

XXII Congreso Mexicano de Meteorología y VII Congreso Internacional de  
Meteorología, 2013

**Implementación de una bodega de datos sobre recursos hídricos, incluyendo  
ciclones**

Velázquez Alvarez, Jaime; Sanvicente Sánchez Héctor; Yolanda Solís Alvarado; Rivera Benítez, Jaime;  
Patiño Gómez, Carlos; Aparicio Mijares Javier  
René Melgar García; Héctor Manuel López Hernández; Javier García Hernández; Oscar Daniel Peña  
Segura; Gerardo Martínez Castillo; Nadia Sámano Nieves; Isabel García Martínez; René Mendoza  
Betanzos; Luis Enrique Gonzáles Hernández; Eduardo Celis Lais

Instituto Mexicano de Tecnología del agua  
Coordinación de Hidrología, Subcoordinación de Gestión Integrada del Agua  
Paseo Cuauhnáhuac 8532 Col. Progreso, Jiutepec, Morelos C.P. 62550  
E-mail: [jaimewel@tlaloc.imta.mx](mailto:jaimewel@tlaloc.imta.mx)

## **1. Introducción**

El desarrollo de estudios relacionados con los recursos hídricos requiere de datos históricos que permitan conocer la evolución o comportamiento de ciertas variables a lo largo del tiempo y, en algunos casos, su distribución en el espacio. Particularmente, en el caso de ciclones es necesario contar con información histórica que permita conocer tendencias u otros aspectos de interés para un cierto fin. La información histórica debe estar almacenada en una base de datos cuyo diseño permita una consulta eficiente de la información. Asimismo, para que la información sea útil, debe estar actualizada, debe ser consistente y debe ser fácilmente accesible a los usuarios a través de una aplicación creada para este propósito.

Las bodegas de datos (data warehouse en inglés) son bases de datos que almacenan primordialmente información histórica de un tema, y cuyo diseño está orientado a hacer más eficiente la consulta y análisis de la misma. En este trabajo se describe la implementación de una bodega de datos que contiene datos históricos sobre diversos temas relacionados con los recursos hídricos, uno de los cuales fue el de ciclones. Este trabajo fue uno de los resultados de un proyecto que el Instituto Mexicano de Tecnología del agua desarrolló para la Comisión Nacional del Agua, que incluyó el desarrollo de herramientas informáticas para consulta y actualización de información estadística asociada al recurso hídrico, con el fin de apoyar la toma de decisiones en este ámbito. Se describen, entre otras cosas, el universo de datos que se consideró, el diseño de la base de datos y una aplicación Web que se desarrolló para consultar la información.

## **2. Objetivos**

Los objetivos planteados para el trabajo fueron los siguientes:

- Desarrollar un sistema de información de apoyo a la toma de decisiones asociadas a los recursos hídricos, compuesto por:

- una bodega de datos conteniendo datos estadísticos históricos asociados a diversos temas, incluyendo la ocurrencia de ciclones en México.
- una base de datos geográfica conteniendo las capas geográficas asociadas a los diferentes temas.
- una aplicación web para consulta y actualización de los datos.

### 3. Metodología

El sistema informático descrito en este trabajo es parte del Sistema Nacional sobre cantidad, calidad, usos y conservación del Agua (SINA), el cual está establecido en la Ley de Aguas Nacionales (LAN) y cuya implementación está a cargo de la Comisión Nacional del Agua. El SINA, según la LAN, debe concentrarse en el procesamiento e integración de la información que sea relevante para el diseño y evaluación de las políticas públicas en el tema del agua. Es decir, el SINA es un sistema concebido para apoyar la planeación y programación hídrica.

La información que describe la situación del agua en México puede representarse en una pirámide, mostrada en la figura 1. Como ahí se observa, la situación del agua debe ser establecida desde el punto de vista del desarrollo sustentable considerando aspectos económicos, sociales y ambientales, así como los aspectos institucionales de todas las dependencias del sector.

