



Universidad de las Ciencias Informáticas

Facultad 8

**Trabajo de Diploma para optar por el título de  
Ingeniero en Ciencias Informáticas**



Análisis, diseño e implementación del mercado de datos del  
Indicador Medio Ambiente para el almacén de datos de la  
Oficina Nacional de Estadísticas

**Autores**

Taimy Borrell Madrazo

Yusdaily Gómez Romero

**Tutor**

Ing. Nilber Barbán Góngora

Ciudad de La Habana, junio de 2010

“Año 52 de la Revolución”



*"El mundo camina hacia la era electrónica... Todo indica que esta ciencia se constituirá en algo así como una medida del desarrollo; quien la domine será un país de vanguardia. Vamos a volcar nuestros esfuerzos en este sentido con audacia revolucionaria"*

*de*

## **INTRODUCCIÓN**

La informática en Cuba, por su crecimiento y expansión, ha venido transformando rápidamente la sociedad actual; en sus inicios sólo facilitaba los trabajos repetitivos y monótonos del área administrativa, gracias a las grandes innovaciones que se han hecho en este campo y con las nuevas tecnologías que han surgido, ha traído como consecuencia el incremento de la producción.

Esta disciplina se aplica a numerosas y variadas áreas del conocimiento o a la actividad humana, como por ejemplo: gestión de negocios, almacenamiento y consulta de información, comunicaciones, control de transportes, investigación, desarrollo de juegos, diseño computarizado, aplicaciones/herramientas multimedia, medicina, biología, física, ingeniería, arte, entre otros.

Una de las aplicaciones más importantes de la informática es proveer información oportuna y veraz, lo cual puede tanto facilitar la toma de decisiones en una empresa como permitir el control de procesos críticos.

Muchas empresas y compañías cubanas o sucursales de transnacionales como la compañía telefónica, empresas jaboneras, firmas comerciales, bancos, emplean sistemas en tareas que implican procesar grandes masas de datos contables y de administración.

Además de las empresas mencionadas, la Oficina Nacional de Estadísticas (ONE) en Cuba está inmersa en todo un proceso de informatización, puesto que necesita un sistema adecuadamente estructurado, que gestione la información referente a los modelos estadísticos de una forma más rápida y accesible para todos los sectores de la sociedad.

La Oficina Nacional de Estadísticas se crea en 1994 con el objetivo de captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país. Para ello tiene una serie de Indicadores estadísticos como son: la Salud, el Turismo, la Industria, la Educación, entre otros.

Debido a los grandes cambios ocurridos en los ámbitos demográficos y ecológicos, el país ha adoptado la medida de llevar estadísticas del Indicador Medio Ambiente con el fin de poner en manos de investigadores y la población en general, una recopilación de los principales datos ambientales, fundamentalmente del período 1990-2008, para brindar una visión global del estado del Medio Ambiente y de nuestros recursos naturales, la presión que se ejerce sobre ellos y las acciones que toma el gobierno y la sociedad para protegerlo y preservarlo. El Indicador Medio Ambiente cuenta con 54 series estadísticas los cuales contienen información sobre las condiciones físicos-geográficas donde se asientan los recursos naturales y ambientales.

Todos los datos relacionados son almacenados en formato Excel, por lo que se hace muy complejo el proceso de acceder. La información es mostrada en tablas, por lo que la manera de divulgarla, no es la deseada. A los trabajadores les dificulta realizar los principales reportes, cruces de variables, tasas y porcentajes porque la forma de almacenar, recuperar y presentar la información es compleja, afectando así la disponibilidad de la información y el proceso de toma de decisiones.

La necesidad creciente de contar con información organizada para la toma de decisiones para Órganos del Estado y satisfacer la demanda de información pública respecto a las estadísticas del Medio Ambiente, son dos causas para perfeccionar el sistema y proveerle una mejor forma de trabajo a los especialistas en la ONE.

Por lo anteriormente planteado surge la siguiente pregunta: ¿Cómo centralizar la información de los Indicadores del Medio Ambiente que facilite el acceso y la divulgación de dicha información en la Oficina Nacional de Estadística para la adecuada toma de decisiones a corto plazo?

Para dar solución a la problemática antes mencionada se ha definido como **objetivo general** realizar el análisis, diseño e implementación del mercado de datos de los Indicadores del Medio Ambiente para la Oficina Nacional de Estadísticas.

El **objeto de estudio** es el almacén de datos.

Los **objetivos específicos** de este proceso investigativo son:

1. Elaborar el marco teórico de la investigación acerca de las principales tendencias de implementación de los almacenes de datos y los mercados de datos.
2. Realizar el análisis y diseño de los Indicadores del Medio Ambiente para el almacén de datos de la Oficina Nacional de Estadísticas.
3. Implementar y cargar los clasificadores para el mercado de datos de los Indicadores del Medio Ambiente para el almacén de datos de la Oficina Nacional de Estadísticas.
4. Validar la solución desarrollada mediante la realización de pruebas.

El **campo de acción** estará centralizado en el mercado de datos estadísticos para los Indicadores del Medio Ambiente.

La **idea a defender** será: con el desarrollo del mercado de datos del Indicador Medio Ambiente se podrán almacenar y recuperar los datos provenientes del Indicador de forma rápida y eficiente, permitiéndole a la Oficina Nacional de Estadísticas en Cuba, realizar los principales reportes y divulgar dicha información.

A continuación se proponen las **tareas** a realizar por cada objetivo planteado anteriormente:

*Elaborar el marco teórico de la investigación.*

1. Realización de un estudio del estado del arte del tema Indicadores del Medio Ambiente.
2. Identificación de las tecnologías existentes en Cuba y el mundo que estén relacionadas con el desarrollo del mercado de datos. (Histórico-Lógico)
3. Realización del estudio de las metodologías y herramientas a utilizar en el desarrollo de la investigación. (Analítico-Sintético)

*Realizar el análisis y diseño de los Indicadores del Medio Ambiente.*

4. Planificación de entrevistas a especialistas en el tema en la ONE. (Entrevistas)
5. Elaboración del levantamiento de los requisitos funcionales y no funcionales. (Observación)
6. Modelación de los requerimientos. (Modelación)
7. Elección de la granularidad del proceso del negocio.
8. Definición de las dimensiones del mercado de datos y los hechos asociados. (Analítico- Sintético)
9. Estructuración del modelo dimensional para transformarlo al diseño físico.

*Implementar y cargar los clasificadores para el mercado de datos de los Indicadores del Medio Ambiente.*

10. Construcción de la base de datos.
11. Montaje de los clasificadores para el mercado de datos de los Indicadores del Medio Ambiente para el almacén de datos de la Oficina Nacional de Estadísticas.

*Validar la solución desarrollada.*

12. Realización de las pruebas de implantación al mercado de datos de los Indicadores del Medio Ambiente.

Para el desarrollo del trabajo se utilizarán los siguientes **métodos científicos** que guiarán el estudio de cada tema y posibilitarán arribar fácilmente a la solución.

## Métodos teóricos

Analítico-Sintético: Se realizará un estudio profundo de las metodologías y herramientas a utilizar en el desarrollo de la investigación basándose en la revisión de documentos, artículos e informes. Además se analizarán las dimensiones y los hechos asociados.

Histórico-Lógico: Será utilizado para analizar la trayectoria de los sistemas existentes relacionados con mercado de datos de los Indicadores del Medio Ambiente.

Modelación: Es considerado de gran importancia para el diseño de la base de datos, por lo que se utilizará para la modelación de los requerimientos el cual demostrará con claridad la forma que se va a presentar la solución.

## Métodos empíricos

Entrevistas: El uso de este método permitirá planificar una conversación con especialistas de la Oficina Nacional de Estadísticas con el objetivo de conocer cómo se comporta el Indicador Medio Ambiente.

Observación: Se utilizará para identificar las funcionalidades del mercado de datos de los Indicadores del Medio Ambiente, con el propósito de definir los requisitos funcionales y no funcionales de la base de datos a implementar.

El presente trabajo está estructurado en tres **capítulos**:

Capítulo 1: Fundamentación teórica.

En este capítulo se describirá los principales conceptos relacionados con el tema. Se realizará un estudio del estado del arte acerca del mercado de datos del Indicador Medio Ambiente para comprobar cuál es la solución factible y similar al sistema a desarrollar. Se estudiará las metodologías y herramientas más utilizadas, analizando sus características, ventajas y desventajas con el propósito de proponer la adecuada para el desarrollo de la investigación.

Capítulo 2: Análisis y Diseño

El capítulo se centra en definir los temas de análisis detallado del funcionamiento interno de la base de datos. Se describirá la definición del negocio centralizado en el Medio Ambiente. Se examinarán las series correspondientes a este Indicador para identificar los requisitos funcionales y no funcionales, además de los requisitos de información y los multidimensionales, luego se definirán las tablas de hechos, las

medidas y las dimensiones para estructurar el modelo dimensional de la solución. Se realizará la matriz BUS para obtener una mejor visión de las relaciones de los hechos y las dimensiones.

### Capítulo 3: Implementación y Prueba

En este capítulo se presenta una explicación de cómo se lleva a cabo el proceso para la implementación del procedimiento propuesto en el Trabajo de Diploma. Se realizará el montaje de los clasificadores para el mercado de datos y se validará la solución desarrollada mediante la realización de las pruebas de implantación, las listas de chequeo y la aceptación por el cliente.

## **CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA**

### **Introducción**

Las tecnologías han cambiado la forma en que operan las organizaciones actuales. A través de su uso se logran importantes mejoras, pues automatizan los procesos operativos, suministran una plataforma de información necesaria para la toma de decisiones.

La informática, está sumida en la gestión integral de la empresa y por eso las normas y estándares propiamente informáticos deben estar sometidos a la misma. La informática no gestiona propiamente la empresa, ayuda a la toma de decisiones, pero no decide por sí misma. Marca una pauta en el desarrollo de esta área laboral, por lo que la ONE ha tenido que recurrir a la Informática para conseguir buena organización de la información, pero el logro de ese objetivo depende fundamentalmente de su arquitectura actual, tanto de hardware como de software.

El almacén de datos o Data Warehouse (DW) es actualmente el centro de atención de esta institución, porque provee un ambiente para que la misma, haga un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales. Las aplicaciones para soporte de decisiones basadas en un Data Warehouse, pueden hacer más práctica y fácil la explotación de datos para una mayor eficacia del negocio. (An. Luis Walter Reynoso Socio Fundador y Aldo Marcelo Socio Gerente, 2005)

### **1.1 Conceptos asociados al tema**

Para la profundización de la temática, es preciso aludir ciertos conceptos que ayudan a un mejor entendimiento de los aspectos relacionados al tema. Los mismos se exponen a continuación.

#### **1.1.1 Almacén de datos**

Un almacén de datos es una colección de datos que recoge información de múltiples sistemas fuentes u operacionales dispersos y cuya actividad se centra en la “toma de decisiones”. Una vez reunidos los datos de los sistemas fuentes se guardan durante mucho tiempo, lo que permite el acceso a datos históricos; así los almacenes de datos proporcionan al usuario una interfaz consolidada única para los datos, lo que hace más fácil escribir las consultas para la toma de decisiones. (María Evelia Casales Cabrera, 2009)

Un almacén de datos se caracteriza por ser un conjunto de datos orientados a temas, integrado, variante en el tiempo y no volátiles, que dan soporte al proceso de toma de decisiones.

1. **Orientado a temas:** orientado a los temas principales de la organización. La información se clasifica en función de los aspectos que son de interés para la organización. Se organizan por temas para facilitar el acceso y el entendimiento por parte de los usuarios finales. Por ejemplo, se tiene información que proviene de los datos operacionales de distintas áreas como son: préstamos, ahorros, nóminas y créditos. Pues los temas que son de interés son: clientes, productos, vendedores y actividad.

2. **Integrado:** en un sistema operacional no hay integración, pero en un almacén de datos todos los datos se encuentran integrados. Esta integración se muestra de diferentes maneras: consistencia en codificación de estructuras, en unidades de medida de las variables, en múltiples fuentes, en convenciones de nombres, en atributos físicos de los datos, entre otros.

3. **Variante en el tiempo:** en los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente, por el contrario, la información almacenada en el almacén de datos refleja un histórico de los datos en un horizonte de tiempo mucho más amplio (del orden de años). De esta forma se puede detectar tendencias en el tiempo.

