. Se deberá definir el patrón de flujo mínimo de información, esto es, quien es responsable de abastecer cierta información y para quien.</p>
--	---	---

¹²⁸ Configuración significa la instalación y adecuación del sistema operativo y determinadas aplicaciones, basadas en requisitos organizacionales, para el óptimo desempeño del servidor al tiempo de ejecución.

	<p>de inactividad mediante la recuperación rápida de interrupciones no planificadas con la capacidad de copia de seguridad y migración de todos los entornos virtuales sin interrupción.</p> <p>11. El formato georrelacional tiene una muy buena escalabilidad y una estructura de datos dual¹²⁹ lo cual implica que la Geobase de Datos puede tomar completa ventaja de las características de las RDBMS.</p> <p>12. El sistema garantiza la interoperabilidad entre aplicaciones GIS tanto en Open Source (OpenGIS Simple Feature Specification) como en plataforma ArcGIS.</p> <p>13. Los servicios de Web conforman todo una red de nodos distribuidos los cuales pueden incluir servidores, estaciones de trabajo, clientes del escritorio y “clientes ligeros”¹³⁰. Los servicios Web implican normas que proporcionan los elementos para que estos dispositivos actúen recíprocamente para formar una red informática de mayor entorno.</p> <p>14. Un SBDD¹³¹ implica un conjunto de programas que operan en diversas computadoras; así que se debe permitir que cada sitio almacene y mantenga su propia base de datos, facilitando de esta forma el acceso inmediato y eficaz a sus datos, mejorando con ello la fiabilidad si la computadora de un sitio no funciona, el resto de la red seguirá operando.</p>	
--	---	--

¹²⁹ Características espaciales almacenadas en un formato basado en archivos con atributos almacenados en una base de datos relacional.

¹³⁰ Teléfonos celulares, PDAs, GPS, Palms, etc.

¹³¹ Sistema de Bases de Datos Distribuidas

	<p>15. Capacidad de ejecutar diferentes versiones de ArcGIS Server en una única máquina física. Esto es especialmente útil en entornos de desarrollo, así como la aplicación de estrategias de migración de versiones.</p> <p>16. El cumplimiento con el plan de la CONAGUA ante recuperación de desastres requieren de aplicaciones ArcGIS Server y servicios que puedan ser rápidamente configurados en el caso de una catástrofe. Esto puede ser efectuado mediante un desarrollo acelerado y el despliegue de equipo autónomo de las aplicaciones de ArcGIS Server.</p>	
<p>Sistema similar desarrollado con técnicas habituales</p>	<p>1. El sistema recibe los datos desde cualquier RDBMS compatible o a partir de ODBC o ADODC.</p> <p>2. Tienen la capacidad de realizar procesamiento autónomo, esto permite realizar operaciones locales o distribuidas.</p> <p>3. Un SBDD implica un conjunto de programas que operan en diversas computadoras; así que se debe permitir que cada sitio almacene y mantenga su propia base de datos, facilitando de esta forma el acceso inmediato y eficaz a sus datos, mejorando con ello la fiabilidad si la computadora de un sitio no funciona, el resto de la red seguirá operando.</p> <p>4. El objetivo en el manejo de transacciones es por un lado tener una transparencia</p>	<p>1. El sistema no garantiza la interoperabilidad entre aplicaciones GIS tanto en Open Source como en plataforma ArcGIS.</p> <p>2. Usan sistemas de archivo¹³² para el almacenamiento y procesamiento de datos.</p> <p>3. Un sistema distribuido de bases de datos se almacena en varias computadoras; lo cual implica que existe un administrador del sistema responsable de la seguridad global y, a su vez, en cada una de las otras bases de datos hay un administrador local quien llega a un nivel de autonomía local.</p>

¹³² Estos sistemas de archivos se refieren al uso de archivos tipo "shape" el cual, a pesar de ser un estándar, no permite la facilidad de acceso hacia un banco de datos geográficos de gran tamaño, tampoco permite trabajar con "versiones" de la misma información y no propicia el trabajo participativo entre otros inconvenientes.

	<p>adecuada de las acciones concurrentes a una base de datos y, por otro, tener una transparencia adecuada en el manejo de las fallas que se pueden presentar en una base de datos.</p>	
--	---	--

5.2. Sobre la geobase de datos

Enseguida se presenta el universo de la información así como los métodos de revisión y validación requeridos para cada tipo de información e implementados en las ETL (Extracción-Transformación-Carga) para evitar que la información se corrompa y se introduzca basura. Indique cuáles serían los mejores indicadores de que la geobase está funcionando adecuadamente.

Descripción del Universo de Información.

La GeoAgua está estructurada para almacenar información tanto de tipo vectorial como raster. Sin embargo se ha observado que al trabajar con información de tipo raster se presentan algunos inconvenientes en cuanto al tamaño de los archivos; particularmente al almacenar archivos tipo LiDAR (Light Detection And Ranging – detección por luz y distancia)¹³³ los cuales requieren de una gran capacidad de almacenamiento y procesamiento; por lo que se decidió almacenar la información raster en un conjunto cluster de dos servidores especializado tipo blade operando sobre LUNS¹³⁴ en forma independiente de los dos servidores en clúster para el almacenamiento de datos vectoriales los cuales también operan con unidades LUN. En particular se cuenta con la siguiente información:

INFORMACIÓN TIPO RASTER	INFORMACIÓN TIPO VECTOR
<p>Imágenes de satélite:</p> <ol style="list-style-type: none"> 1. ASTER GDEM 2. LADSAT NASA (2000, 2002) 3. RAPIDEYE (Río Mayo, Tabasco, Acapulco 2013) 	<p>Ciclo hidrológico:</p> <ol style="list-style-type: none"> 1. Acuíferos 2006 2. Acuíferos 2009 3. Acuíferos Enero 2011 4. Acuíferos Enero 2012

¹³³ <http://www.inegi.org.mx/geo/contenidos/geodesia/lidar.aspx>

¹³⁴ Logical Unit Number es una dirección para una unidad de disco duro de almacenamiento y por extensión, el disco en sí mismo. El término proviene del protocolo SCSI como una forma de diferenciar unidades de disco individuales dentro de un bus SCSI con un arreglo de discos. El término es muy común en el ámbito de las Storage Area Networks (SAN) y en otros campos del almacenamiento corporativo. Un LUN no es normalmente un disco entero sino una partición virtual o un volumen dentro de un conjunto RAID.

<p>4. SPOT (2003, 2008, 2009, 2010, 2011)</p>	<p>5. Acuíferos Febrero 2013 6. Acuíferos Julio 2010 7. Cuencas hidrológicas Febrero 2014 8. Estaciones climatológicas marzo 2007 9. Humedales potenciales INEGI 10. Número de días con granizo 11. Número de días con lluvia 12. Número de días con niebla 13. Número de días con tormenta 14. Red hidrográfica (sin topología de red geométrica) 15. Ríos a nivel nacional escala 1:250,000 16. Sitios de muestro de calidad del agua (2006) 17. Ríos principales 18. Subregiones hidrológicas 19. Vedas (03-02-2010) 20. Vedas (10-06-2010)</p>
<p>INEGI</p> <p>1. Fisiografía. Provincias fisiográficas, Sistema de Topoformas, subprovincias fisiográficas</p>	<p>Curvas de nivel, a nivel nacional, escala 1:50,000 clasificadas conforme a las cartas topográficas de INEGI</p>
<p>LiDAR</p> <p>1. Acapulco 2. Chiapas 3. DF 4. Huixtla, Chiapas 5. Varias marzo 2011 (resolución de 5m) 6. Varias escala 50,000 7. Monterrey 8. Río Bravo (Cd. Acuña, Nuevo Laredo, Piedras Negras)</p>	<p>Contexto</p> <p>1. Densidad de población 2. Entidades federativas (polígonos) 3. Entidades federativas (líneas) 4. Localidades rurales puntos (2011) 5. Localidades urbanas (polígonos) 6. Localidades urbanas puntos (2011)</p>

<ul style="list-style-type: none"> 9. Tabasco (Llanura costera Golfo Sur) 10. Presa Falcón 11. Mexicali 12. Pesa la Amistad (CILA) 13. Tamaulipas 	<ul style="list-style-type: none"> 7. Marginación (2005) 8. Municipios 9. Organismos de cuenca (MGM 2010) 10. Organismos de Cuenca (10-05-2010) 11. Organismos de cuenca líneas 12. Zonas metropolitanas (2000)
<p>Varios:</p> <ul style="list-style-type: none"> 1. Cuencas GASIR 2. Lago de Texcoco 3. TEO 4. PTAR 5. Sequías 6. Mapas de inundación 7. Humedal Málaga 8. Ciénega de Lerma 9. Cuatro Ciénegas 10. Canal de la Compañía 11. Canal de Chalco 12. Coaxtla 13. Cutzamala 	<p>Escala 1:250,000:</p> <ul style="list-style-type: none"> 1. Cuencas INEGI 2. Cuerpos de Agua 3. Curvas de nivel 4. Edafología 5. Geología 6. Ríos 7. Toponimia 8. Carta de uso de suelos y vegetación SIIIg 9. Carta de uso de suelos y vegetación SIIIv 10. Carta de uso de suelos y vegetación SIVv 11. Veredas y brechas 12. Vías de transporte
<p>RAN:</p> <ul style="list-style-type: none"> 1. Guerrero 2. Hidalgo 3. Morelos 4. Nayarit 	<p>Escala 1:50,000:</p> <ul style="list-style-type: none"> 1. Cuerpos de agua INEGI 2. Nombres geográficos 3. Ríos clasificados por región hidrológica con topología de red geométrica. 4. Regiones hidrológicas INEGI 5. Subcuencas INEGI 6. Toponimia INEGI
<p>Cartas tipográficas 1:50,000 Cartas de uso de suelo (INEGI SI a SIV) Cartas hidrológicas</p>	<p>Impacto en la sociedad y el medio ambiente</p> <ul style="list-style-type: none"> 1. Comisiones de cuenca 2. Comités de cuenca

<p>Cartas topográficas 1:250,000 serie II Mapas de humedales 2012 Microcuencas FIRCO</p>	<p>3. COTAS 4. CRAE 5. Sitios RAMSAR</p>
<p>Modelos Digitales de Elevación y Curvas de Nivel (corregidas a partir de datos de INEGI 1:50,000):</p> <ol style="list-style-type: none"> 1. Por Dirección Local 2. Por Organismo de Cuenca 	<p>Índices cartográficos:</p> <ol style="list-style-type: none"> 1. Índice 250k 2. Índice 50k 3. Índice grado por grado 4. Índice Landsat 5. Índice por letra 6. Ortofotos INEGI 2004
<p>Ortofotos digitales:</p> <ol style="list-style-type: none"> 1. 1:50,000 2. 1:20,000 (SEMARNAT 2011) 3. 1:10,000 clasificadas por cartas topográficas de INEGI; en blanco y negro y color (procesadas). 4. 1:5,000 	<p>Mapas mundiales</p> <ol style="list-style-type: none"> 1. Agua potable 2. Países del mundo 3. Saneamiento
<p>Google Earth. Muy diversos proyectos.</p>	<p>Usos del Agua:</p> <ol style="list-style-type: none"> 1. Cobertura de agua potable y alcantarillado 1990 2. Cobertura de agua potable y alcantarillado 1995 3. Cobertura de agua potable y alcantarillado 2000 4. Cobertura de agua potable y alcantarillado 2005 5. Distritos de riego 6. Presas consultivo técnico 7. Zonas de disponibilidad (2007)
	<p>Red hidrográfica con topología de red geométrica implementada sobre ArcHydro por lo que cuenta con estructura y clasificación de ríos, canales, presas, estaciones climatológicas e hidrométricas a las cuales tiene asociadas series</p>

de tiempo para análisis hidrológico.

Adicionalmente la GeoAgua cuenta con un gran acervo documental de metadatos que hace referencia a estudios de aguas subterráneas así como metainformación de todas y cada una de las capas de información que se encuentran en la Geobase. Para su consulta se cuenta con el servicio del Metada Explorer¹³⁵ el cual, en el futuro próximo se migrará a un Geoportal.

5.2.1. Métodos de revisión y validación requeridos para cada tipo de información e implementados en las ETL (Extracción-Transformación-Carga)

Con fundamento en la definición pura de los procesos ETL los cuales son una parte esencial de la integración de datos, cuya función completa el resultado de todo el desarrollo de la cohesión de aplicaciones y sistemas; los procesos de extracción, Transformación y Carga de datos geospaciales se efectúan de la siguiente forma.

FASE DE EXTRACCIÓN

- Extracción de los datos desde la fuente o sistema de origen. Esto puede referirse a información en unidades de almacenamiento de alto volumen como NAS-SAN hasta información en línea o almacenada en DVD o memorias USB.
- Análisis de los datos extraídos efectuando un chequeo de los siguientes elementos:
 - Revisión, corrección de errores geométricos y topológicos. Nodos de polígonos abiertos o colgantes, polígonos cruzados o superpuestos, líneas colgantes, tolerancia clúster, eliminación de huecos, etc. Esto se hace a través de la herramienta de corrección de errores topológicos.
 - Revisión y corrección de ajustes espaciales. Las inconsistencias entre las fuentes de datos en ocasiones requieren que se realice trabajo adicional para integrar un nuevo dataset con el resto de los datos. Algunos datos se distorsionan o rotan geoméricamente con respecto a los datos base. Dentro del entorno de edición, las herramientas de ajuste espacial ofrecen métodos interactivos para alinear e integrar los datos. El ajuste espacial admite una variedad de métodos de ajuste y ajustarán todas las fuentes de datos

¹³⁵ <http://siga.cna.gob.mx/metadataexplorer/>

editables. Generalmente se utiliza cuando ha importado datos desde otra fuente, como el dibujo CAD. Algunas de las tareas que puede realizar incluyen la conversión de datos de un sistema de coordenadas a otro, la corrección de distorsiones geométricas, la alineación de entidades a lo largo del borde de una capa en las entidades de una capa contigua y el copiado de atributos entre capas. Debido a que el ajuste espacial opera dentro de una sesión de edición, puede utilizar la funcionalidad de edición existente, como la alineación, para mejorar los ajustes.

- Revisión, corrección de errores de atributos. Agrega, elimina o actualiza valores de atributo La ventana atributos o bien a través de la ventana de tabla de atributos permiten visualizar y editar los atributos de las entidades que se ha seleccionado. En este caso se pueden corregir errores de sintaxis, ortografía, semántica o paráfrasis.
- Interpretar este chequeo para verificar que los datos extraídos cumplen la pauta o estructura que se esperaba. Si no fuese así, los datos se rechazan.
- Convertir los datos a un formato preparado para iniciar el proceso de transformación. En este paso se realizan cambios en cuanto al sistema de coordenadas requerido y la reproyección de la capa de información en caso de ser necesario.
- Una de las prevenciones más importantes que se deben tener en cuenta durante el proceso de extracción es el exigir siempre que esta tarea cause un impacto mínimo en el sistema de origen. Este requisito está basado en la práctica puesto que, si los datos a extraer son muchos, el sistema de origen se podría ralentizar e incluso colapsar, provocando que no pudiera volver a ser utilizado con normalidad para su uso cotidiano.