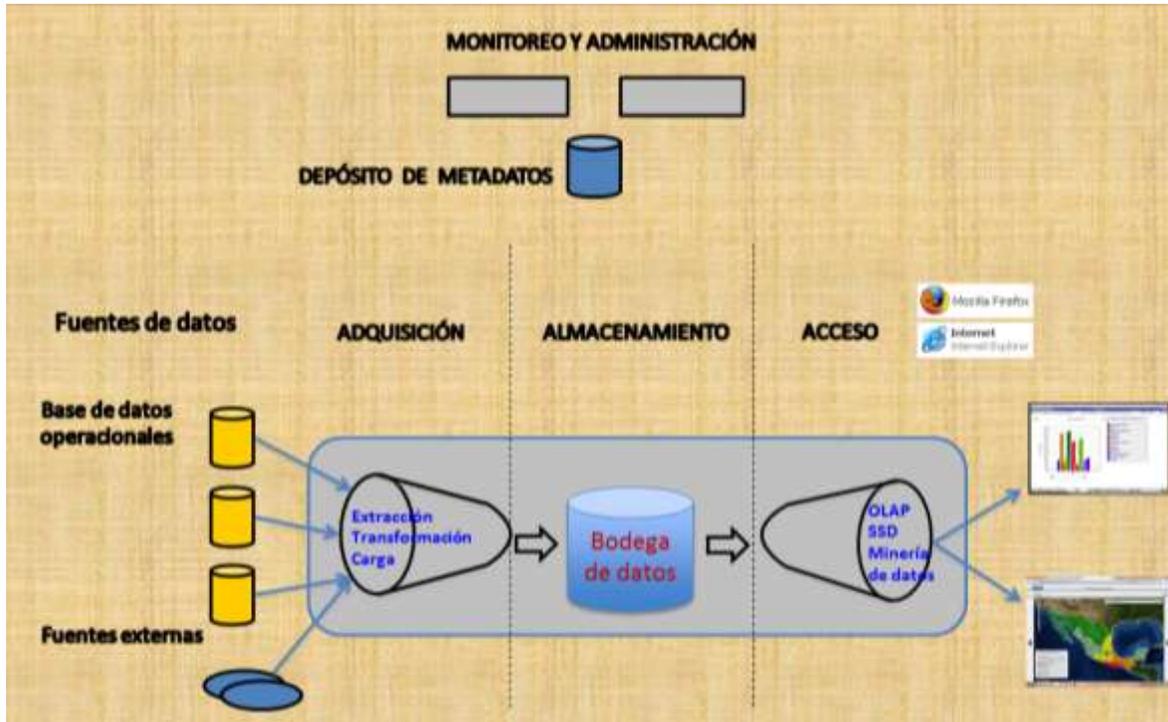


**Figura 1.** Pirámide de información de la situación del agua en México

La base de la pirámide es integrada por datos primarios o básicos que sirven para propósitos de estudios técnicos y proyectos de investigación; estos datos son generados por las áreas técnicas y se almacenan y administran en sistemas específicos dentro de éstas áreas (por ejemplo: CLICOM, BANDAS, SIGMAS, etc.). El segundo nivel de la pirámide es información ya procesada y que está destinada a directivos y analistas. Finalmente, el nivel superior está compuesto por indicadores, los cuales están destinados a tomadores de decisiones y público en general.

El SINA es un sistema que se enfoca en el nivel intermedio y superior de la pirámide de información, en donde los datos poseen un valor agregado para la toma de decisiones y planificación hídrica.

Por otro lado, una bodega de datos es una base de datos históricos, resultante de la integración de datos provenientes de diversas fuentes, orientada a un tema específico, cuyo diseño está enfocado a soportar aplicaciones que permiten realizar análisis, descubrimiento de conocimiento y toma de decisiones. Los componentes básicos asociados a una bodega de datos se muestran en la figura 2.