4. **No volátil:** el almacén de datos existe para ser leído y no para ser modificado, la información es por tanto permanente, la actualización del almacén de datos significa la incorporación de los últimos valores que tomaron las distintas variables contenidas en él, sin ningún tipo de acción sobre lo que ya existía. Los datos más recientes no sustituyen a los precedentes, sino que suelen mantenerse con un mayor nivel de detalle. Por ejemplo: en el año actual se guardarían las ventas diarias de los diferentes artículos, mientras que los datos anteriores se mantendrían accesibles de manera agregada. (Ana Buigues, 2010)

El almacén de datos ha venido a poner solución al problema que surge como consecuencia de que muchas aplicaciones tengan diferentes versiones de la misma información distribuida en datos con diferentes formatos. El almacén de datos sirve como soporte de integración para esta problemática. Cabe destacar que no es un producto y por tanto no puede comprarse, debe ser construido paso a paso. Existen aplicaciones comerciales y productos que pueden personalizarse en cierta medida para un problema concreto, no hay un producto comercial actualmente, que se adapte a las necesidades de cualquier problema de esta índole. El hecho de que haya que construirlo implica que requiera un período de tiempo, que variará en función de las necesidades de la organización, hasta que el sistema esté en producción. (Juan Luis Martínez Albuerno, 2008)

La arquitectura de un almacén de datos está formada por tres componentes básicos que son:

- El proceso de extracción, transformación y carga de los datos que permite mover datos desde múltiples fuentes tomando aquellos que son de utilidad para el almacén de datos, una vez que los datos son extraídos, éstos se transforman a una forma presentable y de valor para los usuarios, al terminar se cargan los datos en el almacén de datos.
- Los procesos de gestión de datos, encargados de almacenar y procesar los datos en el almacén de datos.
- Los servicios de consulta y acceso a la información almacenada. (Juan Luis Martínez Albuerne, 2008 )

El almacén de datos surgió como respuesta a las necesidades de extraer información y conocimiento de los datos, que se utilizaría como apoyo en la toma de decisiones de las organizaciones. El principal objetivo es hacer que la información sea lo más accesible posible así como ayudar a los usuarios a encontrar la información que necesitan. El modelo en que se recopilan los datos en el almacén de datos se conoce como modelo dimensional. Este modelo provee un acceso sencillo y eficiente a los datos.

### **1.1.2 Mercado de datos**

El mercado de datos o data marts es un subconjunto de datos de un almacén relativo a los requisitos de un departamento o área de negocio concreto. Este subconjunto de datos puede funcionar de forma autónoma, o bien enlazado al almacén de datos. El motivo por el cual se crean mercados de datos es el crecimiento que tiene el almacén y así facilitar su construcción y utilización. (María Evelia Casales Cabrera, 2009). Las características de los mercados de datos son:

- Se centran en los requisitos de los usuarios asociados a un departamento o área de negocio concreto.
- Como diferencia con los almacenes de datos, los mercados no contienen datos operacionales detallados.
- Son más sencillos a la hora de utilizarlos y comprender sus datos, debido a que la cantidad de información que contienen es mucho menor que en los almacenes de datos. (María Evelia Casales Cabrera, 2009)

¿Qué diferencia existe entre un mercado de datos y un almacén de datos? Su alcance. El mercado de datos está hecho para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización. Es el almacén natural para los datos departamentales. En cambio, el ámbito del

almacén de datos es la organización en su conjunto. Es el almacén natural para los datos corporativos comunes. (Josep Curto, 2007)

Para crear el mercado de datos de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP (Procesamiento de Transacciones en Línea) como el propio almacén de datos o sobre una base de datos OLAP (Procesamiento Analítico en Línea). La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento. De esta forma se pueden plantear dos tipos de mercados de datos:

**OLTP:** el procesamiento transaccional en tiempo real constituye el trabajo primario en un sistema de información. Este trabajo consiste en realizar transacciones, es decir, actualizaciones y consultas a la base de datos con un objetivo operacional: hacer funcionar las aplicaciones de la organización, proporcionar información sobre el estado del sistema de información y permitir actualizarlo conforme va variando la realidad del contexto de la organización. Muestras de este tipo de trabajo transaccional en una empresa es: la inserción de un nuevo cliente, el cambio de sueldo de un empleado, la tramitación de un pedido, el almacenamiento de una venta, la impresión de una factura, entre otros. Una de las desventajas de OLTP es que proporciona capacidades limitadas para la toma de decisiones. (Leandro, 2009)

**OLAP:** los datos son clasificados en diferentes dimensiones y pueden ser vistas unas con otras en diferentes combinaciones para obtener diferentes análisis de los datos que contienen. En este modelo los datos son vistos como cubos los cuales consisten en categoría descriptivas (dimensionales) y valores cuantitativos (medidas). El modelo multidimensional de datos simplifica a los usuarios formular consultas complejas, arreglar datos en un reporte y cambiar de datos resumidos a datos detallados. (María de los Ángeles Ibarra, 2005). OLAP puede trabajar con tres tipos de almacenamiento:

**MOLAP (OLAP Multidimensional):** en estos sistemas los datos se encuentran almacenados en una estructura de datos multidimensional. Algunas ventajas de este modelo es que puede escribir sobre la base de datos y posibilita hacer cálculos más complicados. Y tiene como desventaja: un tamaño limitado para la arquitectura del cubo y no puede acceder a los datos que no se encuentren en el mismo.

**ROLAP (OLAP Relacional):** son sistemas en los cubos, en los cuales los datos se encuentran almacenados en una base de datos relacional. Algunas ventajas de ROLAP son: los datos pueden ser compartida con aplicaciones SQL y cuenta con datos y estructura más dinámicos. Dentro de sus desventajas se encuentran: los cálculos están limitados a las funciones de las bases de datos y la construcción es cara.

HOLAP (OLAP Híbrido): estos sistemas mantienen los registros detallados en la base de datos relacional, mientras que los datos resumidos o agregados se almacenan en una base de datos multidimensional separada. Son conocidos como híbridos debido a que contiene las características de los sistemas anteriores.

### 1.1.3 Base de datos

Una base de datos (BD) se define como un fichero en el cual se almacena información en campos o delimitadores, teniendo acceso a ella posteriormente tanto de forma separada como de forma conjunta. Se utiliza normalmente para recoger grandes cantidades de información. (Paqui Valle Pérez, 2009) El contenido de una base de datos engloba a la información concerniente de una organización, de tal manera que los datos estén disponibles para los usuarios. Las bases de datos tienden a ser organizados por:

- Campo: es la unidad más pequeña a la cual uno puede referirse en un programa. Desde el punto de vista del programador representa una característica de un individuo u objeto.
- Registro: colección de campos de iguales o de diferentes tipos.
- Archivo: colección de registros almacenados siguiendo una estructura homogénea. (Guillermo Espinosa, 2008)

Las bases de datos tiene varias características, algunas de ellas son:

- Independencia de los datos: los datos no dependen del programa y por tanto cualquier aplicación puede hacer uso de los datos.
- Reducción de la redundancia: se llama redundancia a la existencia de duplicación de los datos, al reducir ésta al máximo se conseguirá un mayor aprovechamiento del espacio y además se evita que existan inconsistencias entre los datos. Las inconsistencias se dan cuando se encuentran con datos contradictorios.
- Seguridad: un sistema de base de datos debe permitir un control sobre la seguridad de los datos.
- Se visualiza normalmente como una tabla de una hoja de cálculo, en la que los registros son las filas y las columnas son los campos.
- Permite la programación a usuarios avanzados y realizar un listado de la base de datos. (Paqui Valle Pérez, 2009)

Se afirma que las bases de datos son uno de los recursos más importantes para el desarrollo de la empresa, se ha demostrado que son de gran influencia en el progreso de esta área, brindando un almacenamiento y acceso confiable, eficiente y práctico en el uso de la información que se produce. Las bases de datos no solo han servido como repositorios de información, es una de las herramientas más importantes en la gestión de información.

## **1.2 Sistemas similares**

Un sistema informático es el conjunto de partes interrelacionadas, hardware y software, que al unirse se utiliza para capturar, almacenar y procesar información. Los sistemas informáticos constituyen una parte esencial en las diferentes áreas de trabajo, por las disímiles oportunidades que esta ofrece; pues en la actualidad la mayoría de las empresas utilizan estos sistemas para el procesamiento de los datos.

En el mundo, existen muchos programas informáticos que almacenan las estadísticas ambientales con el objetivo de llevar un mejor control y organización de la información, entre ellos se encuentran los que se mencionan a continuación.

### **1.2.1 Instituto Nacional de Estadística en España**

El sitio se encarga de recopilar toda la información estadística de varios Indicadores como son: Medio Ambiente, Sociedad, Economía, Agricultura, entres otros. Los Indicadores ambientales ofrecen información de diferentes aspectos como son los gases de efecto invernadero, al agua como recurso natural renovable y la gestión de residuos. Los datos relacionados son mostrados en HTML y permite descargarlo en diferentes formatos entre ellos Excel.

Este sistema se asemeja al que se desea desarrollar, proporciona una organización de los datos y permite acceder a dicha información de una forma rápida y coordinada, pero no presenta los datos de algunos Indicadores ambientales, como son los suelos, el clima y los desechos sólidos, esto trae consigo que sea completamente diferente al sistema que se va a construir, por lo que esta no es una solución factible.

### **1.2.2 EUROSTAT**

El sistema EUROSTAT de Europa recoge información estadística de los países europeos. La base de datos ofrece varios Indicadores: Industria, Economía, Transporte, entre otros. El Indicador Medio Ambiente, ofrece información de varias temáticas: las emisiones de gases, energía y los desechos. Los datos referentes al Indicador, se puede presentar de distintas formas en PDF, tablas, gráficos, mapa, y se puede descargar en los formatos, XMS, HTML, XML, entre otros.

En este sistema se presenta características similares a las que necesita el software a desarrollar, se accede fácilmente pero solo brinda datos de muy pocos Indicadores ambientales, por lo que el sistema no logra cumplir con todos los requisitos que se necesita para elaborar el mercado de datos del Indicador Medio Ambiente para la Oficina Nacional de Estadística de Cuba.

### 1.2.3 CEPALSTAT

CEPALSTAT constituye el portal de acceso a la información estadística de los países de América Latina y el Caribe que recopila, sistematiza y difunde la CEPAL. Esta ha venido contribuyendo en la difusión de datos relevantes de los países de América Latina y el Caribe a través de diversos medios en las áreas económica, social y ambiental. (CEPALSTAT, 2009)

Los datos relacionados de los Indicadores ambientales son almacenados en una base de datos. La misma contiene información de las siguientes áreas temáticas que se han estructurado por tema: biodiversidad, agua, mares y borde costero, tierras y suelos, energía, aire, desastres naturales, desechos sólidos, asentamientos humanos, transporte y gestión ambiental.

La plataforma en que reside la base de datos de Medio Ambiente permite: consultas para cada Indicador a través de la selección que haga el usuario de períodos, países y desagregaciones; la grabación de los resultados de cada consulta en una planilla Excel; la generación de gráficos de líneas e histogramas a partir del resultado de la consulta en HTML y la visualización de las fichas técnicas, que contienen enlaces a bases de datos y páginas Web que proporcionan datos específicos sobre los Indicadores incluidos. (CEPALSTAT, 2009)

El sistema cuenta con buena organización de la información pero solo brinda información estadística de las áreas económica, social y ambiental, en cuanto al Indicador Medio Ambiente propone diferentes temáticas, por lo cual el usuario puede acceder a los datos de su interés de manera fácil y cómoda. Es una aplicación muy sencilla y rápida donde solo se controlan algunos datos de cada país Latinoamericano.

### 1.2.4 UNdata

La División de Estadísticas de las Naciones Unidas (UNSD) ha puesto al servicio de la comunidad internacional una amplia base de datos en temas como Población, Industria, Energía, Comercio y Medio Ambiente que abarcan a todos los países. La base de datos estadísticas de los Indicadores ambientales recoge información con respecto a temáticas: aire, desechos y los gases de efecto invernadero. El sistema muestra la información en HTML, brinda la posibilidad de descargar la información en el formato XML además de otros, es fácil y rápido el acceso a la información.