FASE DE TRANSFORMACIÓN

En la fase de transformación del proceso de ETL se aplican una serie de reglas de negocio o funciones, sobre los datos extraídos para convertirlos en datos que serán cargados en la Geobase de Datos. Estas directrices deben basarse en excepciones o restricciones y a fin de potenciar su pragmatismo y eficacia, hay que asegurarse de que sean:

- Declarativas.
- Independientes.
- Claras.
- Inteligibles.
- Con una finalidad útil para el negocio.

FASE DE CARGA

En esta fase, los datos procedentes del paso anterior son cargados en la Geobase de Datos. En esencia existen dos formas básicas de desarrollar este proceso:

- Acumulación simple. Consiste en realizar un resumen de todas las transacciones comprendidas en el período seleccionado y transportar el resultado como una única transacción hacia la Geobase de Datos. Es la forma más sencilla y común de llevar a cabo el proceso de carga.
- Rolling: este proceso es recomendable en los casos en que se requiere mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud almacenada.

Sea cual sea la manera elegida de desarrollar este proceso, hay que tener en cuenta que esta fase interactúa directamente con la Geobase de Datos del Agua y, por eso, al realizar esta operación se aplicarán todas las restricciones que se hayan definido en ésta. Si están bien definidas, la calidad de los datos en el proceso ETL estará garantizada.

Cabe aclarar que los procesos antes referidos son efectuados tanto para integración, en la Geobase de Datos, de datos vectoriales como de raster y/o tablas o relaciones.

5.2.2. Mejores indicadores de que la geobase está funcionando adecuadamente.

Una de las funciones indispensables de la administración de la Geobase de Datos Institucional del Agua (GeoAgua) es el monitoreo del rendimiento de los servidores de bases de datos. Para ello, a fin de definir indicadores idóneos, se

hace uso de los contadores de rendimiento¹³⁶ los cuales dan medida de numerosos parámetros. Existen muchísimos contadores que permiten la monitorización, tanto del sistema operativo como propios de SQL Server sin embargo solo se hará mención de los que, a juicio del autor son los más recomendables.

El sistema operativo de los servidores donde está instalada la GeoAgua, cuenta con una herramienta denominada Performance Monitor. Esta herramienta permite ver los contadores en directo. Sin embargo, la forma más saludable es arrancar los contadores, acumular datos y luego estudiarlos a detalle. Lo normal es realizar una captura cada cierta cantidad de minutos (1, 2 o 5) de varios contadores. Esto permite realizar estudios de tendencias, identificar picos así como valles de carga durante el día pero, sobre todo, para no tener que poner contadores cuando se reporte un problema de rendimiento, momento en el que quizás ya sea demasiado tarde. Windows tiene su propia herramienta de registro y alertas de rendimiento. Se pueden configurar un conjunto de contadores, una frecuencia de los mismos, así como un formato de salida para los datos y luego empezar a recoger datos. El formato de los datos puede ser binario, texto plano e incluso pueden almacenarse en una base de datos, a través de un DNS. Esto nos abre la puerta a una cantidad enorme de posibilidades cuando se requiere explotar la información recogida.

En ambos casos, ya sea en directo o acumulando datos, no es bueno recoger estos datos desde el propio servidor que se desea monitorear. Esto no es sólo por la alteración en los datos recogidos por la toma de los propios contadores, sino por el impacto en el rendimiento general que se puede causar.

INDICADORES

En general se deben monitorear el propio servidor a nivel de hardware (memoria, procesador, red y disco), para localizar problemas propios del desarrollo y particularidades del sistema. En cuanto a la Geobase de Datos en sí misma es necesario analizar contadores de objetos propios de SQL Server. Es

¹³⁶ Un contador, en general, es una medida de un parámetro determinado. Cada contador pertenece a un objeto. Por objeto, se entiende un ámbito de medición, existiendo muchos de estos objetos. Cada objeto tiene uno o más contadores y es necesario que se instale o active el objeto para que podamos recoger los datos de sus contadores. Hay objetos que se instalan con el sistema operativo, como el objeto "system"; otros hay que instalarlos, como los de red; y otros los instalan las aplicaciones, como los de SQL Server. También hay contadores que sólo están disponibles bajo determinadas circunstancias, como que un servicio concreto esté arrancado o se esté realizando un backup. Por último, existen contadores que permiten la medición de diferentes instancias o partes concretas o del total de las mismas. Por ejemplo, los contadores que miden el número de transacciones, los cuales permiten recoger el número de transacciones de cada base de datos independientemente.

fundamental señalar que aunque aparentemente el análisis está dividido en áreas, todas estas están relacionadas. Un problema de contención de disco suele repercutir en el CPU, podría estar causado por una mala gestión de la memoria y, finalmente, saturar la red. Es muy importante tener en cuenta estas interrelaciones al momento de realizar un diagnóstico certero que permita localizar el origen, o los orígenes, del déficit de rendimiento.

MEMORIA

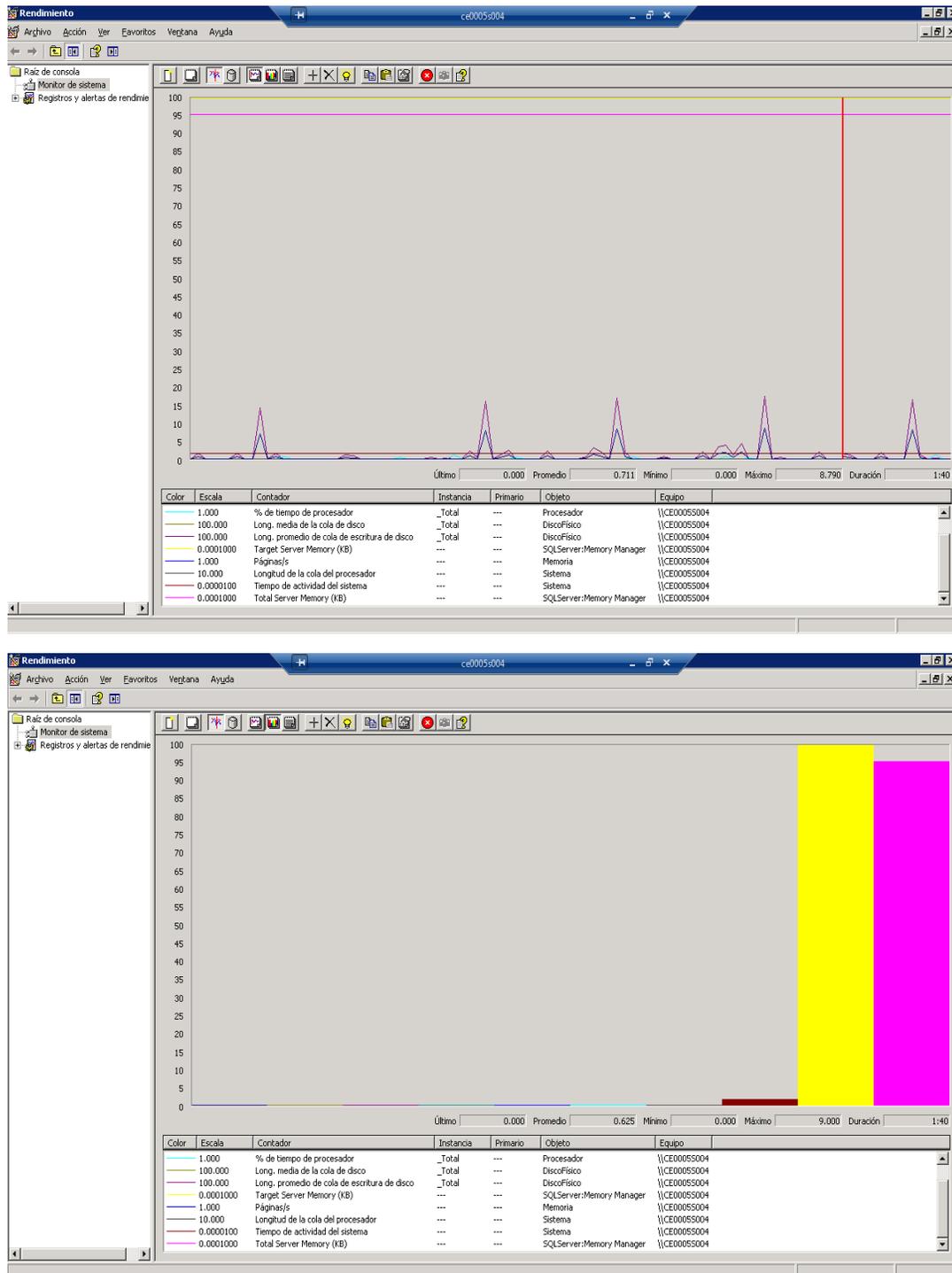
El correcto uso de la memoria es lo que brindará un servidor más ágil en su respuesta. Con relación a la operación de la GeoAgua SQL Server debe tener suficiente memoria disponible como para emplear un poco más en caso de hacer falta; pero a la vez, debe ocupar la suficiente memoria como para evitar que las operaciones se realicen en el área de swap del disco físico. Si la mejor consulta es aquella que no se realiza, la segunda mejor es aquella que se realiza contra la memoria caché.

Memory Pages/sec. Indica el número de páginas que entran y salen de la memoria caché en cada segundo. Su valor debe situarse muy cercano a 0. Si es mayor a 20 de forma continua, tal vez no exista un problema de rendimiento; pero lo que es seguro, es que la memoria no está siendo gestionada adecuadamente.

Memory: Availability Mb (Kb ó Bytes). En general, se debe contar con 50 MB de memoria libres, y disponibles, como mínimo.

SQL Server: Memory Manager: Total Server Memory y Target Server Memory. Estos dos contadores, del mismo objeto, indican el total de memoria con la que se cuenta así como la memoria que se necesita. El valor del parámetro “Total” debe ser igual que “Target”; de no ser así, y “Total” es menor que “Target”, es un indicio claro de un problema en la memoria, ya que se tendría menos de la que se necesita.

En la Gráfica 8 se presentan dos ejemplos de la medición de los parámetros señalados en los párrafos anteriores para el caso particular de la GeoAgua. La línea roja representa la evolución temporal de todos los parámetros en medición.



Gráfica 8. Gráficas de medición de indicadores de operación de la Geobase de Datos.

PROCESADOR

El uso del procesador es un punto básico para el monitoreo de cualquier servidor, sea o no de base de datos. Dentro de los parámetros que se pueden consultar se recomienda poner atención en dos de los más importantes: el porcentaje de uso y la cola del procesador, más un tercero que permita conocer que cantidad de memoria es utilizada por el sistema operativo. Cuanto menor sea la utilización de los CPU que se tiene en el server, mucho mejor. Este uso comienza a ser un problema cuando de forma sostenida se sobrepasa el 80%. Para solventarlo, antes de pensar en añadir más procesadores y/o cambiarlos por otros más veloces o tecnológicamente más avanzados, es preciso revisar las re-compilaciones y los planes de ejecución de las consultas, entre otras cosas. Luego, para monitorizar la CPU se puede optar por los siguientes contadores, los cuales se muestran en la Gráfica 9:

Processor % CPU Usage (% de tiempo de procesador) (instancia _Total si se cuenta con más de un procesador y si todos ellos están desempeñando el mismo rol): Este parámetro se debe mantener por debajo del 80%.

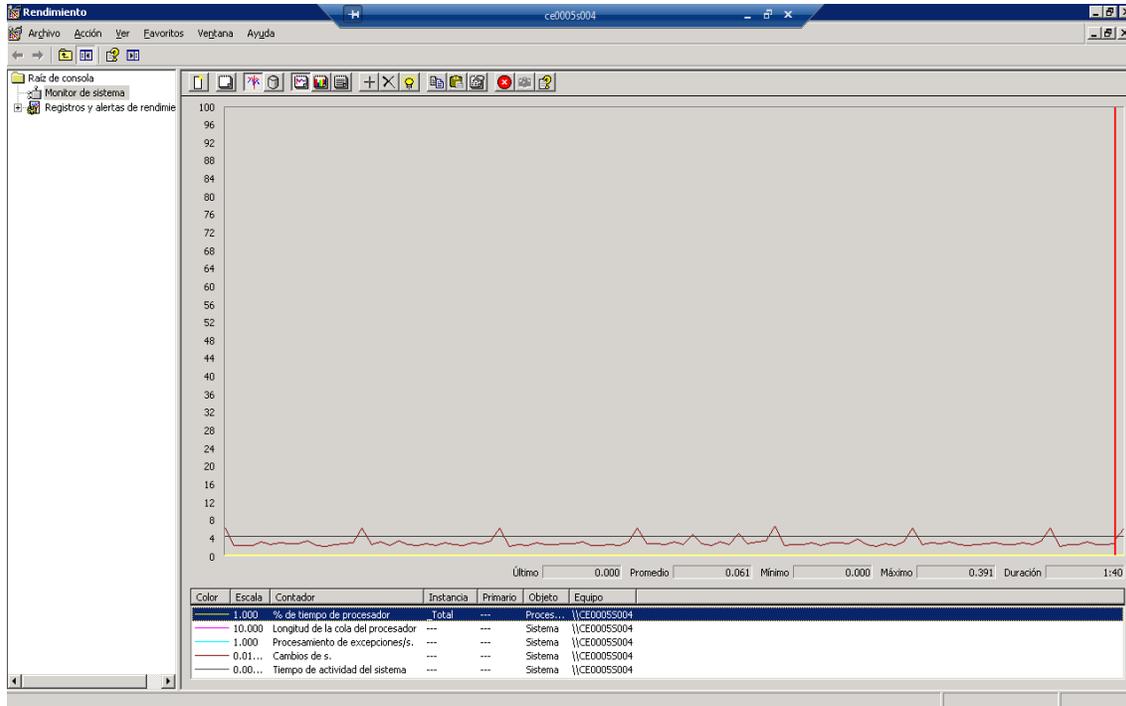
System Processor queue length (Longitud de la cola del procesador). Es la cola de procesador, debe permanecer por debajo de 2 por CPU. En la Gráfica 9 se muestra la medición de estos dos parámetros.

Si estos dos contadores están por encima de lo normal, es necesario observar muy detalladamente un par de parámetros adicionales para afinar el diagnóstico. Uno es **System: Context Switches/sec** (sólo en caso de tener más de una CPU). Este contador indica las veces en las que un mismo proceso, medianamente pesado, cambia de procesador para completarse. El cambio tiene lugar por cuestiones de balanceo de carga y aunque puede forzarse a que no se efectúe, lo más lógico es dejar que sea el servidor el que gestione estos saltos. Pero cada movimiento tiene un costo el cual puede verse reducido si configuramos el servidor para que use fibras¹³⁷, que suavizan de forma considerable el costo del cambio entre procesadores. Si este contador se sitúa por encima de 8000, es momento de pensar en modificar esta configuración y muy posible de server.