**Figura 2.** Componentes asociados a una bodega de datos

De acuerdo a la figura 2, las fuentes de datos que alimentan a una bodega de datos pueden ser bases de datos ya operando o fuentes externas (por ejemplo, archivos de texto o en otros formatos). Un componente asociado a la bodega de datos es el de adquisición, el cual se compone de un conjunto de herramientas que permiten la adquisición de datos de las diversas fuentes, mediante un proceso de extracción, transformación y carga de los datos al data warehouse. Este conjunto se denomina comúnmente herramientas ETL (por Extraction, Transformation and Loading).

El segundo componente es el de almacenamiento, el cual está constituido propiamente por la base de datos histórica o data warehouse. Un tercer componente es el de acceso, el cual se constituye por herramientas para análisis de los datos, por ejemplo herramientas OLAP (On-Line Analytical Processing), sistemas de soporte a las decisiones (SSD, por sus siglas en inglés) o herramientas de minería de datos. Un cuarto componente está constituido por herramientas de monitoreo y administración que permiten conocer por ejemplo los temas consultados, operaciones realizadas sobre la base de datos y el estado actual de la actualización de datos; las herramientas de administración permiten dar de alta usuarios y conceder permisos, dar de alta nuevos temas, etc. Finalmente, un quinto componente es el depósito de metadatos, en donde se almacena la información que describe a los datos almacenados en el data warehouse (fuente, fecha de inserción, persona que insertó el dato, etc.) y descripción de las tablas que componen el data warehouse.

A continuación se describen las actividades desarrolladas en este trabajo. La primera consistió en definir los datos estadísticos y geográficos que se requerían almacenar en la base de datos histórica y geográfica respectivamente, y elaborar el modelo conceptual conteniendo ambos tipos de datos. Posteriormente se desarrollaron las actividades encaminadas a implementar la bodega de datos y las componentes asociadas a ésta, los cuales se muestran en la figura 2.

### 3.1 Definición de requerimientos de datos geográficos y estadísticos, y elaboración del modelo conceptual de la base de datos

El primer paso en el desarrollo del sistema de información fue determinar el conjunto de datos estadísticos y geográficos que comúnmente se requieren en la planificación hídrica y toma de decisiones en este sector. Para este fin, se revisaron diversos documentos como las Estadísticas del agua en México, Estadísticas del agua a nivel de región hidrológico-adminva, Plan Nacional Hídrico y otros documentos. El resultado de esta actividad fue un documento que contenía el listado de temas, variables e indicadores reportados en esos documentos, donde se indicaban sus unidades, resolución temporal, fuente y otros aspectos. Asimismo, se obtuvo la lista de capas geográficas que están asociadas a cada tema.

Una vez que se identificó el conjunto de datos, se procedió a definir una forma de organizarlos con el fin de elaborar un modelo conceptual de la base de datos y así tener claridad, tanto para los usuarios como para los desarrolladores del sistema, durante todo el desarrollo del proyecto. Los datos fueron organizados de acuerdo a los 3 ejes del desarrollo sustentable: ambiental, económico y social. A su vez, cada eje se dividió en los siguientes sistemas:

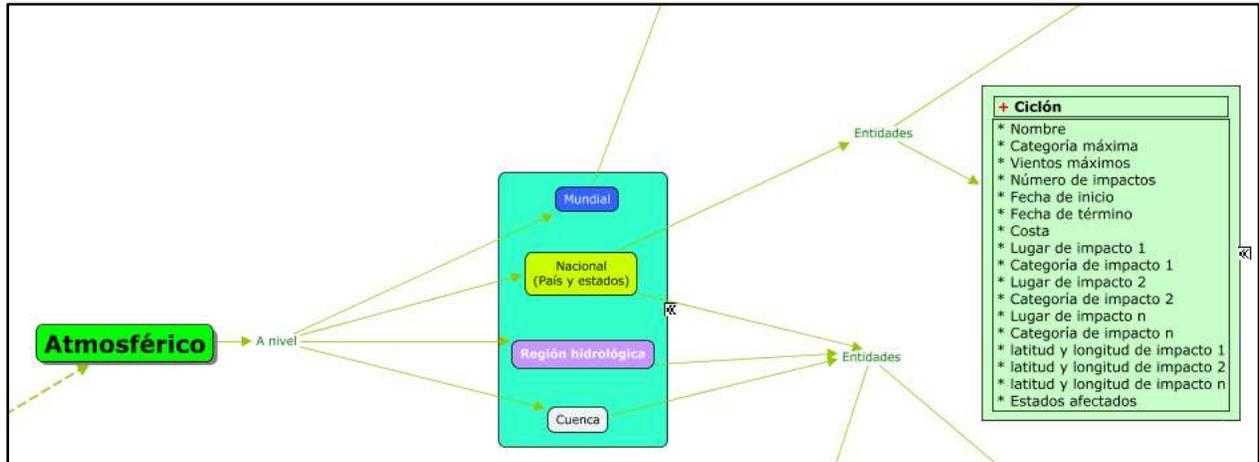
| <b>Eje</b> | <b>Sistemas</b>                      |
|------------|--------------------------------------|
| Ambiental  | Acuático                             |
|            | Atmosférico                          |
|            | Terrestre                            |
| Económico  | Productivo                           |
|            | Financiero                           |
| Social     | Normativo, de organización y gestión |
|            | Desarrollo humano                    |

Posteriormente se procedió a realizar un análisis de todos los datos que se identificaron como necesarios en la planificación y se asignó cada dato estadístico y geográfico al eje que le correspondía. Por otro lado, para tomar en cuenta el componente geográfico en la organización de los datos se consideró la cuenca como la unidad básica de planeación de los recursos hídricos (como está especificado en la LAN), la cual a su vez es parte de una región hidrológica, la cual es parte de un país y el cual es parte del ámbito mundial.

Finalmente, se procedió a elaborar un modelo conceptual de la base de datos usando el software CMap en donde se plasmaron los ejes, los sistemas y los datos, tanto geográficos como estadísticos, que fueron considerados en cada sistema.

En el caso de los ciclones, esta entidad de información se colocó en el eje ambiental y en el sistema atmosférico, sus datos se tienen a nivel nacional y algunos atributos que se consideraron fueron su nombre, categoría máxima y vientos máximos alcanzados durante su ocurrencia en el país, fecha de inicio, fecha de término, costa donde se originó (océano Pacífico, océano Atlántico, mar Caribe), datos de los lugares donde impactó (localidad,

categoría, latitud, longitud) y estados afectados. La figura 3 muestra la sección del mapa mental que corresponde correspondiente a los datos de ciclones.



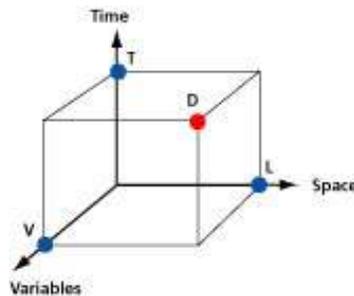
**Figura 3.** Sección del modelo conceptual de la base de datos correspondiente a los datos de ciclones ocurridos en el país