El sistema posee una buena estructura de los datos pero no contiene información estadística de algunos Indicadores ambientales, por lo que el sistema no cumple con todas las recomendaciones que se requiere en el sistema a realizar.

### **1.3 Metodologías de desarrollo**

El desarrollo de un software no es una tarea fácil, es por eso que surge la necesidad de usar una metodología. Las metodologías son un conjunto de procedimientos, técnicas y un soporte documental que imponen un proceso disciplinado sobre el desarrollo del software con el objetivo de hacerlo más predecible y eficiente. Para la elaboración del mercado de datos del Indicador Medio Ambiente se debe identificar una metodología que guie el proceso de desarrollo a este tipo de solución.

Para hacer uso de la misma en soluciones de almacenes de datos e inteligencia de negocios (DW&BI) solo se hará referencia a los criterios de Kimball e Inmon por marcar claramente su tendencia en cuanto al tema de almacenes de datos. Aunque existen otras que se basan en estas propuestas, pero no siguen obligatoriamente una específica sino que realizan una selección de lo mejor de cada una; algunas de ellas son: SQLBI, DM2, CRISP-DM, Hefesto y Metodología para el diseño conceptual de almacenes de datos.

#### **1.3.1 Inmon**

Conocida como la Metodología de Inmon, en honor a su creador William H. Inmon, uno de las personalidades más influyentes en el área de los almacenes de datos. Fue el creador del término DW y es considerado como el padre de la disciplina.

La visión de Inmon se basa en un enfoque descendente (top-down), propone construir primero el almacén de datos y a partir de este los mercados de datos, plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Al ser construido descendentemente los mercados de datos se nutren del DW Corporativo, convirtiéndose en un complejo empresarial de bases de datos relacionales.

Inmon afirma que la creación de una base de datos relacional con una leve normalización necesita ser las bases para los mercados de datos. Por lo que no se crean los mercados de datos directamente desde los sistemas de procesamiento de transacciones en línea a través de un área de ensayo. En lugar de ello, se crean a partir de la arquitectura relacional de los datos corporativos.

### 1.3.2 Kimball

Ralph Kimball es reconocido a nivel mundial en el diseño de DW y creador del enfoque multidimensional. Se ha dedicado desde hace unos cuantos años al desarrollo de su metodología para que éste concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de estos proyectos.

La propuesta de Kimball se basa en dividir el mundo de inteligencia de negocio entre los hechos y las dimensiones, esta es eficaz y conduce a una solución completa en un corto período de tiempo. Entre sus características principales es que su arquitectura es ascendente (bottom-up), plantea que se debe crear por cada departamento un conjunto de mercado de datos independientes orientados a los temas que estén relacionados con él. Y “el almacén de datos es la unión de todos los mercados de datos de una entidad”. (Curto, 2008)

#### Selección de la Metodología

De las metodologías observadas la vital discrepancia es el enfoque de estas personalidades. El Centro de Tecnologías de Almacenamiento de Datos (DATEC) seleccionó la metodología de Kimball porque propone construir un mercado de datos por cada temática y luego unirlos para conformar el almacén de datos. Basada en la misma, DATEC la adaptó a la necesidades del entorno de trabajo y trazó su propia metodología designándola: Metodología para el desarrollo de soluciones de almacenes de datos e inteligencia de negocio en la línea de almacenes de datos e inteligencia de negocio (DW&BI).

### 1.4 Sistema gestor de bases de datos

Un sistema gestor de bases de datos (SGBD) es una colección de datos relacionados entre sí, estructurados y organizados y un conjunto de programas que acceden y gestionan esos datos. La colección de esos datos se denomina base de datos. (Editorial McGraw-Hill, 2008)

Los sistemas gestores de bases datos están diseñados para gestionar grandes bloques de información, que implica tanto la definición de estructuras para el almacenamiento como de mecanismos para la gestión de la información. El SGBD es una aplicación que permite a los usuarios definir, crear y mantener la base de datos y proporciona un acceso controlado a la misma. Brinda los siguientes servicios:

- Creación y definición de la base de datos, especificación de la estructura, el tipo de los datos, las restricciones y relaciones entre ellos mediante lenguajes de definición de datos. Toda esta información se almacena en el diccionario de datos, el SGBD proporcionará mecanismos para la gestión del diccionario de datos.

- Manipulación de los datos realizando consultas, inserciones y actualizaciones de los mismos utilizando lenguajes de manipulación de datos.
- Acceso controlado a los datos de la base de datos mediante mecanismos de seguridad de acceso a los usuarios.
- Mantener la integridad y consistencia de los datos utilizando mecanismos para evitar que los datos sean perjudicados por cambios no autorizados.
- Acceso compartido a la base de datos, controlando la interacción entre usuarios concurrentes.
- Mecanismos de respaldo y recuperación para restablecer la información en caso de fallos en el sistema. (Editorial McGraw-Hill, 2008)

Existen muchos SGBD que son usados por profesionales de la información para diseñar trabajos colaborativos e interdisciplinarios. Están disponibles para el uso y desarrollo de aquellos interesados en el área de la gestión de la información. Para seleccionar el gestor de bases de datos adecuado para la construcción del mercado de datos del Indicador Medio Ambiente, se realizará un estudio de algunos SGBD más usados donde se hará una descripción de las ventajas y desventajas de cada uno de ellos.

### 1.4.1 MySQL

Es un sistema de gestión de bases de datos relacional, fue creada por la empresa sueca MySQL AB, la cual tiene el copyright del código fuente del servidor SQL, así como también de la marca. MySQL es un software de código abierto licenciado bajo la GPL de la GNU, aunque MySQL AB distribuye una versión comercial, en lo único que se diferencia de la versión libre, es en el soporte técnico que se ofrece y la posibilidad de integrar este gestor en un software propietario, ya que de otra manera se vulneraría la licencia GPL. El principal objetivo de MySQL es la velocidad y robustez. Entre sus principales características se destacan las siguientes: soporta gran cantidad de tipos de datos para las columnas, gran portabilidad entre sistemas, puede trabajar en distintas plataformas y sistemas operativos, aprovecha la potencia de sistemas multiproceso gracias a su implementación multihilo. (Dr. Marco Antonio Cruz-Chávez, 2009)

Este sistema tiene muchas ventajas como las que se muestran a continuación.

- Bajo costo en requerimientos para la elaboración de bases de datos, ya que debido a su bajo consumo puede ser ejecutado en una máquina con escasos recursos.
- Los sistemas se pueden conectar usando socket TCP/IP en cualquier plataforma.

- Facilidad de configuración e instalación. (Dr. Marco Antonio Cruz-Chávez, 2009)
- Ofrece mejor rendimiento al ser usado en aplicaciones Web, en Drupal y con su conexión programada en el lenguaje PHP.

Al igual que otros sistemas tiene desventajas:

- Puede provocar problemas de integridad y la información se puede dañar en entornos donde se gestione mucha información simultáneamente, altas concurrencias en las modificaciones de datos.
- No es viable para su uso con grandes bases de datos a las que se acceda continuamente ya que no implementa una buena escalabilidad.
- Carece de soporte para transacciones, rollback y subconsultas.

### **1.4.2 Oracle**

Oracle es básicamente una herramienta cliente/servidor para la gestión de bases de datos. Es un producto vendido a nivel mundial, aunque la gran potencia que tiene y su elevado precio hacen que sólo se vea en empresas muy grandes y multinacionales. En el desarrollo de páginas web ocurre lo mismo: como es un sistema muy caro no está tan extendido como otras bases de datos, por ejemplo, Access, MySQL, SQL Server, entre otras. Consume gran cantidad de recursos en el servidor, lo que trae consigo que se necesite de un buen sistema de hardware. Cuenta con herramientas necesarias para su utilización así como con herramientas de programación básicas de Oracle. (Marco A Cruz-Chávez, 2009)

Ventajas de Oracle:

- Puede ejecutarse en varias plataformas, desde una PC hasta un servidor que cuente con diferentes sistemas operativos.
- Permite crear los formularios de forma similar a otras herramientas de programación conocidas como Visual Basic, Visual C.
- Capaz de almacenar una amplia variedad de información y recuperarla de forma rápida y fácil.
- Es un sistema muy complejo que permite desarrollar diseños con triggers y procedimientos almacenados con una integridad referencial bastante potente.

Como otros sistemas de gestión de base de datos cuenta con ciertas desventajas:

- Es muy riesgoso el controlador de versiones ya que puede perderse o dañarse la información.

- Una de sus mayores desventajas radica en su precio, incluso las licencias son excesivamente caras.
- Otro problema es la necesidad de ajustar el sistema a las necesidades del cliente, no basta con instalar el Oracle en un servidor y conectar las aplicaciones al cliente, ya que si es mal configurado puede resultar extremadamente lento.

### 1.4.3 PostgreSQL

PostgreSQL es un Sistema de Gestión de Bases de Datos Objeto-Relacionales (ORDBMS) que ha sido desarrollado de varias formas desde 1977. Comenzó como un proyecto denominado Ingres en la Universidad Berkeley de California. En 1986 otro equipo dirigido por *Michael Stonebraker*<sup>1</sup> continuó el desarrollo del código de Ingres para crear un sistema de bases de datos objeto-relacionales llamado Postgres. En 1996, debido a un nuevo esfuerzo de código abierto y a la incrementada funcionalidad del software, Postgres fue renombrado a PostgreSQL. El proyecto PostgreSQL sigue actualmente un activo proceso de desarrollo a nivel mundial gracias a un equipo de desarrolladores y contribuidores de código abierto. Está ampliamente considerado como el sistema de bases de datos de código abierto más avanzado del mundo. (Carlos D. González , 2010)

El PostgreSQL corre en casi todos los sistemas operativos: Linux, Unix, BSDs, Mac OS, Beos y Windows. Cuenta con una documentación muy bien organizada, pública y libre, con comentarios de los propios usuarios y es altamente adaptable a las necesidades del cliente. Trabaja con varios lenguajes de programación como son: PHP, C, C++, Perl, Python, entre otros. Soporta todas las características de una base de datos profesional (triggers, funciones, secuencias, relaciones, reglas, tipos de datos definidos por usuarios, vistas materializadas). PostgreSQL funciona con grandes cantidades de datos y una alta concurrencia, con muchos usuarios accediendo a la vez al sistema. Sus características técnicas la hacen una de las bases de datos más potentes y robustos del mercado. Su desarrollo comenzó hace muchos años y durante ese tiempo, estabilidad, potencia, robustez, facilidad de administración e implementación de estándares han sido las características que más se han tenido en cuenta durante su desarrollo. (Ralfm , 2007)

---

<sup>1</sup> Es un científico especializado en la base de datos de investigación y desarrollo. Fue profesor de Ciencias de la Computación en la Universidad Berkeley de California y ayudó a crear la mayoría de la base de datos relacionales del mercado existente. Es el fundador de Ingres, ilustraciones, Cohera, entre otros. También es un editor para el libro de lecturas en sistemas de bases de datos.

Entre sus ventajas se puede encontrar:

- Es capaz de ajustarse al número de CPU y la cantidad de memoria que ofrece el sistema de forma óptima, siendo capaz de soportar mayor cantidad de peticiones simultáneas sin provocar errores.
- Comprueba la integridad referencial y así como también la de almacenar procedimientos en la propia base de datos.
- Permite la creación de funciones personalizadas, el manejo y configuración de disparadores.

Algunas desventajas son:

- La velocidad de respuestas que ofrece este gestor es un poco deficiente.
- Tienen un límite de 8k por filas que se pueden aumentar hasta 32 lo que trae consigo que disminuya su rendimiento.