El otro contador que puede aportar información adicional es **System: % Total Privileged Time** éste contador indica qué porcentaje de tiempo se dedica a tareas del sistema operativo. Si su valor está por encima del 20%, es posible que el problema no esté en el procesador, sino en el disco; para confirmarlo será

¹³⁷ Se pueden definir como conjuntos de hilos o hilos gruesos

necesario verificar si el contador **PhysicalDisk: % Disk Time** está por encima del 55%.



Gráfica 9. Porcentaje de uso del procesador y la cola del procesador y cantidad de memoria que es utilizada por el sistema operativo

DISCO

El disco es el punto en el que más frecuentemente se localizan los cuellos de botella, así como la causa final de los problemas. Discos potentes y rápidos, una buena configuración en el RAID que asegure la disponibilidad, y una ubicación correcta de los archivos en los diferentes grupos de discos es la clave para que los servidores rindan a plena potencia. Aún con todo esto, pueden presentarse problemas de contención, que con una adecuada monitorización se podrá detectar; es más, la monitorización nos dará pistas para paliar la situación. Así, además de saber si cada conjunto de discos va bien o mal, se puede determinar si el problema está en las lecturas o en las escrituras, si cambiar el fill factor de los índices ayudaría.

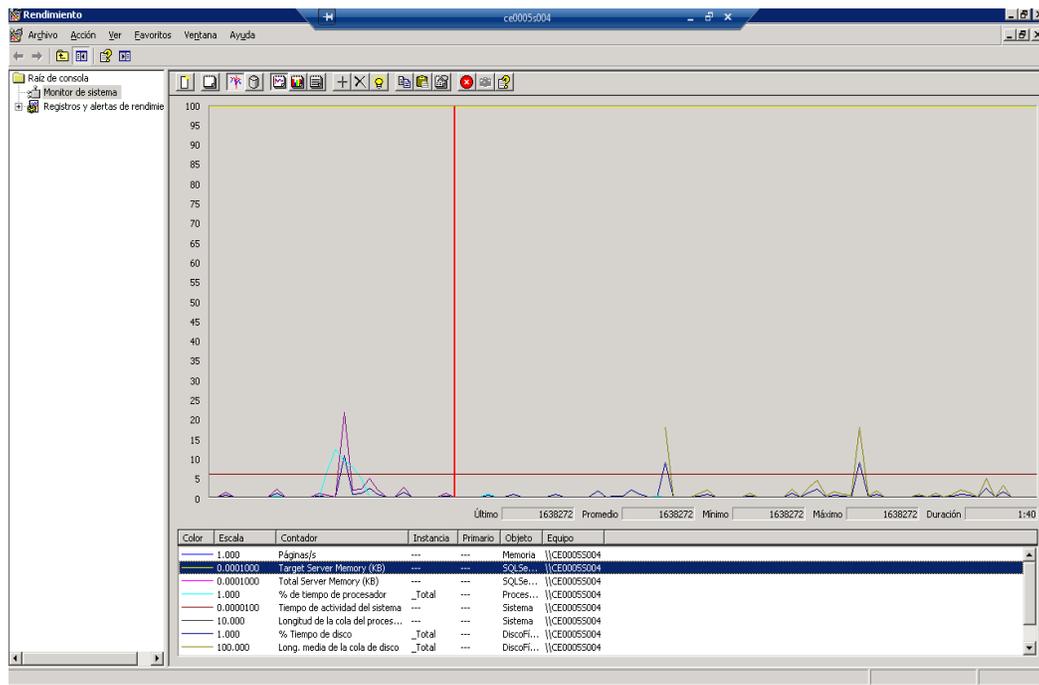
Para poder observar los contadores de disco físico, es necesario que éstos estén activos.

El contador más relevante, indicado en la Gráfica 10, es el de la cola de disco ya sea la media o la actual. Si durante períodos prolongados (más de 10 minutos) se mantiene por encima de 2, se puede decir que existe un problema, aunque

teniendo en cuenta dos detalles importantes: primero, el contador funciona a nivel lógico. Si por ejemplo se cuenta con una unidad en RAID 5 compuesta por 5 discos, será necesario dividir el valor del contador entre 5; y segundo, por sentido común, es necesario permitir determinados procesos, sean o no de mantenimiento. La realización de backup y las re-indexaciones son operaciones que ocupan intensamente el disco. En caso de existir muchos de estos procesos la forma de proceder será encontrar un momento de baja carga para la ejecución de estos, como durante las noches o los fines de semana.

PhysicalDisk Avg. Disk Queue Length. Debe estar por debajo de 2 en cada unidad, tras ponderar el número de discos del RAID. La instancia común puede ser de ayuda para calibrar el estado general.

PhysicalDisk % Disk Time. Indica qué porcentaje de tiempo se emplea en el disco. Si está por encima del 55%, puede que haya un problema, habría que mirar también **PhysicalDisk: Disk Read Time** y **PhysicalDisk: Disk Write Time** para ver si hay un importante desequilibrio no esperado entre las lecturas y las escrituras, que podría regularse variando el fill factor de los índices. En una base de datos de sólo lectura o eminentemente de lectura, es lógico que haya un desequilibrio; si las lecturas están muy por encima de las escrituras, el fill factor puede que sea muy alto. Si ocurre lo contrario y las escrituras llevan la mayor parte del peso, el aumentar el fill factor podría paliar el problema. Es más una cuestión de afinar y probar hasta encontrar un equilibrio.



Gráfica 10. Para el monitoreo habitual los parámetros de desempeño se encuentran dentro de límites aceptables.

RED

Los objetos que permiten colocar contadores de red son aquellos que empiezan por **Network**, concretamente en las llamadas **Herramientas de administración y supervisión**.

La red en ocasiones puede ser un cuello de botella y es imposible no tenerla en cuenta, ya que los datos van y vienen por allí. No es infrecuente que se produzcan cortes en la comunicación, que pueden ser detectados por métodos tan simples como poniendo una estadística de ping con paquetes ligeros. La aplicación de las normas básicas de la tecnología cliente-servidor aporta la mejor forma de optimizar el uso de la red. Es decir, enviar sólo aquello que el cliente pide e impedir que el cliente pida más de lo que necesita con una correcta paginación, parámetros limitados en los reportes, búsquedas controladas, etcétera.

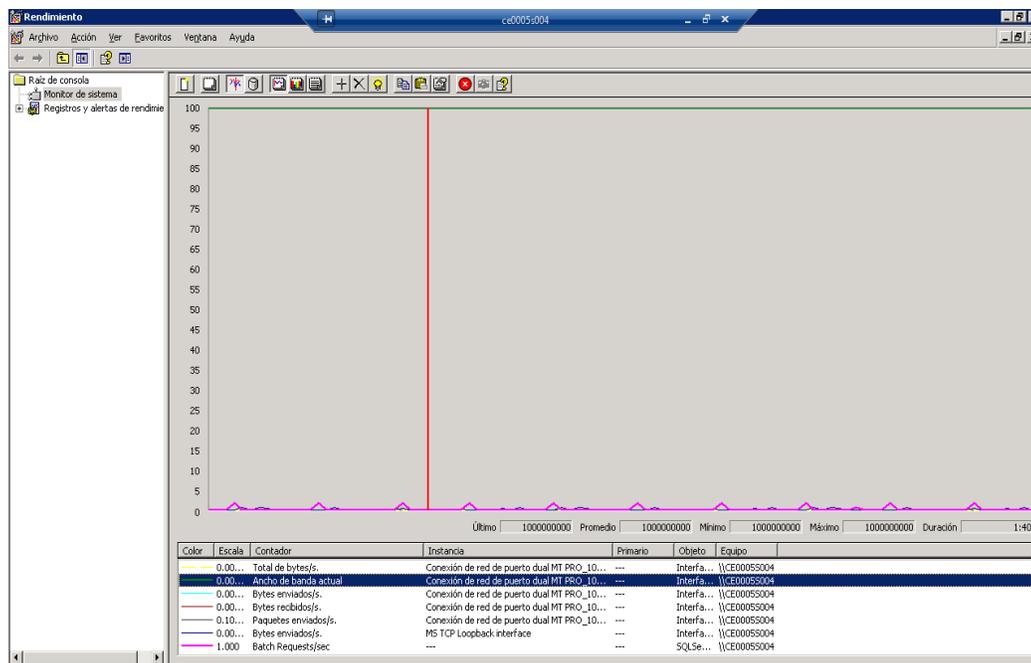
La red debe estar dimensionada acorde con lo que se le va a exigir, para lo cual se cuenta con la ayuda de los siguientes contadores:

- **Network Interface Bytes Total/sec.** Depende esencialmente de la red, con lo que es difícil dar una cifra que pueda ser usada de forma general. Se puede aplicar una sencilla regla, en combinación con el contador

Current Bandwidth. El objeto **Bytes Total/sec** dividido entre **Current Bandwidth** debe ser menor que 6.

- **Network Segment % Network Utilization.** Permite verificar el uso de cada tarjeta de red con las que cuente el servidor.
- **Server Bytes Received/sec y Server: Bytes Transmitted/sec.** Permiten comprobar si es el servidor de base de datos es el que está saturando la red y perjudicando sus otros usos.
- **SQLServer SQL Statistics Batch Request/sec.** Una tarjeta de red de 100Mbs soporta, aproximadamente unos 3000 comandos por segundo. Si este contador está por encima, se precisa una segunda tarjeta o una de mayor capacidad.

Los contadores señalados en los párrafos anteriores se muestran en la Gráfica 11.



Gráfica 11. Medición de indicadores de rendimiento de la interface de red (NIC)

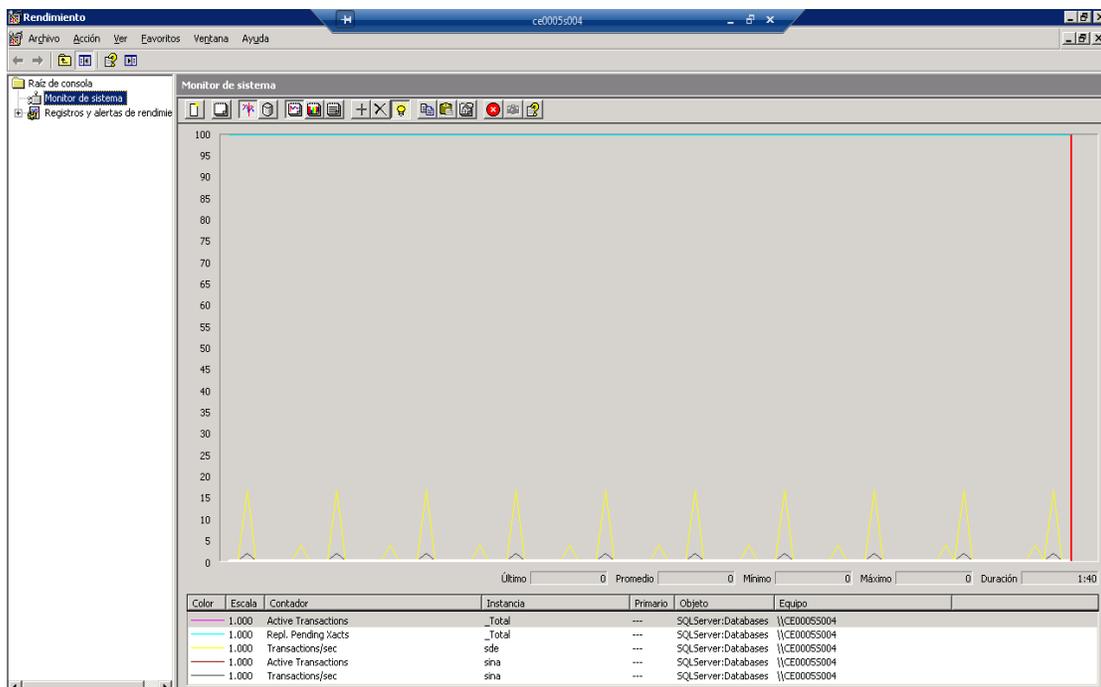
SQL SERVER

SQL Server posee un importante número de objetos, con sus contadores, que permiten monitorear a detalle lo que se requiera en un servidor. Tenemos objetos para SQL Server Agent, replicación, bloqueos, memoria, caché, conexiones, paginación, backup, etc. No hay más que ir al monitor de rendimiento y en la lista de objetos ver todos los que empiezan con **SQLServer**. Aquí se citarán algunos de ellos, los más generales quizás. Pero en función del

uso que tengamos en cada servidor deberemos ampliar o cambiar los contadores a seguir.

Los siguientes son algunos de los contadores de uso más extendido, intentando hacer un barrido por lo más significativo para conseguir una monitorización básica:

- **SQLServer: Access Methods: Page split/sec.** Si está por encima de 100, viene acompañado de problemas de disco. Un aumento en el fill factor puede resolver la situación.
- **SQLServer: Buffer Manager: Cache Hit Ratio.** Indica el porcentaje de veces que el motor usa la caché frente al disco. Es un valor medio desde el último reinicio. Este valor debe permanecer por encima del 99%, casi en 100, para servidores OLTP. En servidores OLAP, debe estar por encima del 80%.



Gráfica 12. Análisis de transacciones: activas, pendientes y número de transacciones por segundo

Los siguientes 5 contadores sirven para afinar problemas de caché y memoria:

- **SQLServer: Buffer Manager: Page Life Expectancy.** Tiempo en segundos que permanece una página en memoria sin tener ninguna

referencia que la retenga allí. Cuanto más tiempo, mejor. Un valor de referencia, por encima de 300, es decir 5 minutos.