### 3.2 Implementación de la bodega de datos

#### 3.2.1 Elaboración del modelo lógico de la bodega de datos

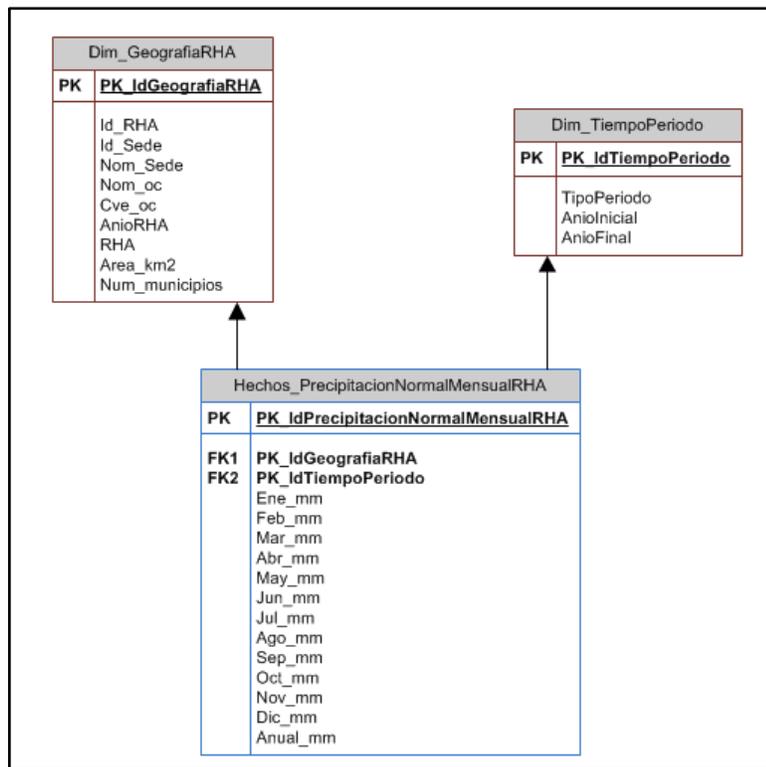
Para la elaboración del modelo lógico de la bodega de datos se utilizó el modelo dimensional, el cual es el modelo comúnmente usado en este tipo de desarrollos. Un cubo de datos es una técnica utilizada en sistemas de información, en la cual el usuario percibe los datos almacenados en una base de datos como un cubo, en donde cada eje que conforma el cubo representa una dimensión a lo largo de la cual se desean analizar los datos. En la figura 4, mostrada abajo, se presenta un cubo de datos con tres dimensiones: tiempo, espacio y variable. Un punto en el cubo representa un dato medido en una cierta fecha, un lugar y de una cierta variable.

Los cubos de datos proporcionan muchas ventajas, entre ellas la flexibilidad para consultar los datos, simplicidad del modelo lógico subyacente a los cubos y la posibilidad de programar interfaces intuitivas para su consulta por los usuarios.



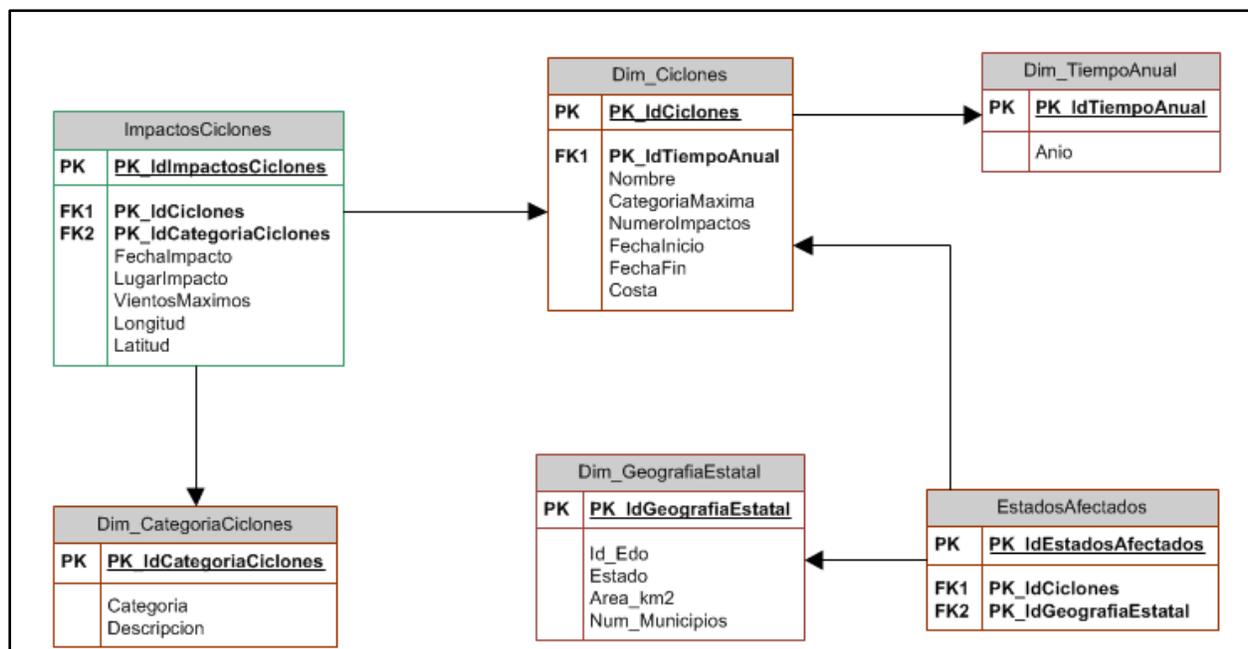
**Figura 4.** Modelo dimensional de datos. Tomado de (Maidment, 2002).