PostgreSQL tiene varias versiones disponibles, la última (versión 8.4) fue lanzada el primero de julio del 2009. La mayoría de los cambios en PostgreSQL 8.4 son herramientas y órdenes de administración, nuevas o mejoradas. Entre las mejoras más populares están:

- Configuración de idioma y ordenamiento por base de datos. Para que se pueda seleccionar la configuración más adecuada dependiendo del idioma que se requiera.
- Nuevas herramientas para monitorear las consultas, entregando mayor información a los administradores para saber lo que está sucediendo en la base de datos.
- Implementación de funcionalidades avanzadas de ANSI SQL: 2003 que permiten realizar consultas complejas en una sola expresión en donde antes se requerían varias.
- Mejoras en procedimientos almacenados, por ejemplo usar valores por omisión en la declaración de parámetros o listas de argumentos de largo variable, al más puro estilo C++/Java. (Franco Catrin, 2009)

### **Selección del Sistema Gestor de Bases de Datos**

Después de realizar una comparación (ver Tabla 1) con las diferentes características de los sistemas de gestión de bases de datos, el Centro de Tecnologías de Almacenamiento de Datos seleccionó PostgreSQL por ajustarse a todos los requerimientos mínimos que se desea para el desarrollo del mercado de datos, además es el gestor solicitado por el cliente. Este manejador de base de datos es libre, a diferencia de Oracle y MySQL que son privativos y hay que pagar por una licencia. Oracle consume gran

cantidad de recursos por lo que se necesita de una PC con buenas características. En cuanto al MySQL cuando almacena o gestiona gran cantidad de datos pueden existir pérdidas o daños en la información, su integridad es baja a diferencia del PostgreSQL.

Crterios	MySQL	Oracle	PostgreSQL
<b>Sistemas operativos</b>	Linux/Windows	Windows, Unix, Macintosh y Mainframes	Linux/Windows
<b>Consumo de Recursos</b>	Consume pocos recursos	Consume muchos recursos	Consume pocos recursos
<b>Velocidad</b>	Rápido	Rápido	Lento
<b>Costo</b>	Costoso/Gratis	Muy costoso	Gratis
<b>Rendimiento</b>	Alto	Alto	Alto
<b>Comercialización</b>	Privativo/Libre	Privativo	Libre
<b>Volumen de datos</b>	Alto	Alto	Alto
<b>Integridad</b>	Baja	Alta	Alta

**Tabla 1:** Comparación entre los SGBD

El PostgreSQL es elegido porque aventaja los otros SGBD en varios aspectos aunque es más lento al ejecutar los procesos, pero se adapta al sistema que se necesita para el desarrollo de la aplicación. Específicamente se seleccionó la versión 8.4 porque tiene características más avanzadas, contando con 293 mejoras para que la administración, consulta y programación en PostgreSQL sea más fácil.

## 1.5 Herramienta manejadora de base de datos

Para administrar las bases de datos en PostgreSQL 8.4, el Centro de Tecnologías de Almacenamiento de Datos escogió la herramienta PgAdmin3 versión 1.10 porque es un software libre que accede a todos los objetos del PostgreSQL y responde a las necesidades de todos los usuarios.

### 1.5.1 PgAdmin3

PgAdmin3 es una aplicación gráfica para gestionar el gestor de bases de datos PostgreSQL, siendo la más completa y popular con licencia Open Source. Está escrita en C++ usando la librería gráfica multiplataforma wxWidgets, lo que permite que se pueda usar en Linux, FreeBSD, Solaris, Mac OS X y Windows. Es capaz de gestionar versiones a partir de PostgreSQL 7.3 ejecutándose en cualquier plataforma, así como versiones comerciales de PostgreSQL como Pervasive Postgres, EnterpriseDB, Mammoth Replicator y SRA PowerGres. Está diseñado para responder a las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y facilita enormemente la administración. (Ubuntu, 2008)

Algunas características de este manejador de datos son:

- Cuenta con una sintaxis SQL editor, servidor en el editor de código.
- La conexión es utilizando TCP/IP o Unix Domain Sockets y no requiere de drives para comunicarse con el servidor.
- Muy rápido para la visualización y entrada de datos.
- Procedimientos almacenados.
- Transacciones. (Ubuntu, 2008)

La versión 1.10 de PgAdmin3, tiene nuevas características muy interesantes como son: construcción gráfica de una determinada consulta; inclusión del framework pgScript para el desarrollo de scripts y así ejecutar las consultas; buscador de objetos; menú para habilitar/deshabilitar reglas y borra o reasigna un rol a una determinada base de datos. (GUI , 2009)

### 1.6 Herramienta CASE

En la actualidad, muchas empresas se han extendido a la adquisición de herramientas CASE (Ingeniería Asistida por Computadora), con el fin de automatizar los aspectos clave de todo el proceso de desarrollo de un sistema. CASE proporciona un conjunto de herramientas automatizadas que están desarrollando una cultura de ingeniería nueva para muchas empresas. Uno de los objetivos más importante del CASE es conseguir la generación automática de programas desde una especificación a nivel de diseño. De

acuerdo con *Kendall y Kendall*<sup>2</sup> la ingeniería de sistemas asistida por ordenador es la aplicación de tecnología informática a las actividades, las técnicas y las metodologías propias de desarrollo, su objetivo es acelerar el proceso para el que han sido diseñadas; automatiza o apoya a una o más fases del ciclo de vida del desarrollo de sistemas. (Ana Mercedes Caceres, 2006)

Cuando se hace la planificación de la base de datos, la primera etapa del ciclo de vida de las aplicaciones de bases de datos, se puede escoger una herramienta CASE que permita llevar a cabo el resto de tareas del modo más eficiente y efectivo posible. El uso de las herramientas CASE puede mejorar la productividad en el desarrollo de una aplicación de bases de datos. Una herramienta CASE suele incluir:

- Un diccionario de datos para almacenar información sobre los datos de la aplicación de bases de datos.
- Herramientas de diseño para dar apoyo al análisis de datos.
- Herramientas que permitan desarrollar el modelo de datos corporativo, así como los esquemas conceptual y lógico.
- Herramientas para desarrollar los prototipos de las aplicaciones. (Ana Mercedes Caceres, 2006)

La utilización de herramientas de modelado permite gestionar y mantener aplicaciones que trabajan con un volumen de datos. Entre estas herramientas se encuentra ER/Studio de Embarcadero y Visual Paradigm que permite transformar, migrar e integrar amplias cantidades de datos procedentes de diversas fuentes de la empresa.

### **1.6.1 Er/Studio**

Es una herramienta de modelado de datos fácil de usar y multinivel para el diseño de bases de datos a nivel físico y lógico. Mejora la calidad del modelo de datos y reduce los riesgos asociados a la seguridad de la información. Está equipado para crear y manejar diseños de bases de datos funcionales y confiables. Ofrece fuertes capacidades de diseño lógico, sincronización bidireccional de los diseños físicos y lógicos, construcción automática de bases de datos, documentación y fácil creación de reportes. (Equipo Danysoft, 2006)

Ofrece numerosas características, entre ellas se encuentran:

---

<sup>2</sup> "Análisis y diseño de sistemas" de Kenneth Kendall y Julie Kendall, es un libro que presenta de manera precisa los métodos, herramientas y técnicas de desarrollo de sistemas.

- Automáticamente crea diagramas altamente navegables y legibles.
- Reversa las más recientes secuencias de operación en orden cronológico.
- Permite muchos diseños físicos desde una arquitectura central lógica.
- Publica modelos y reportes en una variedad de formatos incluyendo HTML, RTF, XML Schema, DTD Output, y ER/Studio Viewer, un ambiente interactivo.
- Genera código fuente para algunos diseños de base de datos.
- Construye modelos gráficos desde una base de datos o esquema existente.
- Impulsa diseños de esquemas estrella y copo de nieve y soporta la importación de elementos de metadatos desde una variedad de plataformas de BI y almacenes de datos.
- Permite modelado de usuarios, roles y permisos en niveles lógico y físico.
- Soporta varios SGBD como son: MySQL, Oracle, PostgreSQL, Sybase, entre otros. (Embarcadero T, 2009 )

ER/Studio es una herramienta de base de datos que ayuda a diseñar, generar y mantener aplicaciones de base de datos de calidad y alto rendimiento. Desde un modelo lógico de los requerimientos de información y reglas del negocio que definen la base de datos, hasta un modelo físico optimizado por las características específicas de la base de datos de destino. ER/Studio permite visualizar la estructura adecuada, los elementos clave y un diseño optimizado de la base de datos. (Equipo Danysoft, 2006)

### **1.6.2 Visual Paradigm**

Es una herramienta CASE de modelado que utiliza UML (Lenguaje Unificado de Modelado). Soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Los sistemas de modelado UML ayudan a una rápida construcción de aplicaciones de calidad y a un menor coste. Permite dibujar todos los tipos de diagramas de clases, ingeniería inversa, generar código a partir de diagramas, generación de objetos a partir de bases de datos y generación de bases de datos a partir de diagramas de entidad-relación. (Descargas de software, 2007) Posee una plataforma capaz de integrarse con Eclipse, NetBeans, Oracle JDeveloper, JBuilder, IntelliJ IDEA, WebLogic Workshop, Microsoft Visual Studio, entre otras.

La herramienta es colaborativa, es decir, soporta múltiples usuarios trabajando sobre el mismo proyecto; genera la documentación automáticamente en varios formatos como web o PDF y permite el control de

versiones. Cabe destacar su robustez, usabilidad y portabilidad. (Blog de Eduardo León, 2007) El Visual Paradigm es un producto de calidad, soporta aplicaciones web, genera código para Java y es fácil de instalar y actualizar. (Luis Giraldo y Yuliana Zapata, 2005)

### **Selección de la herramienta para el modelado de los datos**

La herramienta Er/Studio es una excelente herramienta para el diseño de almacenes de datos y mercados de datos pero a desventaja de Visual Paradigm, Cuba no pagó su licencia, lo cual implica que no se puede utilizar para el desarrollo de aplicaciones. Con este aspecto mencionado, se decidió trabajar con la herramienta CASE de modelado el Visual Paradigm versión 6.4. La misma realiza de forma íntegra los planes de construcción del software. Genera de forma automática códigos desde diagramas y ofrece la posibilidad de generar documentación.

## **1.7 Modelo de datos**

Un modelo de datos es un conjunto de herramientas conceptuales para describir datos, sus relaciones, su significado y sus restricciones de consistencia. Es básicamente una descripción de un contenedor de datos así como de los métodos para almacenar y recuperar información de estos contenedores. Los modelos de datos no son cosas físicas: son abstracciones que permiten la implementación de un sistema eficiente de base de datos. (Carlos Minay, 2009) ¿De qué modo deben diseñarse las bases de datos que conforman un almacén de datos para soportar eficientemente los requerimientos de los usuarios?

Para dar respuesta a la siguiente interrogante, se realizará un estudio de los modelos de datos con el objetivo de seleccionar la estructura más óptima para desarrollar el mercado de datos.

### **1.7.1 Modelo entidad-relación**

El modelo de entidad-relación también denominado diagrama E-R es un modelo de datos basado en una percepción del mundo real que consiste en un conjunto de objetos básicos llamados entidades y relaciones entre estos objetos, implementándose en forma gráfica a través del diagrama entidad-relación. (Guillermo Storti y otros, 2007) El modelo entidad-relación fue propuesto por *Peter Chen*<sup>3</sup> y sus principales elementos son las entidades con sus atributos y las relaciones entre entidades. (Sara Álvarez, 2007)

---

<sup>3</sup> Peter Pin-Shan Chen en 1976, crea el Modelo Entidad-Relación y desde entonces se viene utilizando de una forma muy global. Ha sido la base para diversas metodologías sobre análisis y diseño de sistemas, herramientas de ingeniería de software asistida por computador y repositorios de sistemas.