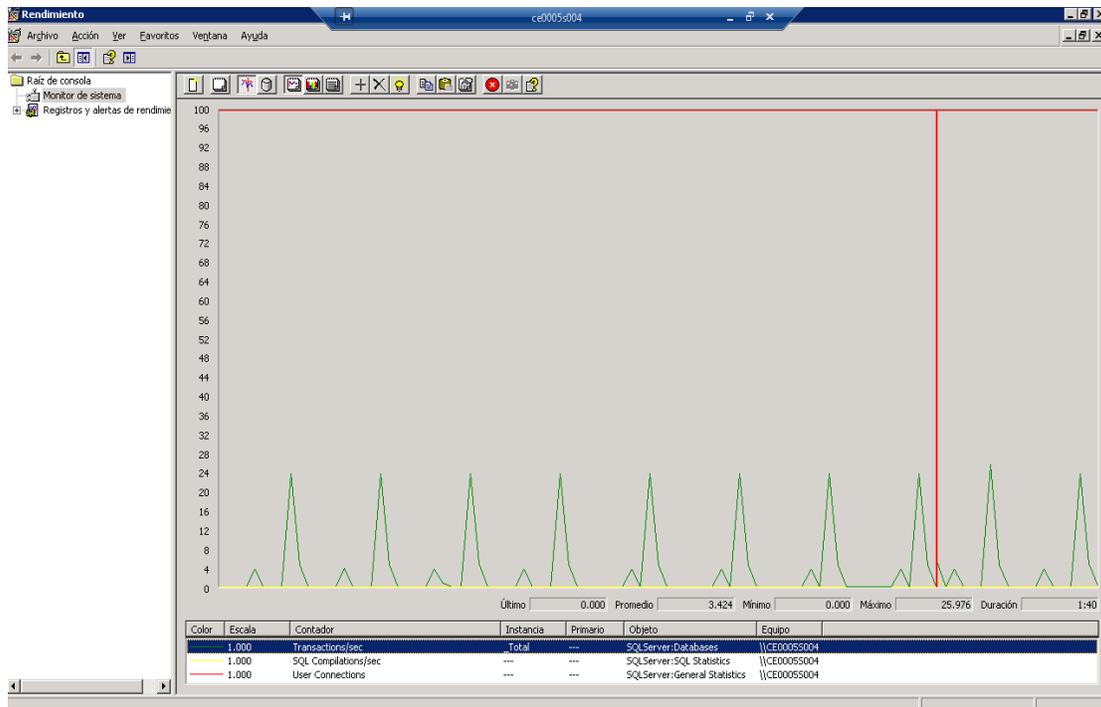
- **SQLServer: Buffer Manager: Lazy Write/sec.** Páginas que salen de la caché por segundo. Al contrario que el anterior, cuanto más alto, peor. Por debajo de 20.
- **SQLServer: Buffer Manager: Checkpoint Pages/sec.** Este checkpoint obliga a bajar a disco todas las páginas que se tengan en memoria. Sólo debe ejecutarse en determinadas circunstancias, la mayoría de ellas, tareas administrativas, por lo que si se ejecuta con mucha frecuencia, estaremos mal utilizando la memoria.
- **SQLServer: Buffer Manager: Procedure Cache Pages.** Este contador indica las páginas de memoria dedicadas a almacenar planes de ejecución de procedimientos almacenados. Un descenso brusco en este contador puede venir acompañado de un descenso del rendimiento, causado por la re compilación de procedimientos almacenados.
- **SQLServer: Databases: Log Flushes/sec.** Indica las veces por segundo que las páginas pasan de caché al archivo de log. Funciona en paralelo al número de transacciones y, como el número de transacciones, cuanto menor sea mayor será el rendimiento.

Para tener una idea del número de usuarios que se maneja, los siguientes 3 contadores son muy útiles:

- **SQLServer: General Statistics: User connections.** Indica el número de conexiones. Sirve para identificar horas de alta y baja actividad y para saber si un pico en otras áreas puede estar relacionado con un mayor número de usuarios en ese momento.
- **SQLServer: Databases: Transaction/sec.** Además de un uso similar al contador anterior, permite determinar qué bases de datos tienen más carga de transacciones. Un caso a tratar de forma particular es el de tempdb, ya que es muy común que sea ésta una de las bases de datos que más transacciones por segundo soporta. Es necesario vigilar y optimizar en lo posible este hecho. Su reducción puede venir de muchas formas, siendo algunas tan obvias como la revisión de los procesos que usan tablas temporales, pero no sólo eso, también hay que observar las consultas muy pesadas y complejas, que usan tempdb para completarse.
- **SQLServer: SQL Statistics: SQL Compilations/sec.** Da una idea de la carga real del servidor, en cuanto a compilaciones se refiere. Si el valor está alrededor a 100, es buena señal. Si pasa de ahí, se estarán

consumiendo muchos recursos en la preparación de planes de ejecución. Por lo general, la detección de bloqueos no se observa de una forma fácil con contadores, salvo que la situación sea realmente caótica.

Estos dos últimos contadores, mostrados a manera de ejemplo, se muestran en las Gráfica 12 y 13.

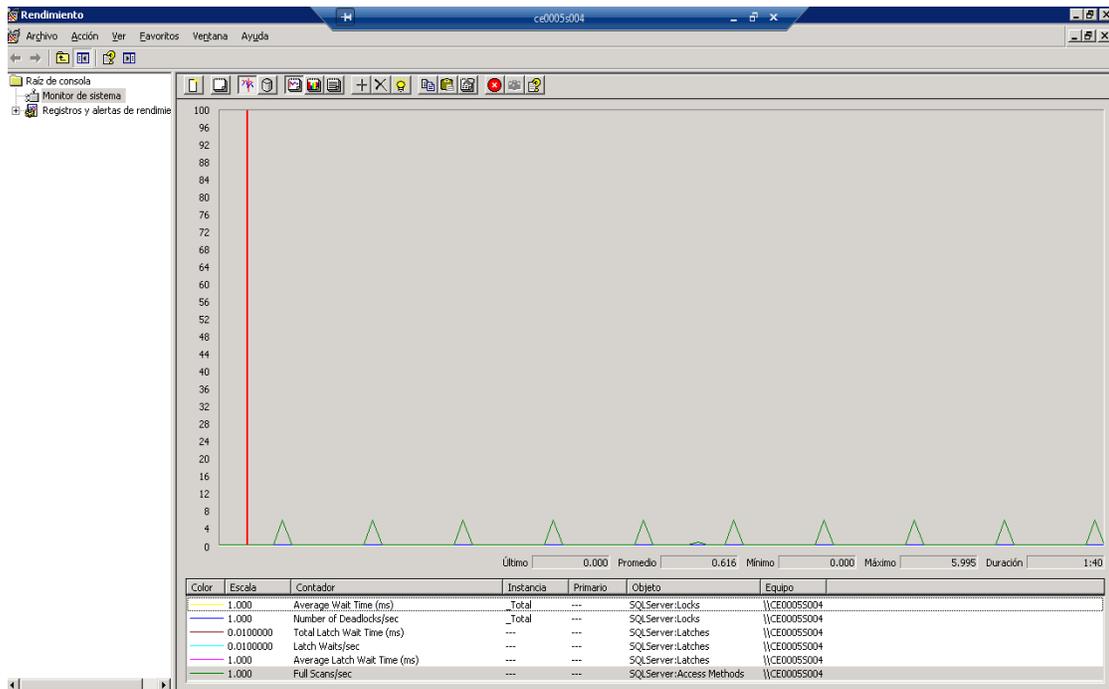


Gráfica 13. Afinación de problemas de caché y memoria

Los siguientes contadores, mostrados en la Gráfica 14, aportan datos sobre los bloqueos e indización que es mejor estudiar, a detalle, con el profiler:

- **SQLServer: Access Methods: Full scans/sec.** Este contador presenta una aproximación de las veces en las que se realizan recorridos de índice, mucho peor que realización de búsquedas en los mismos. Si arroja cifras relevantes, es mejor capturar con profiler y estudiar planes de ejecución de las consultas con más lecturas lógicas.
- **SQLServer: Locks: Number of Deadlocks.** El número de interbloqueos debe ser 0. Si se detecta alguno, hay que revisar y erradicar las sentencias que estén provocando ese problema.
- **SQLServer: Locks: Avg. Wait Time (ms).** Indica la media de milisegundos que hay que esperar para la liberación de un bloqueo.
- **SQLServer: Latches: Average Latch Wait Time (ms), Latch Waits/sec y Total Latch Wait Time (ms).** Estos tres contadores hacen referencia a los

minibloqueos, es decir, bloqueos que son tan cortos que no llegan ni a ser un bloqueo realmente. Eso no significa que no estén ahí. Un elevado número de latches suele venir acompañado de un importante número de bloqueos. Así que estos contadores son de gran ayuda para anticiparse a problemas de este tipo; debido a que lo que hoy son latches, en el futuro pueden ser bloqueos. La forma de gestionar estos valores es obtener una línea base durante un periodo que se considere “normal” y que luego se usará como referencia, para poder saber si la situación se degrada. Si existe alejamiento de esa línea base (por arriba), además de la revisión de las consultas y el nivel de aislamiento entonces se puede deducir que el problema esté en la memoria o en el disco.

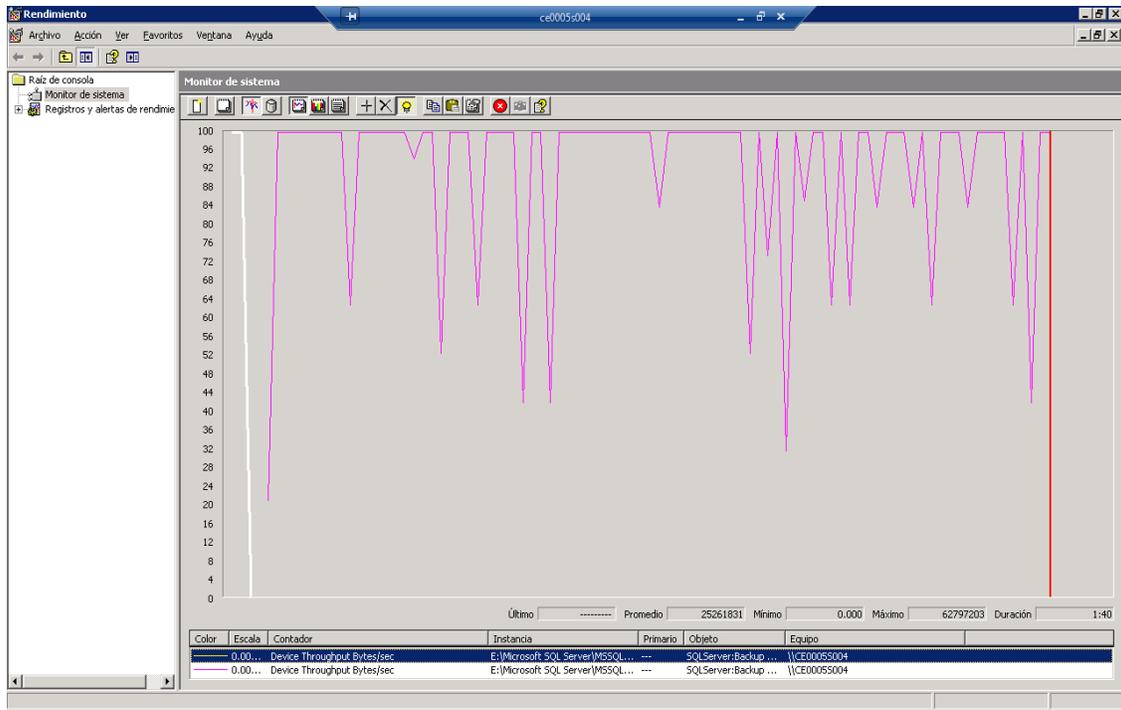


Gráfica 14. Contadores que aportan datos sobre los bloqueos e indización

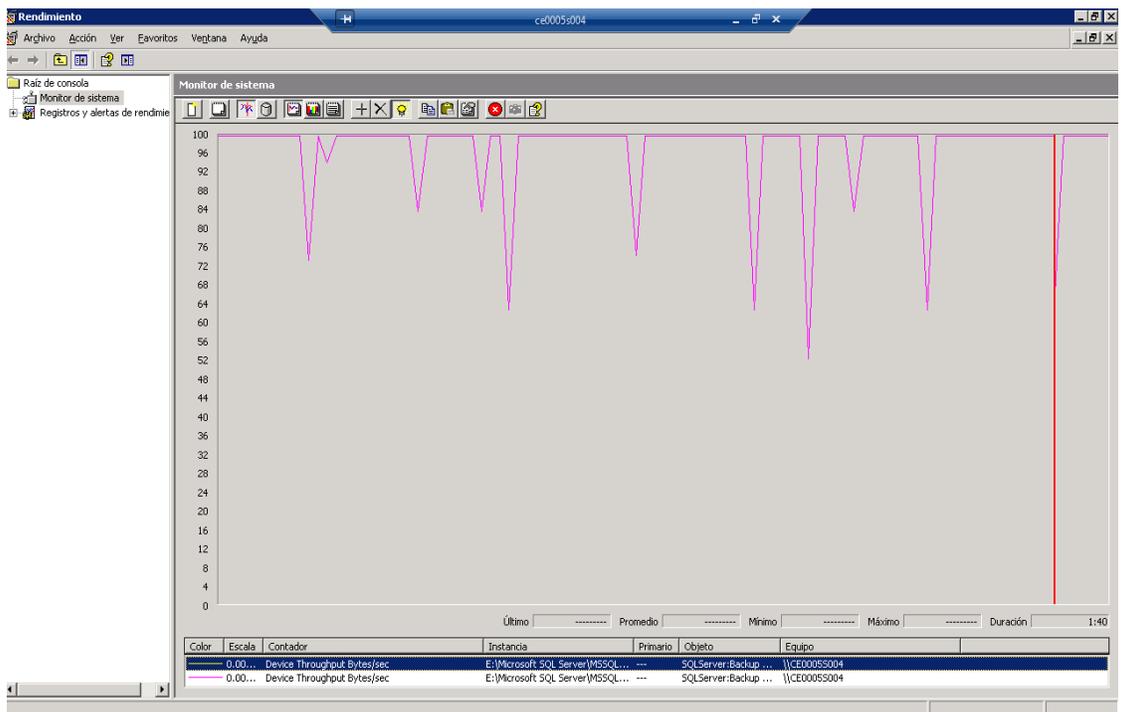
- SQLServer: Backup Device: Device Throughput Bytes/sec.** Cuando se lanza un backup¹³⁸, lo más normal es que se produzca un importante pico en la cola de disco. Si se cuenta con varias bases de datos, lo cual es el caso, es típico lanzar todos los respaldos a la vez. Esto puede implicar un problema de rendimiento que es preciso vigilar. Esto se podría evitar lanzando los backup uno a continuación del anterior. Este contador mide el rendimiento de los backup, y con una regla de tres simple. También puede

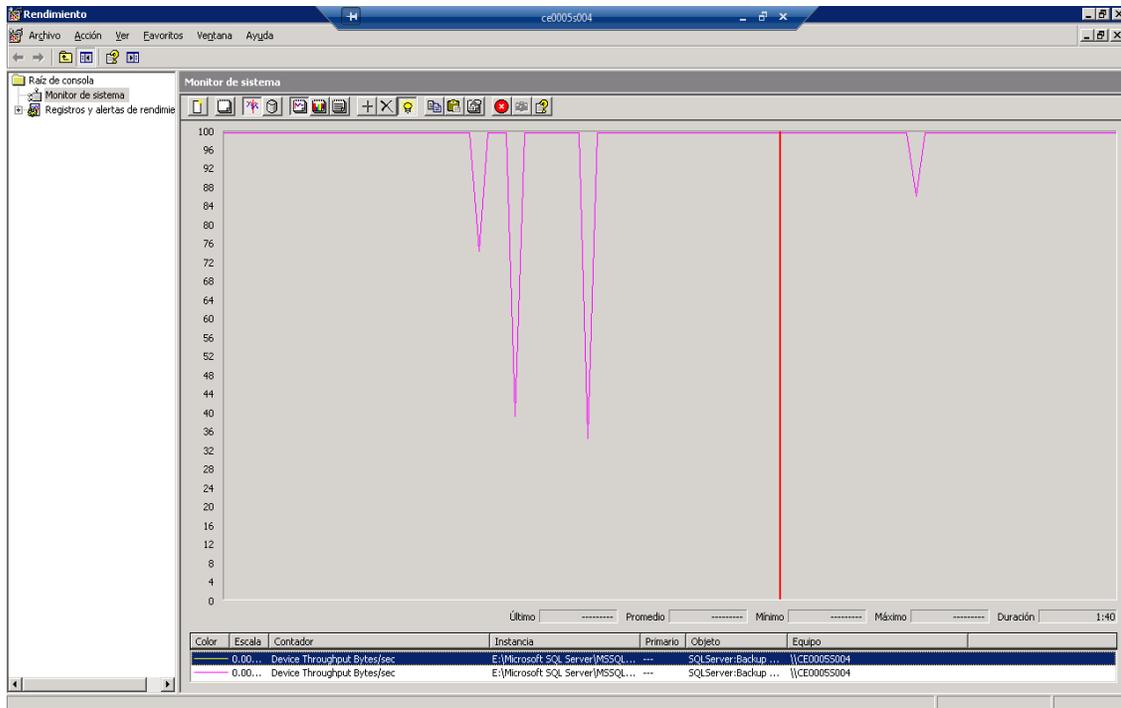
¹³⁸ Respaldo de información, ya sea parcial o total de alguna tabla dentro de la geobase de datos.

ser usado para saber cuánto nos queda para que un backup finalice, por ejemplo. Este contador se presenta en las Gráfica 15 y Gráfica 16.