El modelo dimensional de datos produce modelos lógicos en forma de estrella, en donde existe una tabla principal que almacena los datos históricos de interés (denominada comúnmente “tabla de hechos”) y tablas de dimensiones conectadas a esta tabla de hechos, los cuales determinan el contexto en el cual ocurrieron esos datos, por ejemplo la fecha, el lugar geográfico y la variable. En la figura 5 se muestra un modelo lógico correspondiente a los datos de precipitación normal mensual, considerando el modelo dimensional de datos. Se tienen ahí dos dimensiones: la espacial, representada por la región hidrológico-admva. (RHA) a la cual pertenecen los datos y la dimensión temporal, representada por el periodo de años que se utilizó para el cálculo de la normal. La tabla de hechos en este caso contiene el valor de la normal de precipitación para cada mes, ocurrida en una RHA y calculada para un periodo determinado.



**Figura 5.** Ejemplo del modelo lógico correspondiente a datos de precipitación normal mensual, usando el modelo dimensional de datos

El modelo lógico para los datos de ciclones se muestra en la figura 6. Como ahí se puede ver, se definieron dimensiones para la entidades de ciclones, estados y categorías que el ciclón puede tener (H1,H2,H3, etc.), las cuales son las tablas Dim\_Ciclones, Dim\_GeografiaEstatal y Dim\_CateogriasCiclones respectivamente; estas dimensiones finalmente constituyen catálogos en la base de datos. Se definió también una dimensión para almacenar los años, representada por la tabla Dim\_TiempoAnual, la cual se utiliza para indicar el año de ocurrencia del ciclón.



**Figura 6.** Modelo lógico de datos correspondientes a ciclones, usando el modelo dimensional de datos

### 3.2.2 Creación de la bodega de datos

La base de datos dimensional de la bodega de datos fue creada con el software SQL Server 2008 en un servidor de cómputo que se dedicó para este fin. Inicialmente se había considerado que la base de datos dimensional estuviera separada de la base de datos geográfica, sin embargo se optó por implementar una sola base de datos conteniendo ambas. Esto se hizo para facilitar la consulta de la información y para facilitar el respaldo de los datos.

### 3.3 implementación del componente de acceso y visualización de la información

En esta parte del proyecto se desarrolló una aplicación Web para acceder y consultar la información. Se consideró que las consultas deberían poder visualizar el comportamiento de una variable a lo largo del tiempo y su distribución en el espacio. Particularmente, en el caso de los ciclones se consideró dentro de las consultas el poder visualizar, para un año o un periodo de años, lo siguiente:

- un resumen de los ciclones tropicales que han impacto a México, agrupados por costa donde se originaron (atlántico, pacífico) y por tipo (Depresión tropical, Tormenta tropical, Huracán moderado (H1 y H2), Huracán intenso (H3 a H5).
- lista de huracanes que han impactado a México en ese año o periodo.
- Un desglose del número de ciclones tropicales que han impactado a México por cada año del periodo de interés y por categoría.