- **Entidad:** se trata de un objeto del que se recoge información de interés de cara a la base de datos. Gráficamente se representan mediante un rectángulo. Dentro de las entidades pueden ser fuertes o débiles. Las fuertes son las que no dependen de otras entidades para existir, mientras que las entidades débiles siempre dependen de otra entidad sino no tienen sentido por ellas mismas.
- **Relación:** se puede definir la relación como una asociación de dos o más entidades. A cada relación se le asigna un nombre para poder distinguirla de las demás y saber su función dentro del modelo entidad-relación.
- **Atributo:** se define como cada una de las propiedades de una entidad o relación. Cada atributo tiene un nombre y todos los posibles valores que puede tener. Dentro de una entidad tiene que haber un atributo principal que identifica a la entidad y su valor tiene que ser único. (Sara Álvarez, 2007)

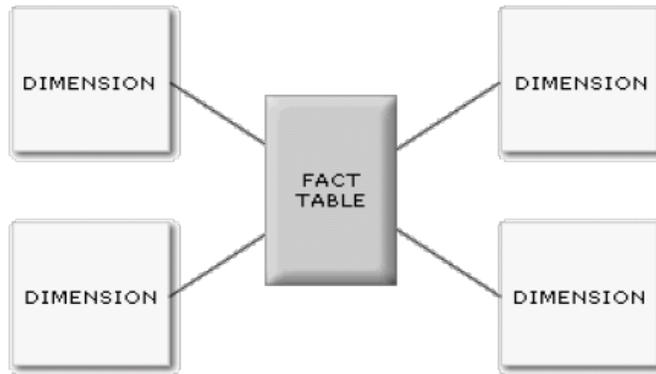
Este modelo se desarrolló para facilitar el diseño de bases de datos permitiendo la especificación de un esquema de la empresa que representa la estructura lógica completa de una base de datos. Debido a esta utilidad, muchas herramientas de diseño de bases de datos se basan en los conceptos del modelo E-R pero no son recomendables para la delineación de los almacenes de datos debido a que no garantizan la recuperación óptima del gran cúmulo de información que se almacena. Además estos diagramas tienden a resultar en un diseño normalizado mientras que en un almacén de datos este aspecto no es un requisito a tener muy en cuenta. (Hobbs, 2005)

### 1.7.2 Modelo dimensional

Se recurre al modelo dimensional cuando se diseñan mercados de datos, almacenes de datos y sistemas de inteligencia de negocios. Teniendo un fin analítico es necesario utilizar esta técnica de modelado que permiten registrar las relaciones existentes entre los datos para ofrecer una fuente de consulta que facilite el análisis y toma de decisiones. (Hazbleydi C. Verástegui, 2007) El modelo dimensional divide el mundo de los datos en dimensiones, medidas y hechos.

- **Hechos:** son una colección de datos. Cada uno representa una transacción o un evento y se registran en las tablas centrales.
- **Dimensiones:** son una colección de miembros del mismo tipo, cada punto de entrada de la tabla de hechos está conectado a una dimensión además determinan el contexto de los hechos.
- **Medidas:** es un atributo numérico de un hecho que representa el comportamiento del negocio relativo a la dimensión.

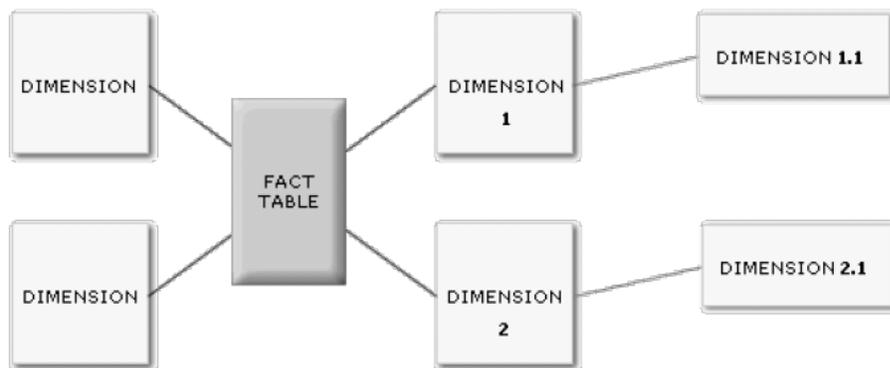
Todos estos objetos mencionados siguen un patrón llamado el modelo estrella (Figura 1), en que una tabla de hechos es el centro de todo el modelo y está unida a múltiples dimensiones que contextualizan las medidas registradas.



**Figura 1: Esquema de estrella**

Las dimensiones se encuentran en segunda forma normal y se organizan en una o más jerarquías para categorizar los datos. Una jerarquía está compuesta de niveles y cada nivel contiene un conjunto de atributos relacionados que describen a las medidas ubicadas en los hechos.

Existen otras versiones del modelo estrella: constelación y copo de nieve o snowflake. En este último esquema las dimensiones se encuentran en tercera forma normal y pueden crearse jerarquías con niveles pertenecientes a más de una dimensión. Se dice que una dimensión está en copo de nieve (Figura 2) cuando los atributos de baja cardinalidad se llevan a tablas separadas. (Hazbleydi C. Verástegui, 2007)



**Figura 2: Esquema de copo de nieve**

Este tipo de esquema se usa para ahorrar espacio de almacenamiento aunque no es recomendable su utilización en los almacenes debido a que hace las prestaciones complejas y más lentas las aplicaciones de consulta.

### **Selección del modelo de datos**

Dadas las características de los almacenes de datos es ideal la utilización del modelo dimensional. Este tipo de diseño tiene como ventajas sobre el modelo entidad-relación que es muy flexible y orientado a los intereses de un usuario final. Mediante la utilización de un modelo dimensional se disminuye la cantidad de tablas y relaciones entre ellas, lo que agiliza el acceso a los datos. Se utilizará el esquema de estrella porque es el más simple, tiene mejor rendimiento y velocidad que el de constelación y copo de nieve.

### **Conclusiones**

En el presente capítulo se abordaron conceptos elementales relacionados con el tema, en el mismo se observó que la tecnología de almacenamiento adecuada para el desarrollo de la herramienta es el mercado de datos que sirve para acumular grandes volúmenes de datos de oficinas a nivel departamental, además se definió como modo de almacenar la información: ROLAP porque se integra al gestor de base de datos seleccionado. Se realizó una investigación sobre el estado del arte de las distintas bases de datos existentes y se pudo observar que ninguna de las soluciones se ajusta a las necesidades porque la mayoría están suscritos a restricciones legales y brindan información solo de algunos Indicadores ambientales. Para guiar el proceso de construcción del mercado de datos se seleccionó la Metodología de Kimball, adaptándola a las necesidades propias de DATEC. La misma fue escogida porque tiene una amplia documentación y posee bien definidas los roles, las actividades, las etapas y los artefactos. Después de un análisis de varias herramientas y apoyados en las características de cada una, se seleccionó PostgreSQL 8.4 por ser libre y brindar una serie de funcionalidades, como herramienta de administración PgAdmin3 1.10 y para el diseño de la base de datos Visual Paradigm 6.4, además se utilizará el modelo dimensional por ser el más óptimo para el desarrollo del mercado de datos. Una vez definidos estos aspectos es permisible pasar al posterior apartado donde se presenta el análisis y diseño de los indicadores del Medio Ambiente.

## **CAPÍTULO 2: ANÁLISIS Y DISEÑO**

### **Introducción**

A pesar de las grandes ventajas del almacén de datos, parecen existir unas importantes barreras para su utilización en empresas de tamaño mediano. Los productos Data Warehouse han nacido para resolver problemas de análisis de grandes masas de información, contando con hardware muy potente y la masiva intervención de consultores externos, expertos en la realización de la puesta en marcha. Un proyecto de este tipo resulta en todos los aspectos excesivo para un departamento de ventas que necesita analizar la información de 500.000-3.000.000 de líneas de pedidos o una cantidad equivalente de información financiera, que es lo normal para una empresa mediana.

Para resolver este tipo de necesidades han surgido los mercados de datos, productos que utilizan la tecnología Data Warehouse adaptada a las necesidades de las empresas medias. El mercado de datos resuelve aplicaciones a nivel departamental, se destaca por una definición de requerimientos más fácil y rápida. También se simplifica el desarrollo de todo el mecanismo de su base de datos y con ello baja substancialmente todo el coste del proyecto, así como su duración. Por lo tanto se realizará un mercado de datos de cada Indicador plasmado en la ONE; el Medio Ambiente es uno de ellos y para lograr una adecuada recopilación de los datos es necesario realizar el análisis y diseño de los Indicadores ambientales, el cual corresponde al objetivo fundamental de la presente sección.

### **2.1 Definición del Negocio**

La Oficina Nacional de Estadísticas es una entidad creada para proponer, organizar y ejecutar, según corresponda, la aplicación de la política estatal en materia de estadística del país. La misión de la ONE es garantizar la producción de estadísticas de calidad a través del Sistema Estadístico Nacional (SEN) ejerciendo una adecuada dirección, ejecución y control de la captación de las cifras económicas y sociales, así como su adecuada difusión de acuerdo con las necesidades de la economía del país en información estadística. En el SEN participan coordinadamente los órganos, organismos, instituciones y entidades que elaboran estadísticas en el país. Está integrado por tres subsistemas:

1. Sistema de Información Estadística Nacional (SIEN): dirigido a la elaboración de estadísticas destinadas a satisfacer los requerimientos informativos de los más altos niveles del Estado y el Gobierno e informar a los organismos internacionales.

2. Sistema de Información Estadística Territorial (SIET): dirigido a la elaboración de estadísticas destinadas a satisfacer los requerimientos informativos de los organismos a los efectos del control administrativo de sus entidades.
3. Sistema de Información Estadística Complementaria (SIEC): dirigido a la elaboración de estadísticas destinadas a satisfacer los requerimientos informativos del Gobierno en el territorio y otros actores locales.

La ONE tiene una estructura institucional distribuida territorialmente en las provincias y municipios del país. Existen 16 oficinas provinciales y 169 oficinas municipales, subordinadas a las ONE provinciales, las cuales son las encargadas de interactuar directamente con los centros informantes. El mecanismo de captación de la información está compuesto por varios procesos orientados a los tipos de modelos existentes. Los modelos de Estadísticas Continuas y Encuestas Periódicas son unos de estos procesos definidos. Las diferencias entre ambos se basan en los períodos de captura que puede ser mensual, trimestral, semestral, anual, entre otros. Cada uno de los mencionados se ajusta a los Indicadores relacionados con cada Indicador (educación, transporte, economía, entre otros) en dependencia a la labor que cada cual realice.

Los datos estadísticos de los Indicadores del Medio Ambiente se capturan mediante las Encuestas Periódicas del modelo 1003 de cada centro informante en los períodos mensual y anual. Estas oficinas entregan en una fecha establecida la información solicitada por la ONE; algunas de las fuentes son: Instituto Nacional de Recursos Hidráulicos; Oficina Técnica del Ozono, CITMA; Instituto de Meteorología; Dirección de Planeamiento y Servicios Comunes, MEP; Centro Nacional de Biodiversidad; Instituto de Suelos; Centro Nacional de Áreas Protegidas; Centro de Gestión, Información y Educación Ambiental; Cuerpo Nacional de Guardabosques del Ministerio del Interior y el Centro Nacional de Investigaciones Sismológicas.

Después de describir el proceso del negocio se destaca que la fundamental tarea de la ONE es absorber los datos estadísticos de todos los sectores para luego almacenarla y permitir la disposición de la misma, facilitando consultarla de una manera alígera y eficaz.

## 2.2 Temas de análisis

Los temas de análisis es un aspecto que hay que tener en cuenta para la elaboración de los mercados de datos. En la definición del negocio se mencionan varios Indicadores, pero solo se centralizará en el Medio Ambiente con el fin de plasmar la solución propuesta.

Cuba se destaca por una relativa riqueza de recursos naturales y una gran diversidad de paisajes y especies de la flora y la fauna. La situación ambiental sufrió más de cuatro siglos de explotación de los recursos naturales, lo que trajo consigo el deterioro de valiosos ecosistemas y la acumulación de daños ambientales, algunos irreversibles o de difícil recuperación.

Asociado al proyecto de elevar el nivel y la calidad de vida del pueblo, surgió el interés creciente por la protección del Medio Ambiente y su incorporación a los planes de desarrollo desde los albores de la Revolución cubana.

Muchos de los Indicadores han sido definidos en función de proveer una información útil, precisa y estructurada de acuerdo a criterios técnicos de los usuarios y aportan información a numerosos proyectos iniciados por el país para los Programas de Desarrollo, sobre todo a partir del establecimiento en 1997 de la Estrategia Ambiental Nacional como guía para el desarrollo sostenible y la política ambiental del país. (ONE, 1958-2008)

Con la aplicación de los diferentes programas y planes se ha trabajado para:

- Incrementar la superficie forestal.
- Reducir la carga contaminante de las aguas terrestres y marinas.
- Intensificar la recuperación de las tierras afectadas.
- Perfeccionar y avanzar en la gestión ambiental en territorios y ecosistemas priorizados (áreas protegidas, cuencas hidrográficas, montañas y ambientes costeros y estuarios).
- Implementar y consolidar un Sistema Nacional de Áreas Protegidas.
- Disminuir la intensidad energética en el desempeño económico, que influye positivamente en la situación ambiental.
- Disminuir los consumos de Sustancias Agotadoras de la Capa de Ozono (SAO).
- Reducir las emisiones de dióxido de carbono provocadas por las actividades humanas.