Gráfica 15. Contadores de análisis de Throughput Bytes/sec al hacer un respaldo en SQL





Gráfica 16. Evolución de un full-backup de la Geobase de datos

Hay que tener cuidado, porque a veces puede arrojar resultados engañosos, apareciendo lecturas cuando hay una importante actividad, generalmente relacionada con tempdb y no hay ningún backup en curso. Los backup del log propios de tempdb, pueden ser detectados en este contador cuando existe un fuerte uso de esta base de datos. Podemos distinguirlos de los backup normales; además de saber que se están haciendo ningún backup en ese momento, porque el contador arroja picos muy breves e intensos, en lugar de valores intermedios durante periodos prolongados.

5.2.3. Procedimientos de actualización y mantenimiento de la geobase para garantizar su actualización a lo largo del tiempo.

En la Tabla 8 se presenta los procedimientos de actualización y mantenimiento de la geobase para garantizar su actualización a lo largo del tiempo. Estos pasos estructurados son esenciales a fin de poder garantizar la validez, calidad y actualidad de la información geoespacial la cual es eje esencial dentro del manejo de la información intrainstitucional.

Tabla 8. Procedimientos de actualización y mantenimiento de la geobase

Procedimiento de	Descripción	Inversión requerida	Políticas de administración
<p>Actualización</p>	<p>Información vectorial: Se solicita en forma anual la actualización de la información a cada una de las áreas generadoras. En casos particulares como REPGA; Calidad del Agua, Agua Potable y Alcantarillado, la información se actualiza cada 6 meses. Información Raster: Esta se solicita a varias dependencias en función de los requerimientos específicos como atención a emergencias y a solicitud de áreas prioritarias como la Dirección General.</p>	<p>Por lo general la inversión se reduce a la gestión de la información en si misma; sin embargo hay que mencionar que para el caso de la información raster, en particular la tipo LiDAR, ha sido necesaria la adquisición de discos externos de alta capacidad para facilitar el transporte de la información. Una vez que se cuenta la información en CONAGUA, esta es trasladada a la Unidad NAS-SAN de SIGA en donde es</p>	<p>Existen los siguientes perfiles (tipo) dependiendo del tipo de usuario:</p> <p>Estándar de lectura Roles: Public, db_datareader, db_denydatawriter, Sin_acceso_a_cartografía_50k Asignación de herramientas:</p> <p>Estándar de procesamiento Roles: Public, db_datareader, db_denydatawriter, Sin_acceso_a_cartografía_50k Asignación de herramientas: Permisos de escritura de tablas en bases de datos de trabajo como por ejemplo en la geobase de datos denominada constructor.</p> <p>DBO Roles: Public, db_datareader, db_datawriter, db_accessadmin, db_backupoperator, db_securityadmin.</p> <p>SDE Roles: Public, db_datareader,</p>

		administrada a través de unidades LUN.	<p>db_datawriter, db_accessadmin, db_backupoperator, db_securityadmin, ddladmin, db_owner.</p> <p>DBA Roles: sysadmin</p>
Mantenimiento	Respaldos incrementales y completos conforme a esquema abuelo-nieto en cintas así como en unidad NAS-SAN a través de LUNs.	La renta de la unidad NAS-SAN es de \$98,399.00 mensuales y la renta de los servidores blade es de \$94,980.61	<p>DBO Roles: Public, db_datareader, db_datawriter, db_accessadmin, db_backupoperator, db_securityadmin.</p> <p>SDE Roles: Public, db_datareader, db_datawriter, db_accessadmin, db_backupoperator, db_securityadmin, ddladmin, db_owner.</p> <p>DBA. Roles: sysadmin</p>

Capítulo 6. CONCLUSIONES Y RECOMENDACIONES

6.1. Sobre el tema de tesis

Innovaciones del tema de tesis

En los siguientes párrafos se hace referencia a las aportaciones que el trabajo hace al conocimiento Universal en los temas de manejo de la información, procesamiento de la información, bases de datos y sistemas de información geográfica.

- **Manejo de Información:** Al trabajar con un recurso natural con una amplia y heterogénea distribución sobre el territorio nacional, la CONAGUA requiere conocer la cantidad, calidad y usos del agua en cada una de las zonas geográficas donde esta se encuentra, permitiéndole una mejor administración del vital líquido. Entre las herramientas que esta Institución emplea para cumplir con su objetivo, está la cartografía digital a través de la cual se identifican y clasifican las distintas regiones hidrológicas, administrativas, cuerpos de agua y demás recursos hídricos existentes en el país, así como la información documental, estadística y financiera.

Para contar con información geográfica vinculada a las grandes bases de datos alfanumérica de la Institución; desde hace más de una década, la Institución se dio a la tarea de generar un acervo cartográfico digital y almacenarlo en una geobase de datos¹³⁹ la cual es factible vincularla con bases de datos alfanuméricas, implementadas en distintas fuentes y distribuirla y ubicar cada nodo en cada uno de los Organismos de Cuenca y Direcciones Locales que conforman la CONAGUA. Esta y otras actividades del quehacer geográfico recaen en la Subgerencia de Información Geográfico del Agua (SIGA); que ha tenido como tareas primordiales la recopilación, manejo, gestión y análisis de datos espaciales. Su misión es “administrar y normar en Geomática, así como en Sistemas de Información Geográfica, a fin de permitir el análisis y consulta de información relacionada a la ubicación de los recursos hidrológicos dentro de su ámbito geográfico, considerando tanto

¹³⁹ Base de datos implementada sobre un sistema administrador de base de datos relacional, el cual permite , a través de un motor de consulta geográfica (ArcSDE), almacenar datos cartográficos y vincularlos tanto con tablas de atributos como con otras bases de datos relacionales las cuales pueden ser geoespaciales o alfanuméricas.

entidades gráficas como estadísticas, apoyando a todas las áreas que conforman la Comisión Nacional del Agua para lograr el uso sustentable del recurso¹⁴⁰. No obstante la misión del SIGA, algunos de sus objetivos específicos principales son la administración del acervo cartográfico digital, servicios de consulta y visualización de datos georreferenciados¹⁴¹, mediante herramientas de consulta de cartografía digital interactiva vinculados a muy diversas fuentes y temas hídricos. Es importante mencionar que para lograr la adecuada administración de esta acervo cartográfico de una forma eficiente, y con base en el artículo 12 bis 6 de la LAN¹⁴², el SIGA debe apoyarse en el trabajo realizado a nivel regional por los Organismos de Cuenca¹⁴³ (OC).

- Procesamiento de información:
- Bases de datos:
- Sistemas de Información Geográfica:

Varias de estas actividades no existen únicamente en la CONAGUA. Áreas con actividades semejantes a las realizadas por el SIGA se encuentran en INEGI, SEMARNAT, CONAM, CONABIO, CONAFOR, por mencionar sólo algunas dependencias desconcentradas del Sector Medio Ambiente. Estos mismos esquemas de Organización Gubernamental se presentan en varias dependencias estatales. Dado lo anterior, se deduce que la implementación de esquemas de administración de información geoespacial con base a los lineamientos señalado en el presente trabajo serían elementos coadyuvantes para la integración de la información en ambientes de alta disponibilidad evitando la duplicidad de actividades al distribuir la carga de trabajo y manejar esquemas interoperables. Además, la alta densidad de información que se requiere en un SIG hace necesario el uso de servicios y aplicaciones que permitan el trabajo con grandes volúmenes de datos y una gran variedad de los mismos, en particular los datos tipo raster. A este gran volumen de datos se le conoce como “Big Data”¹⁴⁴ para los cuales se requiere que las aplicaciones que analizan estos datos tengan una alta velocidad para lograr obtener la información correcta en el momento preciso. Extendiendo la idea, es total y altamente

¹⁴⁰ <http://siga.conagua.gob.mx/Contenido/Acerca/MisionSIGA.aspx>

¹⁴¹ Cualquier elemento hídrico puede ser asociado a sus coordenadas geoespaciales (longitud y latitud) a fin de ubicarlo de forma precisa sobre la superficie terrestre y, en particular, sobre la República Mexicana.

¹⁴² Cf. LAN.

¹⁴³ Ley de Aguas Nacionales (LAN), artículo 12 bis 1. Los Organismos de Cuenca, en las regiones hidrológico - administrativas son unidades técnicas, administrativas y jurídicas especializadas, con carácter autónomo que la LAN les confiere, adscritas directamente al Titular de la CONAGUA, cuyas atribuciones, naturaleza y ámbito territorial de competencia se establecen en la LAN y se detallan en sus reglamentos, cuyos recursos y presupuesto específicos son determinados por la CONAGUA.

¹⁴⁴ <http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

factible implementar el esquema de operación a cualquier nivel y lugar en donde se requiera un *uso eficiente, de bajo costo y estructurado de la información sin importar el ámbito de esta.*

- **Procesamiento de Información:** Para hacer frente a estas actividades a nivel nacional, según lo señalado en el párrafo anterior, el SIGA ha tenido que enfrentar un problema complejo debido a la falta de personal en los OC, lo cual ha hecho difícil la administración de la información geográfica en ámbitos locales. Es por esta razón que ha surgido la necesidad de contar con geobases de datos distribuidas que permitan brindar acceso a este tipo de información al personal que constituye cada uno de los OC. Cabe señalar que la administración de estas geobases de datos, requiere de personal con la suficiente capacidad técnica tanto en el manejo de los sistemas de información geográfica como en la operación de servidores geoespaciales; sin embargo, es difícil contar con tales personas debido a la falta de presupuesto y a las políticas internas de la Institución para la contratación de personal todo ello asociado al problema que implica la adquisición o renta de equipo de cómputo nuevo dado que, debido a que las políticas administrativas de TICs¹⁴⁵ en la CONAGUA deben alinearse a lo dispuesto en el Decreto de Austeridad en materia de TICs¹⁴⁶ derivado del Decreto que establece las medidas de austeridad y disciplina del gasto de la Administración Pública Federal¹⁴⁷. Así, con la ayuda de servidores virtualizados¹⁴⁸ es como se pretende apoyar en el manejo de la información geográfica administrada por la CONAGUA dentro del ámbito nacional. De esta forma, se contaría con una red “distribuida” de servidores geoespaciales vinculados a la Geobase de Datos central permitiendo la migración a servidores físicos conforme se cuente con el personal necesario para su operación y puesta en marcha que, en caso de no poder operar en forma física, se pueden seguir trabajando desde el entorno central en forma remota. Así la idea del trabajo es optimizar los procesos para la interacción de un ambiente de Geobases de Datos:
 1. Operando en una red de área amplia con canales VPN
 2. Virtualizado
 3. Distribuido
 4. Interoperable
 5. Con soporte para Big Data

¹⁴⁵ Tecnologías de Información y Comunicaciones

¹⁴⁶ <http://dgi.sedesol.gob.mx/index/index.php?sec=40>

¹⁴⁷ <http://www.funcionpublica.gob.mx/unaopspf/doctos/adquisiciones/auste041206.pdf>

¹⁴⁸ cf. pág. 16

6. Con acceso a cómputo de nube

- **Bases de datos:** En el artículo de la Ley Nacional del Aguas Nacionales se declara de interés público la descentralización y mejoramiento de la gestión de los recursos hídricos por cuenca hidrológica, a través de Organismos de Cuenca de índole gubernamental esto implica contar con una infraestructura de TI¹⁴⁹ la cual permita el manejo integrado del agua, desde el ámbito de los OC hacia unidad central de CONAGUA así como a la inversa. Para ello es necesario el trabajo coordinado, en equipo, entre éstos y las gerencias que se ubican en oficinas centrales para que, con base en un análisis detallado de la información que poseen, se implementen líneas de acción así como estrategias enfocadas a contribuir al bienestar social y al desarrollo sustentable del recurso hídrico sin descuidar la preservación del medio ambiente. Para el efecto es necesario el intercambio de información entre los OC y oficinas centrales de una forma automática, expedita, segura, confiable y de calidad. Para avanzar en este sentido es que surge la necesidad de generar una red que vincule a los principales bancos de información, tanto alfanuméricos¹⁵⁰ como geográficos¹⁵¹, de cada una de las áreas que conforman a la CONAGUA.

Esta implementación permitirá distribuir la información referente al recurso hídrico a través de medios de TI para presentar un conjunto único de datos relativos a la cantidad, calidad, usos y conservación del agua dentro del ámbito nacional. Esto facilitará el fortalecimiento de los vínculos de trabajo entre los OC y oficinas centrales de la CONAGUA; propiciando la confianza de la sociedad en esta última al contar con la seguridad de que la información consultada es única, oficial y con un alto grado de seguridad de que el sistema opera exitosamente en el ambiente distribuido; trascendiendo límites políticos y geográficos permitiendo con ello cumplir con compromisos de desarrollo, mejorando la cooperación, colaboración y el compromiso de la gestión integrada del agua.

El volumen, la diversidad y la complejidad de datos básicos que se requieren para la administración del recurso hídrico exigen del uso de sistemas de información eficientes capaces de apoyar en forma sustantiva el proceso de tomas de decisión. Un sistema de información hídrico unificado es, desde la perspectiva del autor, "...un conjunto de hardware, software, información y procedimientos computarizados, que permite y facilita el análisis, gestión o

¹⁴⁹ Tecnologías de Información.