Adicionalmente se consideró en la aplicación de consulta la opción de mostrar un mapa con la ubicación de los huracanes intensos ocurridos en México para todo el periodo de datos

contenido en la base de datos del data warehouse. La aplicación se desarrolló con un software de inteligencia de negocios (Business intelligence, por sus siglas en inglés) denominado Web Focus, el cual permite el acceso a los datos y la creación de reportes, ya sea en forma de tabla o gráfica. Asimismo, permite la conexión a los servicios de capas geográficas, creados con el software ArcGIS Server, y la creación de mapas temáticos combinando estos servicios con los datos obtenidos del data warehouse.

### 3.4 Implementación del componente de adquisición de datos (herramientas ETL)

En esta parte del trabajo se desarrollaron programas de cómputo para extraer, transformar y cargar datos a la bodega de datos, lo que se denomina comúnmente herramientas ETL. Se acordó que los datos con los que se actualizaría la bodega, y los metadatos que los describen por ejemplo la fuente de donde provienen, estarían almacenados en archivos de Excel con una cierta estructura. Se definieron también reglas de validación para asegurarse que los datos insertados fueran correctos, por ejemplo que estuvieran en rangos de valores válidos. Se definieron también reglas de actualización en donde se indica en cuales tablas y campos de la bodega de datos se deben insertar los datos. Las herramientas fueron elaboradas con el software Integration Services, el cual es parte del sistema administrador de base de datos SQL Server 2008.

Al momento de ingresar datos al data warehouse, las herramientas ETL leen los metadatos también y los insertan en las tablas que fueron creadas para almacenarlos.

### 3.5 Implementación de herramientas de monitoreo y administración

En esta actividad se elaboraron procedimientos con la plataforma de desarrollo Visual Studio 2010, usando la tecnología .NET, para monitorear el avance en la actualización de los diferentes temas que componen la bodega de datos, administrar usuarios y actualizar catálogos necesarios en el sistema.

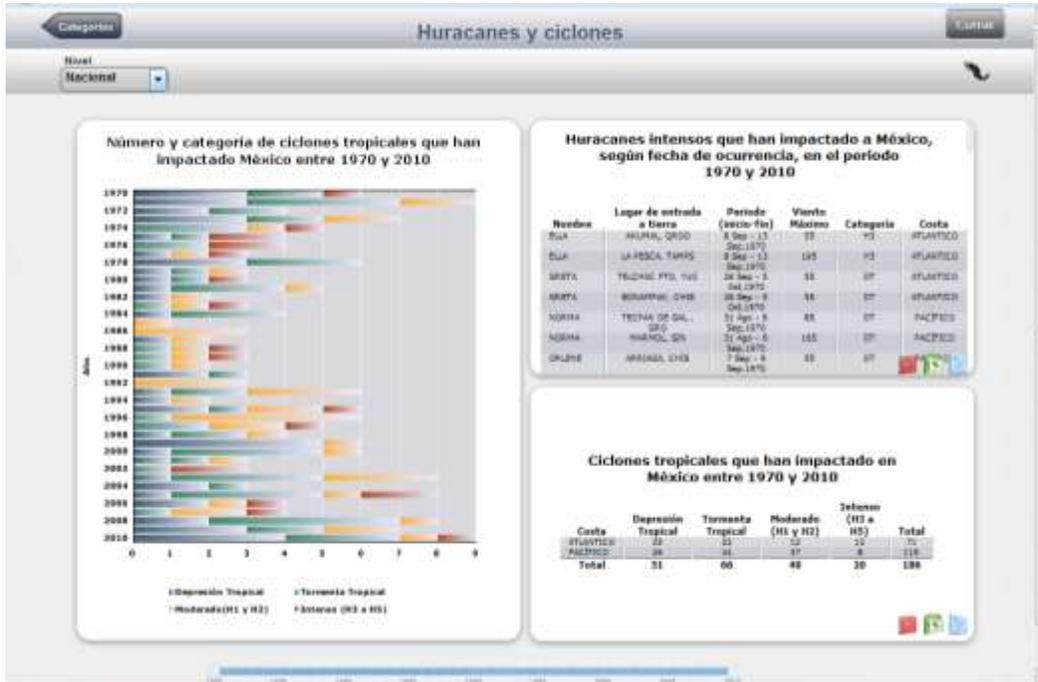
### 3.6 Implementación de la base de datos geográfica

Para implementar esta base de datos se elaboró primero, a partir del modelo conceptual de la base de datos, un modelo lógico usando la técnica de modelación orientada a objetos; el modelo fue elaborado con el lenguaje UML y las clases que ofrece el software ArcGIS. Posteriormente se creó una base de datos de tipo corporativa usando el software ArcSDE y SQL Server 2008, y se cargaron todas las capas geográficas. Finalmente se crearon los servicios de mapas necesarios con el software ArcGIS Server.