En estos cruciales temas, Cuba recoge cifras que dan cuenta de todos sus esfuerzos, fruto del trabajo mancomunado de toda la sociedad cubana, y todos los proyectos y programas en desarrollo, contrastan hoy día con los peligros reales que acechan a la humanidad en medio del desastre ecológico provocado por la irracionalidad y la irresponsabilidad de naciones poderosas. (ONE, 1958-2008)

## **2.3 Roles y permisos**

El rol puede ser entendido como el papel que pone en práctica la persona en el drama social o en un sentido más preciso, como el sistema de expectativas sociales que acompañan a la presentación pública de los sujetos de un determinado estado social o estatus. El lenguaje de la teoría del rol suele ser complejo. El rol que una persona juega dentro de la sociedad va de acuerdo al estatus, que son expectativas que se exigen dentro del grupo social, no todos los roles son de la misma manera, sino que involucran el carácter de cada individuo y la forma en que se asimilan, por ejemplo: un estudiante debe cumplir con sus obligaciones de estudio, presentar exámenes, entre otras tareas. Si es un rol de un sistema informático, sería el papel que juega cada uno de ellos al interactuar con el sistema en función de los permisos que tiene al utilizar el producto. La ONE ha permitido a un grupo de personas, trabajar con la información proveniente del Indicador Medio Ambiente, para ello se definieron dos roles y los permisos de acceso al sistema con el objetivo de llevar un control de los datos. A continuación se nombrarán los mismos:

1. Analista: es la persona que busca información correspondida con los Indicadores ambientales por lo que realiza los reportes y el análisis de los datos relacionados con el tema de análisis. Tiene permiso de consulta.
2. Especialista de ETL: es la persona que realiza el proceso de extraer, transformar y cargar los Indicadores ambientales. Tiene permiso de consulta, escritura y actualización.

## **2.4 Reglas del Negocio**

Las reglas del negocio describen las políticas, normas y restricciones presentes en un software, los cuales son de vital importancia para alcanzar los objetivos propuestos. Para un buen funcionamiento del producto propuesto es necesario seguir las siguientes reglas desglosadas.

1. En la serie 28, el valor de estaciones de cloración se determinará por la suma de cloro de gas e hipoclorito.
2. En la serie 51, el valor de los residuos colectados será la suma de recogidos y dispuestos adecuadamente.
3. Se encontrará el pH en la serie 11, el cual es una medida de acidez o de alcalinidad de una sustancia líquida o sólida. El valor de 0-7 describe acidez y de 7-14 indica alcalinidad, mientras que pH=7 indica neutralidad.

4. Se pueden encontrar fechas que no están completas por lo que se cubrirá con el año solamente o acompañada con el día.
5. En las series se pueden hallar horas que no están registradas por lo que se representará con tres puntos seguidos.
6. En caso de que alguna serie se representa de la siguiente forma:
  - . No se efectuó la operación indicada por falta de algún dato.
  - .. Indicador no aplicable.
  - ... Cifras no disponibles al terminarse la redacción.
  - Resultado igual a cero.
  - 0** La cifra es más pequeña que la unidad de medida utilizada.

## **2.5 Necesidades de los usuarios**

La obtención de los valores estadísticos de los Indicadores ambientales para conocer el comportamiento de cada uno de ellos es una necesidad primordial del usuario. Por lo tanto, se realizó un análisis de los Indicadores y se desprendieron las siguientes penurias:

1. Obtener las temperaturas máximas y mínimas, vientos predominantes, cantidad de días con lluvia, nubosidad y humedad por provincias y estaciones de monitoreo.
2. Obtener las temperaturas máximas y mínimas absolutas registradas, por provincias, estaciones meteorológicas y período. Además de la fecha de cuando se registró.
3. Obtener la cantidad de lluvia media anualmente por provincias.
4. Obtener la cantidad de huracanes de diferentes intensidades que han azotado a Cuba, por períodos, regiones y categorías de huracanes en los meses de temporada ciclónica.
5. Obtener el número de veces que cada región ha sido azotada por huracanes de diferentes intensidades por categorías de huracanes, en un período de 1800-2008.
6. Obtener la cantidad de frentes fríos de diferentes intensidades que han azotado a Cuba en la temporada de invierno, por períodos, regiones e intensidades.
7. Obtener el promedio anual de la concentración de dióxido de azufre por estaciones de monitoreo.

8. Obtener el promedio anual de la concentración de dióxido de nitrógeno por estaciones de monitoreo.
9. Obtener el pH de la lluvia anual por estaciones de monitoreo
10. Obtener las emisiones brutas de efecto invernadero por los años comprendidos desde 1990 hasta 2002.
11. Obtener el consumo y variación de sustancias agotadoras de la capa de ozono por actividades consumidoras y sustancias controladas anualmente.
12. Obtener el consumo de sustancias agotadoras de la capa de ozono de acuerdo al potencial de agotamiento, por sustancias controladas anualmente.
13. Obtener la cantidad de dispositivos generadores de energía y biomasa empleada como combustible por dispositivos, biomasa y anualmente.
14. Obtener la cantidad de oferta renovable por combustible y biomasa anualmente.
15. Obtener la cantidad de agua dulce renovable y precipitaciones por años.
16. Obtener la extracción de agua por destinos anualmente.
17. Obtener la variación de extracción de agua por destinos en los años 2006, 2007 y 2008.
18. Participación de los destinos en la extracción bruta de agua dulce, en los años 2005, 2006, 2007 y 2008.
19. Obtener la carga dispuesta y la carga reducida por cuencas hidrográficas.
20. Obtener la carga dispuesta y la variación de carga contaminante por provincia.
21. Obtener la población con cobertura de agua potable y saneamiento, con acceso y sin acceso en el área urbana y rural.
22. Obtener la cobertura de agua potable por provincias anualmente.
23. Obtener la cobertura de saneamiento por provincias anualmente.
24. Obtener los sistemas de acueductos y alcantarillado anualmente.
25. Obtener la cantidad de especies conocidas, introducidas, endémicas, terrestres, fluviales, marinos, dulceacuícola por grupos taxonómicos.

26. Obtener la cantidad de endemismo vegetal por distritos geográficos y regiones.
27. Obtener la cantidad de flora vascular por categoría de flora.
28. Obtener la cantidad de especies de fauna extintas, en peligro crítico, en peligro y vulnerables, por categorías de fauna.
29. Obtener la cantidad de especies de floras amenazadas y endémicas, por instituciones y provincias.
30. Obtener número de áreas protegidas de acuerdo a su significación de áreas protegidas, superficie y por provincias.
31. Obtener áreas protegidas de reconocimiento internacional, de ello: superficie, zona núcleo, zona de amortiguamiento y zona de transición por áreas y provincias.
32. Obtener áreas protegidas por categoría de manejo por provincias.
33. Obtener la cantidad de pérdidas económicas por años y huracanes.
34. Obtener la cantidad de viviendas dañadas por años y huracanes.
35. Obtener la cantidad de incendios forestales por causas y provincias.
36. Obtener la superficie dañada por incendios forestales y su dinámica por provincias.
37. Obtener terremotos fuertes y perceptibles reportados en Cuba por localidades geográficas, de ello: fecha, magnitud, profundidad e intensidad.
38. Obtener el volumen de desechos sólidos por provincias.
39. Obtener el tratamiento y recolección de desechos sólidos.
40. Obtener la producción de materias primas y recicladas por productos y anualmente.
41. Obtener los gastos de inversión de protección del medio ambiente por sectores ambientales y provincias.
42. Obtener las inversiones para el medio ambiente por actividad económica, actividad ambiental y por sectores.
43. Obtener la cantidad de suelos, área estudiada por clasificación de los suelos y provincias.
44. Obtener la productividad de los suelos por clasificación agroproductiva y provincias.
45. Obtener el área agrícola afectada por provincias.

### **2.5.1 Requisitos de información**

Se estructuraron los Indicadores siguiendo un patrón simplificado de ordenamiento por tema de los cuales se adquirirán los posibles valores que demostrarán el proceder del Indicador Medio Ambiente.

1. Obtener información del clima.
2. Obtener información de los huracanes que han azotado a Cuba.
3. Obtener información de los frentes fríos.
4. Obtener información de los gases de efecto invernadero.
5. Obtener información de las sustancias agotadoras de la capa de ozono.
6. Obtener información de la energía renovable.
7. Obtener información de los desechos sólidos.
8. Obtener información de los gastos para la protección del Medio Ambiente.
9. Obtener información de las inversiones para la protección del Medio Ambiente.
10. Obtener información de los terremotos ocurridos en Cuba.
11. Obtener información de los suelos.
12. Obtener información de la carga contaminante.
13. Obtener información de la fauna cubana.
14. Obtener información de la flora cubana.
15. Obtener información de las lluvias.
16. Obtener información del agua y saneamiento.

### **2.5.2 Requisitos multidimensionales**

Se realizará un análisis de cada uno de los Indicadores ambientales para extraer todos los valores de entrada (VE) y salida (VS). Los VE son las informaciones como provincia, mes, años, entre otras, que están estrechamente relacionadas con los Indicadores ambientales porque dan a conocer su comportamiento y los VS no son más que aquellos datos estadísticos que serán visibles para el usuario como pueden ser la cantidad de huracanes, frentes de frío, lluvias, entre otras.

Principales Indicadores del clima.

VE: provincias con sus estaciones meteorológicas; VS: temperatura máxima y mínima, rapidez del viento, cantidad de días con lluvias; por ciento de humedad relativa y nubosidad media.

Temperaturas máximas y mínimas absolutas registradas.

VE: provincias con sus estaciones meteorológicas y período; VS: temperatura máxima y mínima.

Lluvia total media.

VE: provincias, años y meses; VS: cantidad de lluvia media.

Huracanes de diferentes intensidades que han azotado a Cuba.

VE: temporada ciclónica, períodos y categorías de huracanes; VS: cantidad de huracanes.

Número de veces que cada región ha sido azotada por huracanes de diferentes intensidades.

VE: regiones, períodos y categorías de huracanes; VS: cantidad de huracanes.

Proporción de huracanes por intensidades que han azotado a cada región.

VE: períodos, categoría de huracanes y región; VS: por ciento de huracanes.

Frentes fríos de diferentes intensidades que han azotado a Cuba.

VE: períodos e intensidades de frente frío y meses de frente frío; VS: cantidad de frentes fríos.

Número de veces que cada región ha sido azotada por frentes fríos de diferentes intensidades.

VE: intensidades de frente frío, períodos y regiones; VS: cantidad de frentes fríos.

Valor promedio anual de la concentración de dióxido de azufre (SO<sub>2</sub>) y dióxido de nitrógeno (NO<sub>2</sub>) por estaciones de monitoreo.

VE: estaciones meteorológicas y años; VS: valor promedio de SO<sub>2</sub> y NO<sub>2</sub>.

pH de la lluvia anual por estaciones de monitoreo.

VE: estaciones y años; VS: pH de la lluvia.

Emisiones brutas de gases de efecto invernadero.

VE: años-sector y gases de efecto invernadero; VS: cantidad de emisiones.

Consumo de sustancias agotadoras de la capa de ozono.

VE: sustancias-actividades consumidoras y años; VS: consumo de sustancias.

Consumo de sustancias agotadoras de la capa de ozono de acuerdo al potencial de agotamiento.

VE: sustancias agotadoras y años; VS: consumo de sustancias.

Consumo de sustancias agotadoras de la capa de ozono por actividad consumidora.

VE: actividades consumidoras y años; VS: consumo de sustancias.

Recursos de agua dulce renovables.

VE: fuentes de agua dulce y años; VS: cantidad de agua.

Extracción de agua por destino.

VE: extracción de agua y años; VS: cantidad de agua.

Variación de extracción de agua por destino.

VE: extracción de agua y años; VS: por ciento de agua.

Participación de los destinos en la extracción bruta de agua.

VE: extracción bruta de agua por destino y años; VS: por ciento de los participantes.

Carga contaminante por provincias.

VE: provincia y años; VS: carga dispuesta y variación de la carga.

Carga contaminante por cuencas hidrográficas de interés nacional.