¹⁵⁰ Datos en forma de texto o números organizados en forma tabular.

¹⁵¹ Gráficos de un mapa en formato digital.

representación de datos del recurso hídrico a las Instituciones de gobierno que gestionan el recurso. Este sistema unificado debe funcionar como si fuese una única base de datos con información tanto alfanumérica como geográfica en donde se encuentre representado cada elemento por un identificador único pero común a cada Institución” (ref. [5]). De esta forma, señalando un objeto se deben conocer sus atributos, y a la inversa, preguntando por un registro de la base de datos se debe poder saber su localización en la cartografía. Al asociar los datos a un sistema de información alfanumérica y geográfica es posible separar la información en diferentes capas temáticas y almacenarlas en forma independientemente, permitiendo trabajar con ellas de manera rápida y sencilla facilitando al profesional la posibilidad de relacionar la información existente a través de la topología de los objetos.

Forma de manejo del problema de asignación de recursos

En el gobierno federal los procesos sustantivos y la generación de información están traslapados; esto se ha venido dando desde hace mucho tiempo debido a que los tomadores de decisión le han puesto más atención a la infraestructura así como a la adquisición de servicios operativos y se ha dejado de poner atención a la generación u obtención de la información en sí misma. Tomando en consideración que uno de los objetivos de la CONAGUA es poner a disposición del público en general la información sobre el manejo y cuidado del recurso hídrico se hace imprescindible el adecuado manejo y administración de la información; aunado a ello, en CONAGUA los esfuerzos por presentar esta información al público se han centrado en la organización de la estructura de personal así como cuestiones de carácter administrativo dejando a un lado la tarea preponderante de obtención y generación de datos del recurso hídrico. Dado lo anterior se observa que este tipo de políticas de la administración interna de CONAGUA tiene un impacto negativo en la asignación presupuestal la cual, en esencia, es asignada esencialmente para programas con orientación social; esto no significa que ello sea un error empero si implica que de esta forma la información no aparezca de forma explícita, lo cual debería existir como una componente inherente del programa de acciones.

En el pasado, había procesos que manejaban la información en medios impresos y, obviamente, la información se distribuía mucho más lento. Este retardo en su procesamiento, a pesar de contar en la actualidad con equipo de cómputo mucho más rápido y eficiente, se sigue observando en diferentes áreas de la Institución. Se observa de forma muy marcada que hay un traslape entre el concepto de informática con información de forma tal que se manejan,

erróneamente, como sinónimos dando lugar a asignaciones presupuestales dentro de un programa de trabajo que se enfoca a objetivos de adquisición de infraestructura de TICs no a la generación u obtención de información. En esencia se observa que la información no ha tomado su lugar dentro de los programas de gobierno a pesar de que, el propio INEGI, ha hecho esfuerzos por romper con este paradigma.

La asignación presupuestal que se ha observado que depende de los programas y los temas a los que le son de interés a los tomadores de decisión; esta asignación no es suficiente para atender las necesidades de las áreas, para cada tema y función. Si a esto se le agrega que el manejo de la información requiere de TICs se tiene que, hasta la fecha, se adolece de una clara política o atribuciones de algún área dedicada al manejo de la información.

En el ámbito mundial se habla del CIO¹⁵², que es un puesto de trabajo que se dio al ejecutivo de más alto rango en una empresa responsable de los sistemas de tecnología de información y sistemas informáticos que apoyan las metas de la empresa. En CONAGUA el CIO, reporta al Subdirector General de Administración. De esta forma el tomador de decisión basa parte de sus decisiones en la información que le brinda el CIO empero no se observa claramente que hay medidas sustantivas que no han sido visualizadas del todo por el CIO en cuanto al potencial de los SIG en la Institución. Esto impacta en la asignación presupuestal pues no hay una organización dirigida hacia una adecuada gestión del recurso hídrico al no contextualizar su estudio y análisis con base en un SIG, lo cual en principio debería ser obvio dada la propia naturaleza del recurso natural que administra CONAGUA. Esto está claramente vinculado con la gobernanza del agua¹⁵³ dado que la situación de presión sobre los recursos hídricos no es la misma en las diversas regiones y localidades del país; adicionalmente los cambios climáticos que se prevén para el futuro forzarán la migración a zonas urbanas además de la que ya se origina por la concentración de actividades económica. La interpretación así como el análisis de los diferentes escenarios a futuro (prospectiva) se hacen mucho más enriquecedores y tienen mucho más sentido al analizarlos en un entorno geográfico.

¹⁵² Chief Information Officer (CIO) o Information Technology (IT) Director

¹⁵³ Referida esta como la democracia, la participación en la toma de decisiones y el buen gobierno en la administración del recurso hídrico. En materia ambiental estos tres requisitos son fundamentales para afrontar la crisis que ya presentan sectores ambientales como el agua, con la inclusión de todos los actores sociales en la conformación de las decisiones, de tal forma que permita la aceptación y la eficacia de las medidas que se adopten, pues es aquí donde radica uno de los mayores problemas de gestión del medio ambiente.

Cambiando un poco la percepción de la política del manejo de la información de una forma simple, la asignación presupuestaria depende de las decisiones de mandos superiores que en ciertos casos tienen interés en temas muy particulares los cuales no están realmente asociados al entorno geoespacial, incluso del manejo de información en bases de datos relacionales.

Por otro lado existe una gran disparidad entre el manejo de información entre diferentes áreas que, vistos desde otro ámbito son complementarias, caso concreto: SINA-SIGA. Las demandas políticas de los diferentes niveles, que cambia de un día a otro, la duplicidad de acciones y de planes de trabajo para el manejo de la información, la competencia entre áreas que deberían complementarse genera una gestión y manejo de información de forma diferencial la cual sólo depende de mandos superiores. La parte paradójica es que la asignación de recursos se reduce debido a la falta de visión hacia otras áreas de la propia Institución. De esta forma se tienen dos visiones divergentes:

Nivel operativo.- Sabe que es lo que se necesita pero no que requieren para ello. Los mandos superiores no lo escuchan.

Mando superior.- Tiene sus propias prioridades y se asigna el presupuesto en forma limitada conforme a su propia visión. En ocasiones se generan asignaciones “fortuitas” debido a economías o falta de ejecución presupuestaria en el ejercicio fiscal anterior. Esta situación resulta benéfica toda vez que las áreas que no cuentan con elementos técnicos, hardware o software, pueden tener acceso a esos recursos.

6.2. CONCLUSIONES GENERALES

Como lo expresa el informe de la World Water Vision citado por el Centro Internacional de Investigaciones para el Desarrollo (IDRC) [ref. 6], la crisis del agua es, “una crisis en el manejo de la información sobre el agua”. Dicha reflexión simplifica la interacción más crítica del problema del manejo de la información tanto del recurso hídrico como su vinculación con todo el medio ambiente, presentando un verdadero desafío para las actuales y futuras investigaciones, perfilando estudios posteriores **hacia el manejo de la información asociada al recurso hídrico**, basado s en análisis de información confiable, única, reciente y disponible para la toma de decisiones. Esto nace que los Sistemas de Información se conviertan en verdaderas herramientas funcionales para la planificación del desarrollo.

Las propuestas de alternativas de solución al problema, antes referido, debieran concentrarse en la implementación de redes interoperables de recursos hídricos,

y posteriormente avanzar hacia una integración de vínculos interoperables en el Sector Medio Ambiente. De esta forma se lograría alcanzar mediante el enfoque propuesto en este trabajo, el objetivo principal de incorporar un estudio interdisciplinario al problema ambiental en cuestión, buscando superar las limitaciones de los estudios de carácter sectorial y parcial que ignoran las características sistémicas de los procesos involucrados en la problemática ambiental. Se verifica también la flexibilidad, retroalimentación y posibilidad de reacción y adaptación de cualquier metodología implementada, permitiendo el planteo de acciones que combinen consulta de información geoespacial local y distribuida, en función de la naturaleza misma de considerar al ambiente y la problemática ambiental como un sistema complejo.

La gran facilidad que ofrece el acceso a Internet ha cambiado la forma en que los gobiernos proveen los servicios al público en general y a otras Instituciones. Los ciudadanos que tienen un mayor conocimiento relativo a los recursos y capacidades de las TI están ejerciendo mayor presión para acelerar la oferta gubernamental de estos servicios. Estos servicios no solo permiten una mayor calidad de oferta de servicios al público sino eficiencia y eficacia en la transmisión de información entre dependencias del gobierno visto como un todo y hacen, como consecuencia, más transparentes los procedimientos a los servicios públicos organizados.

Si se analiza la evolución de los sistemas de procesamiento de datos de cualquier gobierno, es claro que la primera prioridad es el desarrollo de soluciones que diseccionen el flujo de datos al interior de gobierno. Las decisiones relativas a la administración de los recursos hídricos para el desarrollo de prioridades están basadas en un análisis del entorno geográfico y en el deseo de demostrar las visiblemente altas mejoras en los servicios públicos prestados por todas y cada una de las dependencias del gobierno. Las aparentes soluciones son efectuadas sobre una base *vertical*, es decir en forma independiente, cada una enfocando el problema hacia una dependencia gubernamental en particular la cual trata en lo posible solucionar el problema. Posteriormente se observa que cada Secretaría del Gobierno Federal tiene su propio sistema de información independiente. Este enfoque vertical puede en algunos casos ser adecuado para la solucionar algunos problemas que atañen a un sector específico, no obstante los muchos de los problemas medioambientales requieren de un enfoque interinstitucional el cual debe considerar de una visión completa del entorno así como las necesidades de los ciudadanos. Cabe señalar que los ciudadanos no tienen solamente necesidades

verticales, más bien tienen necesidades *horizontales*, es decir en forma *interoperable*. Bajo el supuesto de que todos los sistemas verticales estuviesen desarrollados para trabajar eficientemente, con enfoque hacia la ciudadanía, aun así habría una necesidad de efectuar muchas interacciones transaccionales con sistemas independientes y no integrados. Como se observa esta dualidad en las necesidades de servicios interoperables requiere del trabajo en conjunto de varias Instituciones gubernamentales.

Así de acuerdo con un nuevo paradigma de operación basado en un modelo interoperable y la gran cantidad de alternativas que ofrecen los servicios basados en Web se observa un claro patrón a seguir, del cual los pasos a seguir se pueden representar en la siguiente secuencia:

- TI basadas en tecnologías interoperables.
- Presencia en la Web.
- Interacción básica a nivel de geobases de datos interoperables.
- Procedimientos transaccionales.
- Integración y estandarización.
- Revolución.

Actualmente las primera y segunda fases de presencia en la Web son las más simples. Estas se encuentran por lo general desarrolladas, por cada dependencia de gobierno, a través de sus portales o sitios Web los cuales contienen información básica para el público. El desarrollo de estas fases se efectúa independientemente por cada Secretaría de Estado, gobierno Estatal o municipal, sin la presencia de objetivos en común.

La siguiente fase, geobase de datos interoperables o interacción, es la evolución natural de los sitios Web de consulta de información del Sector Medio Ambiente. De esta forma cada dependencia gubernamental y el público en general no solamente puede buscar información, sino comunicarse con autoridades de diferentes ámbitos de gobierno, facilitando el intercambio de información. Todas las consultas podrían interactuar a fin de obtener información de varias dependencias en un solo canal de consulta debido a que todo se encuentra integrado en aplicaciones primarias electrónicamente interoperables.

La tercera fase, transacción, empezará a proveer interacciones y servicios más relevantes. Los usuarios pueden hacer sus consultas totalmente en línea y dejar a los sistemas encargarse de todo. Esta es la fase donde la primera necesidad

planteada en la fase anterior relativa a interoperabilidad se hace visible, cabe aclarar que esta fase no es en donde se inicia el proceso de interoperabilidad. Muchas dependencias del Estado desarrollan sus propias soluciones independientes. Sin embargo, algunas de estas necesitarán interoperar con los sistemas previamente existentes de ahí la necesidad de contar con “gateways” interoperables.

La fase de estandarización integra las aplicaciones entre diferentes Instituciones. La interoperabilidad se hace más importante; las transacciones efectuadas entre ciudadanos y gobierno se hacen más simples y hay un incremento sustancial en producción y calidad de datos, debido a que cada dependencia se hace responsable de su ámbito de competencia.

La última etapa, transformación total, permite una gran variedad de transacciones a ser integradas; se pueden ofrecer nuevos servicios y el valor en conjunto de la integración se incrementa. Este modelo de integración es similar a una red estructura en la cual el valor global de la integración puede ser medido por el número de posibles interconexiones en la red y el valor de los sistemas interoperables puede ser medido como función del número de transacciones conectadas.

Las iniciativas de e-gobierno trabajan principalmente en la disminución del tiempo de proceso de cada transacción, mientras que las soluciones interoperables trabajan fundamentalmente en la disminución de las transacciones mediante la combinación de muchas de éstas con el mismo “esfuerzo transaccional”. Luego entonces, el desarrollo de sistemas de e-gobierno y al mismo tiempo haciéndolos interoperable a través del uso de servicios Web producirá mayores ganancias a la misma o tal vez menor escala de costos.

La implementación de una geobase de datos distribuida operando sobre servidores virtuales sugiere la reducción de recursos al trabajar con servidores donde no los hay y tener personal donde no lo hay. Lo anterior debido a que al procesar administrar la información geoespacial en servidores virtuales se genera hardware “lógico” con lo que se cuenta con servidores tan completos o complejos como el equipo físico lo permita. En cuanto al personal, al ser factible la administración remota de estos equipos virtuales, entonces se hace factible el don de la ubicuidad al poder atender varios servers desde una sola localidad. Adicionalmente al segmentar la información es posible atender problemas de índole regional sin la necesidad de tener que recurrir a bancos de datos centralizados lo cual genera, a nivel central, cuellos de botella innecesarios pues

la respuesta más obvia a este problema es que cada Organismo de Cuenca opere con información geoespacial que solamente sea de su ámbito de competencia. Obteniendo ganancias en tiempo y recursos de acceso.

Por otro lado la implementación de esquemas semejantes de consulta de datos geoespaciales en todas aquellas dependencias del Gobierno Federal que hagan uso de análisis geomático permitirá un desarrollo más rápido en la administración del conocimiento propiciando la implementación de soluciones acordes a las necesidades regionales las cuales, por lo general, difieren sustantivamente de las experiencias en el ámbito central.

Una gran parte del Gobierno Federal está constituido por oficinas desconcentradas y/o descentralizadas lo que genera pequeñas islas de información. Con esta forma de administración de la información se ahorra recursos, se reducen tiempos de actualización y/o migración de sistemas sin afectar la operación del acceso y la consulta a datos prioritarios facilitando tiempos de respuesta mejorando, así tanto la eficacia como la eficiencia en la consulta de datos geoespaciales. De esta forma se cumpliría cabalmente con las necesidades de procesamiento y/ consulta que requieren las grandes organizaciones del Sector Gubernamental. Cabe mencionar que, al sustituir el concepto de cómputo virtualizado por el de cómputo en la nube, por antonomasia, se obtendría una mejora importante en cuanto a la administración de recursos se refiere.