## 4. Resultados

En esta sección se muestran algunas de las consultas que pueden realizarse sobre los datos de ocurrencia de ciclones en México, utilizando la aplicación que se desarrolló en este trabajo. Como se comentó antes, se consideró que la aplicación permitiera mostrar información en forma tabular y gráfica, y en mapas.

La figura 7 muestra la ventana de diálogo de la aplicación conteniendo 3 reportes: una gráfica donde se muestra el número y categoría de ciclones que ocurrieron en un periodo de interés para el usuario, una lista de los huracanes que se presentaron en ese mismo periodo y un resumen de los ciclones ocurridos también en ese periodo. La ventana de diálogo cuenta con una barra en la parte inferior, donde el usuario especifica el período que le interesa visualizar.



**Figura 7.** Reportes a nivel nacional que pueden visualizarse con la aplicación de consulta de datos del data warehouse.

Por otro lado, se programó un módulo para visualizar en un mapa la ubicación de los huracanes ocurridos en el país, clasificados por categoría. Dicho mapa fue implementado usando el conector que ofrece WebFocus para acceder los servicios de mapas creados con el software ArcGIS server. La figura 8 muestra este mapa.



**Figura 8.** Ubicación de los huracanes ocurridos en un periodo de años en el país, clasificados por categoría.

## 5. Conclusiones y recomendaciones

El desarrollo de este trabajo permitió integrar en una sola base de datos, los datos históricos tanto, estadísticos como geográficos, asociados a diferentes temas relacionados con los recursos hídricos en México. Los datos estadísticos fueron almacenados usando el modelo dimensional utilizado que se utiliza en bodegas de datos. Entre otras ventajas, este modelo es que ofrece mayor flexibilidad para realizar las consultas, computacionalmente hace las consultas más eficientes y produce un modelo lógico que es más simple e intuitivo a los usuarios y desarrolladores del sistema.

En el caso del tema de ciclones se almacenaron diferentes variables que caracterizan su ocurrencia en el territorio mexicano. El periodo para el cual se almacenaron datos de este tema fue 1970 – 2010, pero se implementaron las herramientas para que los datos puedan ser actualizados por personal de la CONAGUA a cargo de esa tarea. La interfaz que se desarrolló para acceder al data warehouse, permite la consulta de los datos sólo a nivel nacional, sin embargo, la estructura de la base de datos permite la consulta a diferentes niveles de agregación por ejemplo región hidrológico-adminva., entidad federativa o municipio. Es recomendable en un futuro agregar a la interfaz esta funcionalidad, con el fin de ofrecer a los usuarios mayor versatilidad para conocer la ocurrencia de ciclones en el país.

## 6. Bibliografía

- a. Sanvicente S., Héctor, Solís A. Yolanda, Velázquez A. Jaime y Rivera B. Jaime, "Planeación regional para la sustentabilidad hídrica en el mediano y largo plazos en las regiones hidrológico-administrativas. Actividad 4.5 Desarrollo del soporte informático del Sistema Nacional de Información del Agua (SINA) e incorporación de la información disponibles, incluyendo las curvas de costo generadas". Informe final. Noviembre de 2012.
- b. Velázquez A. Jaime, Diseño e implementación de una bodega de datos climatológicos. Tesis de maestría en ciencias de la computación, ITESM Campus Morelos, 2002.
- c. Maidment, David R., ArcHydro GIS for water resources. ESRI Press. Redlands, California, USA, 2002.