VE: cuencas y años; VS: carga dispuesta y variación de la carga.

Población con cobertura de agua potable y saneamiento.

VE: cobertura de agua potable y saneamiento, años y zona de residencia; VS: cantidad de población.

Cobertura de agua potable y saneamiento por provincias.

VE: provincias y años; VS: por ciento de agua.

Sistema de acueductos y alcantarillados.

VE: años e Indicadores seleccionados de acueductos y alcantarillados; VS: cantidad de agua.

Afectaciones por ciclones tropicales.

VE: años-huracanes y pérdidas económicas; VS: cantidad y por ciento de pesos.

Viviendas dañadas por ciclones tropicales.

VE: años-huracanes y viviendas; VS: cantidad y por ciento de viviendas.

Incendios forestales por causas.

VE: provincias-superficie dañada y causas; VS: cantidad y por ciento de incendios forestales.

Número de incendios forestales y su dinámica por provincias.

VE: provincias y años; VS: cantidad y por ciento de incendios forestales.

Superficie dañada por incendios forestales y su dinámica por provincias.

VE: provincias y años; VS: cantidad y por ciento de superficie.

Terremotos fuertes y perceptibles reportados en Cuba.

VE: localidades geográficas; VS: cantidad de terremotos.

Diversidad de la biota cubana.

VE: grupos taxonómicos con sus nombres comunes de especies de la biota cubana y hábitat donde se desarrollan;

VS: cantidad de biota.

Diversidad y endemismo de la biota terrestre cubana.

VE: grupos taxonómicos con sus nombres comunes de especies de la biota cubana; VS: cantidad de especies endémicas y conocidas y por ciento de especies endémicas.

Endemismo vegetal por distritos geográficos seleccionados.

VE: provincias con sus distritos y endemismo vegetal; VS: cantidad y por ciento de endemismo.

Flora vascular amenazada.

VE: categorías; VS: cantidad y por ciento de flora.

Fauna Amenazada.

VE: categorías; VS: cantidad y por ciento de especies de fauna.

Taxones de la flora y fauna con protección ex situ.

VE: instituciones con sus provincias; VS: cantidad y por ciento de especies.

Áreas protegidas.

VE: provincias, significación de áreas protegidas y superficie; VS: cantidad de áreas protegidas.

Áreas protegidas con reconocimiento internacional.

VE: áreas por provincias; VS: por ciento de zona de núcleo, amortiguamiento y transición.

Áreas protegidas por categorías de manejo.

VE: provincias y categorías de manejo; VS: cantidad de áreas.

Volumen de desechos sólidos recolectados por provincias.

VE: provincias y años; VS: cantidad de desechos.

Tratamiento y recolección de desechos sólidos.

VE: desechos sólidos y años; VS: cantidad de desechos sólidos.

Producción de materias primas recicladas.

VE: productos y años; VS: cantidad y por ciento de materias primas.

Dispositivos generadores de energía renovable y biomasa empleada como combustible.

VE: dispositivos/biomasa y años; VS: cantidad de energía.

Oferta de energía renovable.

VE: fuentes renovables y años; VS: cantidad de energía.

Gastos de inversión para la protección del Medio Ambiente.

VE: sectores ambientales y años; VS: cantidad y por ciento de pesos.

Inversiones para el Medio Ambiente por actividad económica.

VE: actividad económica y años; VS: cantidad de pesos.

Gasto de inversión para la protección del Medio Ambiente por provincia.

VE: provincias y años; VS: cantidad de pesos.

Inversiones para el Medio Ambiente por sector ambiental.

VE: sectores ambientales y años; VS: cantidad de pesos.

Clasificación genética de los suelos de Cuba.

VE: clasificación genética de los suelos y provincias; VS: cantidad de hectáreas.

Clasificación agroproductiva de los suelos de Cuba.

VE: clasificación agroproductiva de los suelos y provincias; VS: cantidad de hectáreas.

Principales factores limitantes edáficos.

VE: factores; VS: cantidad y por ciento de área estudiada.

## **2.6 Requerimientos del sistema**

Un requerimiento es una condición o capacidad que necesita un usuario para resolver un problema o lograr un objetivo. Los requerimientos determinan lo que hará el sistema para satisfacer un contrato, estándar, u otro documento impuesto formalmente. (JACOBSON, y otros, 2000) Para que un sistema funcione correctamente se debe lograr una comunicación efectiva entre los usuarios y el equipo de proyecto con el objetivo de llegar a un entendimiento de lo que hay que hacer, lo que constituye la clave del éxito en la producción de un software.

### **2.6.1 Requisitos funcionales**

Los requisitos funcionales son capacidades o condiciones que el sistema debe cumplir, lo que es muy importante para satisfacer las expectativas del cliente, el producto debe ser capaz de:

RF1: Extraer los datos de los Indicadores ambientales.

RF2: Transformar los datos de los Indicadores ambientales.

RF3: Cargar los datos de los Indicadores ambientales.

### **2.6.2 Requisitos no funcionales**

Los requisitos no funcionales son propiedades o cualidades que el producto debe tener, y que lo hacen atractivo, usable, rápido o confiable. A continuación se listan los que debe cumplir el sistema en proposición.

RNF1: Software

- Sistema Operativo Windows 98 o superior y Linux.

RNF2: Hardware

- PC Cliente: Procesador Pentium III y 256 Mb de RAM.

- Servidor de BD: Procesador Dual Core o superior, 1Gb de RAM para garantizar el correcto funcionamiento del sistema al ser accedido por varios usuarios y 60Gb de capacidad en disco duro para el almacenamiento de la información.

### RNF3: Usabilidad

- El sistema podrá ser usado por la persona que gestione la información referente a los Indicadores estadísticos por lo que es necesario que los usuarios del producto deben tener un conocimiento básico del manejo de las computadoras y los sistemas operativos.
- La aplicación debe ser asequible, amena y funcional tanto para usuarios expertos como para novatos.

### RNF4: Fiabilidad

- El sistema no podrá tener fallos y en caso de que exista, tiene que ser capaz de restaurarse rápidamente.

### RNF5: Rendimiento

- El tiempo de la ejecución de las consultas debe estar entre un rango de cincuenta segundos a un minuto.
- Para cargar los datos el tiempo debe ser corto.
- La base de datos deberá dar respuesta a las solicitudes indicadas por el usuario, de forma ágil y lo más eficientemente posible.

### RNF6: Eficiencia

- Disponible en todo momento debido a la importancia de la información que gestiona.
- Los reportes y estadísticas que se obtendrán deben ser 100% precisos y reales.

### RNF7: Portabilidad

- El producto deberá funcionar en los sistemas operativos Windows y Linux.

### RNF8: Restricciones en el diseño y la implementación

- El producto se desarrollará con PostgreSQL 8.4.
- La herramienta manejadora de datos será PgAdmin3 1.10.

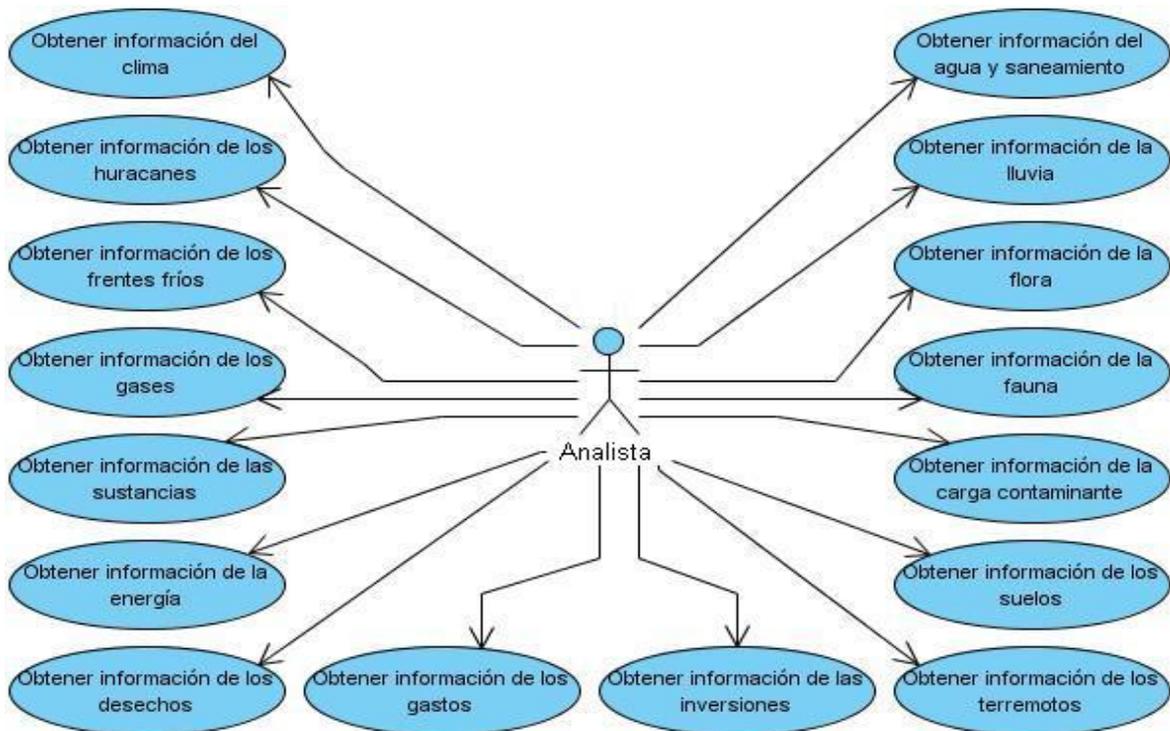
- Para el modelado de los datos se utilizará el Visual Paradigm 6.4.

## 2.7 Casos de uso del sistema

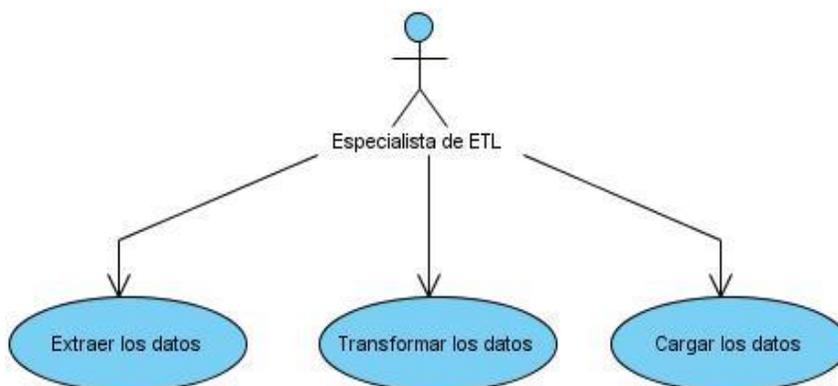
La descripción del sistema es una vereda significativa para el desarrollo del producto mostrando a través de una representación gráfica y descriptiva el funcionamiento de este. En el mismo juega un papel primordial el actor que no es más que aquellas personas clasificadas como terceros que se encuentran fuera del sistema y que trabajan con él. Estos interactúan con los casos de usos que son artefactos narrativos que describen, bajo la forma de acciones y reacciones, el comportamiento del sistema desde el punto de vista del usuario. Por lo tanto, se establece una relación entre los casos de usos y actores sobre las condiciones y requisitos que debe cumplir el sistema.

### 2.7.1 Diagrama de casos de uso

Un diagrama de casos de uso del sistema representa gráficamente a los procesos y su interacción con los actores, los cuales van hacer los roles definidos anteriormente; el especialista de ETL interactúa con los casos de usos funcionales y el analista es el usuario que obtiene la información referente a los Indicadores del Medio Ambiente. A continuación se mostrarán los diferentes diagramas.



**Figura 3: Diagrama de casos de usos de información**



**Figura 4: Diagrama de caso de uso funcional**

### 2.7.2 Descripción de casos de usos

Para entender la funcionalidad asociada a cada caso de uso no es suficiente con la representación gráfica del diagrama de casos de uso. La descripción puede ser elaborada de forma breve o extendida y debe ir acompañada del prototipo respectivo. El prototipo del sistema que se construye en este punto da una visión de las pantallas diseñadas para cada caso de uso que se presenta al usuario para verificar los requisitos de información y funcionales. Ver anexo 1 para observar algunas descripciones de los casos de uso, los restantes se encontrarán en el artefacto Modelo de casos de uso del sistema que estará guardado en el Centro de Tecnologías de Almacenamiento de Datos.