Los interesados ayudarían a crear la demanda que complementaría el impulso de la oferta de las nuevas ideas y tecnologías provenientes del Sector Gubernamental. Con la implementación de esquemas de compartición de información en los cuales se considerasen: bases y geobases de datos distribuidas operando sobre servidores virtuales interaccionando con bases de datos de otras dependencias gubernamentales de forma interoperable tanto los instrumentos como las aplicaciones y los servicios que se generasen como resultado de su trabajo cotidiano aumentarían y mejorarían su productividad así como su eficiencia dado que los conjuntos de datos tendrían una alta disponibilidad vía Internet.

Como una iniciativa adicional se propone la creación de un programa a nivel Federal que haga una diferencia positiva en las vidas de los mexicanos vinculada a las dependencias gubernamentales. Esta iniciativa que podría denominarse Geoinformación Gubernamental Sostenible la cual haría disponibles tecnologías modernas de cartografía a los jefes de gobierno de comunidades rurales e indígenas, a través de redes distribuidas e interoperables

para que estas comunidades puedan efectuar decisiones informadas afectando su desarrollo sostenible. Ello les permitiría a los líderes tomar decisiones y análisis de su entorno mediante:

1. La comprensión de sus propias necesidades y posibilidades de desarrollo.
2. La estimulación de la innovación local.
3. La toma de decisiones relacionadas con la planificación comunitaria y el ordenamiento territorial.

Mediante la entrega de datos, de un sistema de información geográfica y de actividades de capacitación la iniciativa de Geoinformación Gubernamental Sostenible proporcionaría a los líderes comunitarios los instrumentos para entender su situación local y para planificar acciones apropiadas.

6.3. CONCLUSIONES PARTICULARES.

Durante los últimos años el Gobierno en cuanto al sector Medio Ambiente, y en particular en el Sector Hídrico, se refiere ha adquirido considerable experiencia en la modernización de su infraestructura de datos espaciales. El Gobierno de México ha encontrado una manera para construir un IDS, que incluye los siguientes elementos:

- Un enfoque claro sobre el modelo inicial interoperable el cual incluye estudios de caso;
- Un abordaje evolutivo que permite y facilita la substitución tecnológica apropiada a medida que sea necesario;
- El uso de sistemas, de conceptos y de tecnologías abiertos;
- Una infraestructura distribuida para garantizar la interoperabilidad; y
- Lo más importante, fuerte colaboración gobierno-gobierno e Instituciones de Investigación y educativas.

El resultado ha sido una Infraestructura Mexicana de Datos Geoespaciales que trae enormes beneficios a nuestro país.

Las Instituciones gubernamentales han respondido al desafío de trabajar en equipo para construir una nueva infraestructura de datos espaciales y sistema de información. Así el gobierno mexicano, en particular en el ámbito del sector medio ambiente dispone de considerables capacidades y de instrumentos especializados en materia de geomática.

La integración de los conocimientos del área de información puede jugar un rol muy importante en la potenciación del uso de estas facilidades. Es fundamental la integración de la experiencia de especialistas en Sistemas de Información Geográfica e Informática para hacer un uso creativo de las facilidades de las redes interoperables a través de una estrategia de trabajo con este enfoque y no limitarse a establecer solamente la implementación de red de computadoras. Es sumamente importante adecuar los servicios que se han desarrollado en Internet, y otras redes, a las necesidades de los usuarios y al tipo de servicios que brindan las dependencias gubernamentales abre amplias perspectivas en el uso de las redes interoperables para apoyar sus funciones.

Internet asociado a las tecnologías de interoperabilidad constituye un modelo valioso a considerar en la planificación de consulta de información entre dependencias gubernamentales debe combinarse con la utilización de modelos tales como manejo de bases de datos geoespaciales.

Los servicios que se instalen deben buscar permanentemente la normalización y ajustarse a servicios similares internacionales. Las aplicaciones locales que se desarrollen deben enmarcarse en esta línea.

GLOSARIO

API. Interfaz de programación de aplicaciones (del inglés Application Programming Interface) es el conjunto de funciones y procedimientos o métodos, en la programación orientada a objetos, que ofrece cierta biblioteca para ser utilizado por otro software como una capa de abstracción. Se trata del conjunto de llamadas a ciertas bibliotecas que ofrecen acceso a ciertos servicios desde los procesos y representa un método para conseguir abstracción en la programación, generalmente entre los niveles o capas inferiores y los superiores del software. De esta forma, los programadores se benefician de las ventajas de la API haciendo uso de su funcionalidad, evitándose el trabajo de programar todo desde el principio.

INTEROPERABILIDAD. Se entiende como la posibilidad de que los sistemas de las Administraciones Públicas trabajen juntos de forma satisfactoria y productiva independientemente de la tecnología o la aplicación que se utilice, o qué proveedor ha suministrado el sistema subyacente.

Si bien la interoperabilidad puede tener significados diferentes dependiendo del contexto, en el área de las TI el término generalmente es entendido como “la capacidad de diferentes productos y servicios de TI para intercambiar y usar datos e información (es decir "hablar") con el objetivo de funcionar juntos en un entorno conectado en red.”

Varias fuentes de todo el mundo corroboran esta definición:

El Diccionario de Telecomunicaciones de Newton define la interoperabilidad como “la capacidad de gestionar el software y el intercambio de información en una red heterogénea, p. e. una red extensa constituida por varias redes locales diferentes.”

El Acta de e-government de los E.U.A. del año 2002 define la interoperabilidad como “la capacidad con la que comunican e intercambian datos diferentes sistemas operativos y de software, aplicaciones, y servicios de una manera exacta, eficaz y consistente.”

El Acta de Derechos de Autor de Milenio Digital de los E. U. A. (DMCA) del año 1998 define la interoperabilidad como “la capacidad con la que programas de ordenador intercambian información, así como la de utilizar mutuamente esa información que ha sido intercambiada entre ellos.”

Uno de los libros blancos sobre interoperabilidad creado por EICTA (European IT trade association) en junio de 2004, define la interoperabilidad como “la capacidad de dos o más redes, sistemas, dispositivos, aplicaciones o componentes para intercambiar información entre ellos y usar la información intercambiada.”

El Marco europeo de Interoperabilidad, una iniciativa para facilitar la interoperabilidad de servicios y sistemas a nivel pan-europeo, define la interoperabilidad como “la capacidad de los sistemas de Tecnologías de la Información y las Comunicaciones (TIC) y de los procesos de negocio que soportan, para intercambiar datos y permitir la compartición de información y conocimientos.”

E-SERVICES: (e-servicios) en este documento abarca tanto el empleo interno de los sistemas y la información en las Administraciones Públicas, como el despliegue de servicios externos y online a ciudadanos y negocios.

SOA: La Arquitectura Orientada a Servicios (Service-Oriented Architecture o SOA), es un concepto de arquitectura de software que define la utilización de servicios para dar soporte a los requerimientos de software del usuario.

SOA proporciona una metodología y un marco de trabajo para documentar las capacidades de negocio y puede dar soporte a las actividades de integración y consolidación.

En un ambiente SOA, los nodos de la red hacen disponibles sus recursos a otros participantes en la red como servicios independientes a los que tienen acceso de un modo estandarizado. La mayoría de las definiciones de SOA identifican la utilización de Servicios Web (empleando SOAP y WSDL) en su implementación, no obstante se puede implementar una SOA utilizando cualquier tecnología basada en servicios.

Al contrario de las arquitecturas orientado a objetos, las SOAs están formadas por servicios de aplicación débilmente acoplados y altamente interoperables. Para comunicarse entre sí, estos servicios se basan en una definición formal independiente de la plataforma subyacente. La definición de la interfaz encapsula (oculta) las particularidades de una implementación, lo que la hace independiente del fabricante, del lenguaje de programación o de la tecnología de desarrollo (como Plataforma Java o Microsoft .NET). Con esta arquitectura, se pretende que los componentes software desarrollados sean muy reutilizables, ya que la interfaz se define siguiendo un estándar.

PORTABILIDAD: En los primeros días del inicio de la computación los programas y los sistemas eran desarrollados para una plataforma específica, haciendo uso de lenguajes de programación que no eran totalmente. Cada vendedor de hardware tenía su propio y único lenguaje ensamblador, así como su implementación de Fortran y COBOL, que eran entonces los lenguajes de programación estándar. Las herramientas desarrolladas por cada vendedor tenían características únicas. Los programas resultantes no podían ser fácilmente compilados en diferentes plataformas sin tener que hacer un cambio significativo en el código. Para hacer más complejo el problema, había un auge en cuanto al nacimiento de nuevos sistemas operativos y muchos de ellos estaban hechos para correr en un sistema operativo en particular. La mayoría de los programadores utilizaban comandos internos de lenguajes de bajo-nivel los cuales llamaban a funciones internas del equipo usando formatos de archivos que existían solamente en un sistema operativo en particular.

Cambiar una aplicación de una plataforma a otra era muy difícil, en el mejor de los casos sino imposible, a pesar de estar basados en sistemas operativos similares tales como IBM de VSE a MVS, PDP 11 a VAX, o sobre plataformas de diferente formato como UNIX. La tarea de ejecutar un programa en una plataforma diferente fue llamada “porting” o portabilidad y aun hoy en día es usada en ambientes UNIX.

Al correr de los años la adopción de estándares en los lenguajes de programación tales como C, C++ y C#, así como plataformas tales como J2EE y .NET, han reducido en forma significativa la necesidad de hacer portables las aplicaciones de una plataforma a otra; no obstante, aún es necesario cuando se mueve una aplicación sobre diferentes sistemas operativos, como por ejemplo mover de UNIX a MVS o de UNIX a Windows.

El concepto inicial de portabilidad fue definido como una necesidad de correr la misma aplicación en un hardware diferente para cual fue originalmente diseñado. Así la portabilidad fue básicamente relacionada a las restricciones impuestas por el hardware. El sistema operativo fue, entonces, íntimamente vinculado con el hardware.

En la página web <http://whatis.techtarget.com/> se define el término portabilidad, de la siguiente forma:

“La portabilidad es una característica atribuida a un programa de computadora de poder usarse en sistemas operativos diferente para el cual fue creado sin requerir de una reingeniería mayor. Además de las diferencias del idioma, el

hacer portable un programa puede requerir también conversión de los datos y adaptación a los nuevos procedimientos del sistema operativo”.

Por lo tanto, la necesidad de portabilidad existe solamente cuando hay una necesidad imperativa de migrar una aplicación de una plataforma a otra totalmente diferente. Portabilidad no debe ser confundida o mezclada con el término interoperabilidad. Dos sistemas diferentes pueden interoperar sin la necesidad de que alguno de ellos sea portable.

COMUNICACIÓN-CRUZADA: Se entiende como un conjunto de estándares tales como ASCII, BCD y EBCDIC, los cuales fueron desarrollados para estandarizar la forma en que los datos eran almacenados en las computadoras. Los protocolos de datos de las telecomunicaciones tales como Poll Select, BSC1, BSC3 y SDLC así como protocolos de redes de datos tales como SNA, DECNET, ISO/OSI y TCP/IP fueron desarrollados para facilitar la comunicación entre dispositivos.

La existencia de estos protocolos y estándares permitieron, en un principio, a dispositivos tales como impresoras conectarse unas con otras. Posteriormente las computadoras se habilitaron para intercambiar datos usando opciones tales como cintas magnéticas fuera de línea; diálogo computadora-a-computadora; o (tele) comunicaciones on-line, en tiempo-real y canales.

NAT (network address translation). Es el proceso de modificación de la dirección IP en los encabezados de paquetes IP en tránsito a través de un dispositivo de tráfico de enrutamiento o router.

El tipo más sencillo de NAT proporciona una traducción uno a uno de las direcciones IP. Este tipo de NAT se denomina NAT básico, y a menudo también se denomina NAT uno-a-uno. En este tipo de NAT sólo las direcciones IP, la suma de comprobación de encabezado IP y cualesquiera sumas de comprobación de nivel superior que incluyen la dirección IP se cambian. El resto del paquete se deja intacto (al menos para una funcionalidad básica de TCP/UDP; algunos protocolos de nivel superior puede necesitar traducción posterior). La NAT básica se puede utilizar para interconectar dos redes IP que tienen direccionamiento incompatible.

Su utilización es común para ocultar un espacio de direcciones IP entero, que generalmente consiste de direcciones IP privadas, detrás de una sola dirección IP (o en algunos casos un pequeño grupo de direcciones IP) en otro espacio de direcciones, por lo general pública. Para evitar la ambigüedad en el manejo de paquetes devueltos, una NAT uno a muchos debe alterar información de mayor

nivel, como puertos de comunicaciones TCP/UDP salientes y mantener una tabla de traducción para que los paquetes de retorno puedan ser traducidos correctamente de nuevo.

Como se ha descrito, este método permite la comunicación a través de un router sólo cuando la conversación se origina en la red enmascarada, ya que éste establece las tablas de traducción. Por ejemplo, un navegador web en la red enmascarada puede navegar por un sitio externo en la web, pero un navegador externo no podrá navegar en un sitio web en la red enmascarada. Sin embargo, la mayoría de los dispositivos NAT hoy en día permiten al administrador de red configurar entradas de la tabla de traducción para uso permanente. Esta característica se refiere a menudo como "NAT estático" (NATeo) o reenvío de puertos y permite que el tráfico originado en el "exterior" de la red pueda llegar a los anfitriones designados en la red interna enmascarada.

A mediados de la década de 1990 la NAT se convirtió en una herramienta popular para aliviar las consecuencias del agotamiento de las direcciones IPv4. Este tipo de direccionamiento se ha convertido en una característica común, indispensable en los routers para el hogar y la pequeña oficina de conexiones a Internet. La mayoría de los sistemas que utilizan NAT lo hacen a fin de que varios hosts de una red privada puedan acceder a Internet mediante una única dirección IP pública.

La traducción de direcciones de red tiene serios inconvenientes en cuanto a la calidad de la conectividad de Internet y requiere una atención cuidadosa a los detalles de su implementación. En particular, todos los tipos de NAT tienden a romper el modelo de IP originalmente previsto de extremo a extremo a través de la conectividad de Internet y NAT lo cual hace que sea difícil para los sistemas detrás de un NAT el aceptar comunicaciones entrantes. Como resultado, se han implementado métodos de NAT transversal lo cuales se utilizan para aliviar esto problemas encontrados.

RED DE ÁREA DE ALMACENAMIENTO (SAN, STORAGE-AREA NETWORK).

Es un tipo especial de red de área local de alta velocidad destinada a conectar numerosos bancos de dispositivos de almacenamiento (discos) a las computadoras que utilizan los datos. Así, las redes de área de almacenamiento ayudan a construir sistemas de discos compartidos a gran escala. El motivo para utilizar redes de área de almacenamiento para conectar múltiples computadoras a grandes bancos de dispositivos de almacenamiento es esencialmente el mismo que para las bases de datos de disco compartido

SISTEMA DISTRIBUIDO DE BASES DE DATOS. Es un conjunto de computadoras donde se almacena la base de datos. Varios medios de comunicación, como las redes de alta velocidad o las líneas telefónicas, son los que pueden poner en contacto las distintas computadoras de un sistema distribuido. No comparten ni memoria ni discos. Las computadoras de un sistema distribuido pueden variar en tamaño y función pudiendo abarcar desde las estaciones de trabajo a los grandes sistemas.

SITIO O NODO. Se refiere a las computadoras que forman parte de un sistema distribuido, dependiendo del contexto en el que se mencionen. Para enfatizar la distribución física de estos sistemas se usa principalmente el término sitio.

TRANSACCIÓN LOCAL. Es aquella que accede a los datos del único sitio en el cual se inició la transacción.

TRANSACCIÓN GLOBAL. Es aquella que, accede a los datos situados en un sitio diferente de aquel en el que se inició la transacción, o bien accede a datos de varios sitios distintos.

LINCOGRAFÍA

- [1] Álvarez, Carrión G., "Integración de esquemas en bases de datos heterogéneas fuertemente acopladas", http://catarina.udlap.mx/u_dl_a/tales/documentos/msp/alvarez_c_g/capitulo1.pdf, Tesis de Maestría, Colección de Tesis Digitales, Universidad de las Américas, Puebla, México, Mayo 1999.
- [2] Raymond A. Applegate, et. al, "*Noninvasive Measurement of Corneal Topography*", IEEE Engineering in Medicine and Biology, Enero 1995, p. 30-41.
- [3] Data Keeper, "¿Qué es un hypervisor", <http://www.datakeeper.es/?p=716>
- [4] Durand, Hélène. "*Sistemas de Información Geográfica para un Desarrollo Sostenible*", documento web <http://www.fao.org/sd/spdirect/gis/eigis000.htm>, 1999.
- [5] García, Rolando, "*Interdiscipliniedad Sistemas Complejos*", ensayo incluido en E. Leff (comp.) *Ciencias Sociales y Formación Ambiental*. Editorial Gedisa y PNUMA. Barcelona. 1994, p. 86-89.
- [6] INEGI, "Sistemas Nacionales Estadístico y de Información Geográfica", http://www2.inegi.gob.mx/sneig/contenidos/espanol/pronadeig/resumen_sernat.aspx?c=2480
- [7] Madrid, Lorenzo, "Interoperability: A beginner's guide to understanding its past, present, and future and its impact on governments", Microsoft Co., Libro Blanco Octubre 2005, pp. 21.
- [8] Montaña, Espinosa Carlos Raúl, "Vinculación de Sistemas de Información Gubernamentales para la creación de un Sistema Ejecutivo del Agua basado en un Enfoque Sistémico", http://www.imta.edu.mx/images/pdf/tesis_carlos%20raul%20montano%20espinosa.pdf, Tesis de Maestría en Ciencias del Agua, Instituto Mexicano de Tecnología del Agua, Jiutepec, Mor. México, 2009.
- [9] http://es.wikipedia.org/wiki/Ancho_de_banda. Un ejemplo de banda estrecha es la realizada a través de una conexión telefónica, y un ejemplo de banda ancha es la que se realiza por medio de una conexión DSL, microondas, cable módem o T1. Cada tipo de conexión tiene su propio ancho de banda analógico y su tasa de transferencia máxima. El ancho de banda y la *saturación redil* son dos factores que influyen directamente sobre la calidad de los enlaces.
- [10] William J. Cosgrove, Frank R. Rijsberman, "*World Water Vision*" (Visión mundial del agua), Earthscan, 2000. Presenta una revisión excepcionalmente útil del problema mundial de la escasez de agua.

- [11] INSPIRE <http://www.ec-gis.org/inspire/>
- [12] GSDI <http://www.gsdi.org/>
- [13] IDEE <http://www.ideo.es/>
- [14] OGC <http://www.opengeospatial.org/>
- [15] GeoAPI <http://geoapi.sourceforge.net/>
- [16] GeoTools <http://www.geotools.org>
- [17] GeoServer <http://www.geoserver.org>
- [18] INEGI, “Sistemas Nacionales Estadístico y de Información Geográfica”,
http://www2.inegi.gob.mx/sneig/contenidos/espanol/pronadeig/resumen_s emarnat.aspx?c=2480
- [19] <http://www.engr.smu.edu>
- [20] <http://www.edtech.uis.edu>
- [21] Infogeotecnologías, “¿Qué es un GIS?”, documento Web
<http://www.geotecnologias.co.cr/Documentos/GIS.pdf>, 26 de febrero de 2005.
- [22] Rubio, Barroso, Alfonso, et. al., “Los sistemas de Información Geográficos: Origen y perspectivas”, Revista General de Información y Documentación, Vol. 7, No. 01, Universidad Complutense, Madrid, 1997.

BIBLIOGRAFIA

1. Alonso, José Manuel, “*Los Retos del Gobierno Electrónico*”, e-government Lead (W3C/CTIC).
2. Alvarez Carrión, G, “Integración de esquemas den bases de datos heterogéneas fuertemente acopladas”, Tesis de Maestría, Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas, Puebla México, Mayo 1999.
3. Amutio, Gómez, Miguel A., “Aplicaciones Web empresariales, Servicios Web y Estándares. Administración electrónica, estándares e interoperabilidad”, Ministerio de Administraciones Públicas, presentación, W3C, España 2007.
4. Arctur, David, et. al., “Designing Geodatabases Case Studies in GIS Data Modeling”, editorial ESRI Press, 2004.
5. Barrena García, Manuel, “Técnicas de particionamiento multidimensional basadas en índices multiatributo en bases de datos paralelas”, Facultad de Informática, Universidad Politécnica de Madrid, Noviembre 1995.
6. Bell, David Arthur, et. al., “Distributed Database Systems”, Addison – Wesley, Agosto 1992, 410 pp.
7. Bertoné, Rodolfo, Alberto, “Métricas de performance en administración de bases de datos distribuidas en redes LAN y WAN”, Tesis de Magister en Ingeniería de Software, Facultad de Informática, Universidad Nacional de la Plata, Argentina, Septiembre 2004.
8. Bourdon, R., et. al., “*Hipótesis y variables en Metodología de las Ciencias Sociales*”, editorial Laisa, Barcelona, Tomo 1, 1973. pp. 47-62.
9. Castillo, Roger, “Especificación de una Arquitectura para DBMS Distribuido Basada en Componentes”, Tesis de maestría, ITESM, 2002.
10. Camps, Paré; Rafael, et. al., “Software libre. Bases de datos”, Formación de Posgrado, Fundació per a la Universitat Obeta de Catalunya, Eureka Media, Barcelona, Mayo 2005.
11. Canché, Maximiliano, “Control de concurrencia basado en Componentes para Bases de Datos Distribuidas”, Tesis de maestría, ITESM, 2002.

12. Ceri, Stefano, et. al., "Distributed Databases - Principles and Systems", McGraw-Hill, 1984, 393 pp.
13. CONAGUA, "La gestión del agua en México", México, D.F. 2006.
14. C. P. Load, et. al., "Concepts and techniques of Geographic Information Systems", Prentice-Hall, E.U., 2002.
15. Date, C. J., "An introduction to Database Systems", editorial Pearson/Addison Wesley, 8a. Edición, 2004, 1005 pp.
16. Digital Planet, "The Global Information Economy", WITSA, <http://www.witsa.org/news/99mar.htm>
17. DeWitt, David J., Gray, Jim; "Parallel Database Systems: The Future of High Performance Database Processing", Computer Sciences Department, University of Wisconsin, Communications of the ACM, Vol. 36, No. 6, Junio 1992.
18. ESRI, "Spatial Data Standards and GIS Interoperability", White Paper, Editorial ESRI Press, Redlands, California, Estados Unidos, Enero 2003.
19. ESRI, "Introduction to the Multiuser Geodatabase", Course version 4.0, Editorial ESRI Press, Redlands, California, Estados Unidos, Mayo 2010.
20. ESRI, "Managing ArcSDE Application Servers. Introducing the ArcSDE application server", Editorial ESRI Press, Redlands, California, Estados Unidos, 2005.
21. Fillotrani, R, Pablo, "Calidad en el desarrollo de software", Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina, 2006.
22. Fishenden, Jerry, et. al., "Interoperabilidad en la Administración Pública. Conseguir la puesta en marcha de e-services", Microsoft Co., Libro Blanco, <http://www.microsoft.com/spain/interop/govt/govteservices.mspx>, Versión 1.0 RTM. Abril 2005.
23. García, Rolando, et. al., "Interdisciplinariedad y Sistemas Complejos", ensayo incluido en E. Leff (comp.) *Ciencias Sociales y Formación Ambiental*. Ed. Gedisa y PNUMA. Barcelona. 1994. pp. 85-123.
24. García Félix, "2C – El Proceso de Medición Software", Departamento de Tecnologías y Sistemas de Información, Escuela Superior de Informática Universidad de Castilla-La Mancha, Ciudad Real, 2006.
25. Guzmán, Tascón, Andrés, "Interoperabilidad de los Documentos Electrónicos (DE), Estado del Arte", Proyecto de Reforma y

- Modernización del Estado Ministerio Secretaría General de la Presidencia, Chile, 2000.
26. Hernández, Sampieri, Roberto, et. al., *“Metodología de la investigación”*, editorial McGraw Hill, quinta edición, México, 2010.
 27. Hoffer H.A., Severance D.G., *“The Use of Cluster Analysis in Physical Data Base Design”*, First International Conference on Very Large Data Bases, Massachusetts, White paper, septiembre 1975.
 28. IGAC, *“Conceptos básicos sobre Sistemas de Información Geográfica y aplicaciones en Latinoamérica”*, Instituto Geográfico Agustín Codazzi, IGAC, 1995.
 29. J. M. Teo, *“Tecnología de los Sistemas de información Geográfica”*, RAMA, 1995.
 30. Judith Domínguez Serrano, *“La gobernanza del agua en México y el reto de la adaptación en zonas urbanas: el caso de la Ciudad de México”*, Centro de Estudios Demográficos, Urbanos y Ambientales, El Colegio de México, 2005
 31. K. C. Clarke, B. O. Parks, and M. P. Crane, *“Geographic Information Systems and environmental modeling”*, Prentice Hall, Indian edition, 2002.
 32. Linthicum, David S., *“Cloud Computing and SOA Convergence in Your Enterprise”*, ed. Addison-Wesley Technology Series, 2010, pp. 229
 33. L. V. da Rocha, N. Edelweiss, et. al., *“Geoframe-t: a temporal conceptual framework for data modeling in GIS '01: Proceedings of the 9th ACM international symposium on Advances in geographic information systems”*, New York, NY, USA, pp. 124–129, ACM Press, 2001.
 34. M. Tamer, Özsu; et. al., *“Principles of Distributed Database Systems”*, 3a. Edición, Springer, Nueva York, 1999, 866 pp.
 35. Malinowski, Elzbieta, *“Fragmentation techniques for distributes object-oriented databases”*, Master of Science Thesis, University of Florida, 1996.
 36. Martínez, Prieto, Ana Belén, *“Un sistema de Gestión de Bases de Datos Orientadas a Objetos sobre una Máquina Abstracta Persistente”*, Tesis Doctoral, Departamento de Informática, Universidad de Oviedo, España, Mayo 2001.
 37. Martos, Rodríguez, Pablo, *“Ejecución de una base de datos distribuida sobre un entorno de Cloud Computing”*, Tesis de Maestría en Ingeniería de Computadores, Facultad de Informática, Universidad Complutense de Madrid, 2009.

38. Medina Cardona, Luis Fernando, "*Hacia una metodología para Ingeniería de requerimientos en Sistemas de Información Geográfica*", Universidad Nacional de Colombia, Facultad de Ingeniería, Bogotá, Colombia.
39. Medina, Ríos, Violeta, "Desarrollo e implementación de un sistema de cómputo distribuido. Aplicación al manejo y operación de bases de datos", Tesis de Maestría, Universidad Michoacana de San Nicolas de Hidalgo, México, Agosto 2008.
40. Meza, Arellano, Antonio, "Análisis de tráfico de datos en sistemas distribuidos", Tesis de Maestría, Instituto Tecnológico de la Paz, División de Estudios de Posgrado e Investigación, México, Junio 2012.
41. Microsoft, "System Administration for Microsoft SQL Server 7.0", Training and Certification, Workbook, course number 832B, 1998.
42. Montaña, Espinosa Carlos Raúl, "Vinculación de Sistemas de Información Gubernamentales para la creación de un Sistema Ejecutivo del Agua basado en un Enfoque Sistémico", Tesis de Maestría en Ciencias del Agua, Instituto Mexicano de Tecnología del Agua, Jiutepec, Mor. México, 2009.
43. Moreno Escobar, Hernán, CEPAL Comisión Económica para América Latina, "*Jornadas de Gobierno Electrónico. Gobierno Electrónico Interoperable para América Latina y el Caribe*", Noviembre 2006, Santiago de Chile.
44. Muñoz, Garcia, Ana Celina, "Modelo Ontológico para la integración de bases de datos", Tesis Doctoral, Centro de Estudios de Microelectrónica y Sistemas Distribuidos, Universidad de Los Andes, 2010.
45. Ochoa Rosso, Felipe, "*Método de los sistemas*", Cuadernos de Planeación de Sistemas de la Facultad de Ingeniería, Editorial UNAM, D.F., México, 2006.
46. Özsu M.T., Valduriez P., "*Principles of Distributed Database Systems*", Prentice Hall, 1991.
47. Pró, Concepción, Luzmila Elisa, "Bases de datos distribuidas usando algoritmos genéticos para optimización de proceso transacción en la Web", Tesis Doctoral, Escuela Universitaria de Postgrado, Universidad Nacional Federico Villarreal, 2010.
48. Rehimí, Saeed K, "Distributed Database Management Systems: A Practical Approach", John Wiley & Sons, 2010, 896 pp.
49. Santana Rogerio, "*E-gov the Brazilian Vision*", Secretary of Logistics

and Information Technology, Ministry of Planning, Budget and Management, Brasil, 2007. Presentación.

50. Schmelkes, Corina, "*Manual para la presentación de anteproyectos e informes de investigación*", Editorial Oxford University Press, D.F., México, 1998.
51. Tertia, Gavin, et. al., "*Stakeholder Participation & Analysis*", DFID Social Development Division, 1994.
52. Tomlinson Roger, "*Thinking About GIS*", 4a. Edición, Editorial ESRI Press, Redlands, California, Estados Unidos, 2011.
53. Vargas, Mendoza, Juan, et. al.; "Fundamentos teóricos de bases de datos distribuidas",