## 2.8 Modelo dimensional

Se desarrollará el modelo dimensional sobre la información capturada por la fase de análisis. Con relación a las necesidades del Indicador Medio Ambiente, se modelará las 54 series correspondientes a este. Por lo que se identificarán las dimensiones, tabla de hechos y las medidas con el propósito de lograr un buen diseño.

### 2.8.1 Tabla de hechos

Luego de haber estudiado los epígrafes anteriores, se extrajeron las temáticas principales con que cuenta el Indicador Medio Ambiente. A continuación se muestran las tablas de hechos que servirán de repositorio central a diferentes dimensiones.

Tabla de hecho temperatura:

Las temperaturas es el valor más alto o más bajo respectivamente, entre los valores máximos o mínimos de temperatura diaria, la tabla estará asociada a varias dimensiones como son: temporal, dpa, estaciones de monitoreo y el período.

Tabla de hecho clima:

El clima recoge varios valores estadísticos relacionados con las temperaturas, cantidad de días con lluvia, rapidez del viento entre otras, tiene asociación con las dimensiones: dpa, estaciones de monitoreo y años.

Tabla de hecho factor potencial de agotamiento del ozono:

El factor PAO significa potencial de agotamiento de la ozono por lo que está coligados a las sustancias agotadoras de la capa de ozono recogiendo los datos de la misma. Dicha tabla se relaciona con las dimensiones correspondientes para conocer su comportamiento por años.

Tabla de hecho acidez de la lluvia:

Recoge la cantidad de lluvia anual teniendo en cuenta el pH. A dicha tabla le pertenece las dimensiones: estaciones de monitoreo y años.

Tabla de hecho capa de ozono:

Esta tabla recoge el consumo de sustancias y se relaciona con las dimensiones: sustancias controladas y los años.

Tabla de hecho huracanes:

Muestra la cantidad de huracanes que azotaron a Cuba, por lo que es importante conocer su comportamiento por categorías de huracanes, período, regiones, año y temporada ciclónica, las cuales son las dimensiones en cuestión.

Tabla de hecho frentes fríos:

Recopila la cantidad de frentes fríos que entró a Cuba, por lo que está asociada a diversas dimensiones dentro de ellas están: las intensidades de frente frío, el período, las regiones y la temporada que le corresponde.

Tabla de hecho concentración de gases:

Recopila el valor promedio de concentración de  $\text{SO}_2$  y  $\text{NO}_2$  y se relaciona a varias dimensiones que brinda el comportamiento de este por estaciones de monitoreo y años.

### Tabla de hechos gases:

La tabla obtiene las emisiones de gases de efecto invernadero. A la misma le concierne las dimensiones: gases de efecto invernadero, actividades del uso de gases y años.

### Tabla de hechos energía:

Contiene la oferta de energía, los dispositivos y biomasa empleada como combustible y las dimensiones relacionadas son: los diferentes dispositivos y biomasa además del año.

### Tabla de hechos cobertura de agua potable y saneamiento:

La medida recopilada es la cantidad de población con cobertura de agua potable y saneamiento y sus correspondientes dimensiones son: cobertura de agua y saneamiento, zona, dpa y el año.

### Tabla de hecho lluvia:

Las dimensiones asociadas a dicha tabla son: dpa, mes, año y recoge la cantidad de lluvia media.

### Tabla de hecho recursos de agua dulce renovable:

La tabla recoge las precipitaciones y los recursos regulares de agua dulce y se asocia a la dimensión año.

### Tabla de hecho biota cubana:

Esta tabla servirá como fuente de información principal a las dimensiones: grupos taxonómicos, hábitat de especies, especies de la biota cubana, nombres comunes de especies de la biota cubana y año. La misma almacena la cantidad de especies.

### Tabla de hecho sistema de acueducto y alcantarillado:

La tabla recolecta los distintos sistemas de acueducto y alcantarillado, por lo que se asociará a las dimensiones con dichos nombres además del año.

### Tabla de hecho extracción de agua:

Recopila la variación y cantidad de extracción de agua dulce, la misma le corresponde las dimensiones: extracción de agua dulce y años.

### Tabla de hecho endemismo vegetal:

La medida que corresponde a esta tabla es la cantidad de endemismo vegetal, por lo que se asocia a las dimensiones: regiones geográficas, los distritos geográficos y el año.

### Tabla de hecho flora:

Cuenta con varias dimensiones atadas, dentro de ellas se encuentran: categoría de flora, instituciones de la flora, el año y la medida que almacena es la cantidad de especies de flora.

### Tabla de hecho fauna:

Recoge la cantidad de especies de fauna y sus dimensiones correspondientes son: categoría de fauna, la fauna animal y el año en cuestión.

### Tabla de hechos tratamiento y recolección de desechos:

Recopila como medida: la cantidad de tratamiento y recolección, por lo que dicha tabla se asociará a las dimensiones: tratamiento y recolección además del año.

### Tabla de hechos carga contaminante:

La carga dispuesta y la variación de la carga son las medidas correspondidas a este hecho, además de las dimensiones: dpa, año y cuencas hidrográficas.

### Tabla de hechos volumen de desecho:

Como su nombre lo indica recoge el volumen de desechos sólidos y sus dimensiones asociadas es el dpa y el año.

### Tabla de hechos producción de materias primas:

Recoge la cantidad de producción de materias primas y se solidariza a las dimensiones: productos de producción de materias primas y años.

### Tabla de hechos afectaciones por huracanes:

Las dimensiones que corresponden a esta tabla de hecho son: los huracanes, causas de pérdidas económicas, dpa y año. La misma recoge la cantidad de viviendas dañadas y pérdidas económicas.

### Tabla de hecho afectaciones por incendios forestales:

Las medidas que se centran en esta tabla son: la cantidad de incendios forestales y superficie dañada, por lo que se afilia a las dimensiones, dpa y años.

### Tabla de hechos áreas protegidas:

Recopila la cantidad de áreas y las diferentes dimensiones correspondientes a este hecho son: superficie terrestre, significación local, categoría de manejo, dpa y el año.

Tabla de hechos áreas protegidas con reconocimiento internacional:

Esta tabla se relaciona con algunas dimensiones, dentro de ellas se encuentran: las áreas protegidas, el dpa y el año, la misma recoge la cantidad de zona de núcleo, amortiguamiento y transición.

Tabla de hechos agroproductiva de los suelos:

La medida recogida en esta tabla es la productividad de los suelos teniendo en cuenta las dimensiones: clasificación agroproductiva de los suelos, el dpa y el año.

Tabla de hechos terremotos:

La magnitud, profundidad y la intensidad son las medidas recogidas por esta tabla también se asociará a ciertas dimensiones, las cuales se nombran: localidades cubanas, dpa y temporal.

Tabla de hechos gastos para la protección del medio ambiente:

Colecciona los gastos de inversión para el medio ambiente, por lo que está relacionado a las siguientes dimensiones: sectores ambientales, dpa y el año.

Tabla de hechos inversiones para el medio ambiente:

Recopila la cantidad de inversiones para el medio ambiente. Esta tabla servirá de repositorio de información a las dimensiones consecuentes: actividades económicas, actividades ambientales y el año.

Tabla de hechos genética de los suelos:

La medida recogida en esta tabla es la cantidad de suelos teniendo en cuenta las dimensiones: clasificación genética de los suelos, el dpa y el año.

Tabla de hechos principales factores limitantes edáficos:

En esta tabla es donde va a residir, como repositorio central, toda la información existente de las dimensiones factores limitantes edáficos y el año.

### 2.8.2 Dimensiones

Luego de la obtención de las tablas de hechos, se precisaron múltiples dimensiones que posteriormente se utilizará para la realización del diseño dimensional; entre ellas se encuentran:

Año (dim\_anno): Se almacena el año de la información.

Atributos asociados:

anno\_id: Identificador de la dimensión.

anno\_número: Almacena el número de cada año.

Mes (dim\_mes): Se almacenan los meses de la información.

Atributos asociados:

mes\_id: Identificador de la dimensión.

mes\_número: Almacena el número de los meses de la información.

mes\_descripción: Se almacena el nombre de cada mes.

DPA (dim\_dpa): Se almacenan todas las provincias del país.

Atributos asociados:

dpa\_id: Es la llave primaria de la dimensión.

dpa\_cod: Almacena el código oficial de las provincias establecidos dentro de la DPA.

dpa\_nombre: Se almacena el nombre de cada provincia.

Temporal (dim\_temporal): En esta dimensión se almacenan los días, meses y años de la información que se requiere.

Atributos asociados:

temporal\_id: Es la llave primaria de la dimensión.

días\_num: Se almacena el número del día de la información.

mes\_num: Se almacena el número del mes de la información.

anno\_número: Se almacena el número del año de la información.

Categoría de huracanes (dim\_categoría\_huracanes): Se almacenan las diferentes categorías de los huracanes.

Atributos Asociados:

categ\_huracanes\_id: Es la llave primaria de la dimensión.

categ\_huracanes\_tipo: Se almacenan los diferentes tipos de huracanes según la escala Simpson.

Intensidades de los frentes fríos (dim\_intensidades\_frente\_frío): Se almacenan las diferentes intensidades de todos los frentes fríos.

Atributos asociados:

intensidades\_id: Es la llave primaria de la dimensión.

categoría\_intensidades\_frente\_frío: Se almacena las diferentes intensidades de los frentes fríos como pueden ser: moderados, débiles y fuertes.

Regiones (dim\_regiones): Se almacenan las regiones del país.

Atributos asociados:

regiones\_id: Es la llave primaria de la dimensión.

regiones\_nombre: Se almacena el nombre las 3 regiones del país.

Período (dim\_período): Se almacenan los períodos de la información que se desea.

Atributos asociados:

período\_id: Es la llave primaria de la dimensión.

período\_número: Se almacena el número de los períodos de la información deseada.

Dispositivos (dim\_dispositivos): Se almacenan los diferentes dispositivos relacionados con la energía.

Atributos asociados:

dispositivos\_id: Es la llave primaria de la dimensión.

dispositivos\_nombre: Se almacena el nombre de cada uno de los dispositivos de energía.

Biomasa (dim\_biomasa): Se almacenan los diferentes materiales que son usados como combustible para generar energía.

Atributos asociados:

biomasa\_id: Es la llave primaria de la dimensión.

nombre: Se almacena cada uno de los materiales relacionados con la biomasa.

Estaciones meteorológicas (dim\_estaciones\_meteorológicas): Se almacenan las estaciones de monitoreo que existen en todo el país.

Atributos asociados:

estaciones\_id: Es la llave primaria de la dimensión.

estaciones\_nombre: Se almacena el nombre de cada estación de monitoreo del país.

Actividades del uso de los gases (dim\_actividades\_del\_uso\_gases): Se almacenan las actividades en las que se ven involucrados los gases.

Atributos asociados:

actividades\_id: Es la llave primaria de la dimensión.

actividades\_tipo: Se almacenan los tipos de actividades donde se involucran cada uno de los gases de la información que se desea.

Cobertura de agua potable y saneamiento (dim\_cobertura\_agua\_potable\_saneamiento): Se almacena hasta donde el agua potable es capaz de llegar, la población que tiene acceso a esta agua y la población que tiene acceso a saneamiento.

Atributos asociados:

cobertura\_agua\_potable\_saneamiento\_id: Es la llave primaria de la dimensión.

tipo\_agua\_potable: Se almacena la población que tiene acceso al agua potable ya sea por conexión domiciliaria, servicio público o de fácil acceso.

tipo\_saneamiento: Se almacena la población que tiene acceso a saneamiento ya sea por alcantarillado y fosas y letrinas.

Zona residencial (dim\_zona\_residencial): Se almacena la zona donde está presente la información.

Atributos asociados:

zona\_id: Es la llave primaria de la dimensión.

tipo\_zona: Se almacenan los tipos de zonas que pueden ser rural o urbana.

Extracción del agua dulce (dim\_extracción\_agua\_dulce): Se almacenan los diferentes lugares de donde es extraída el agua y las actividades en las cual se utiliza.

Atributos asociados:

destino\_extracción\_id: Es la llave primaria de la dimensión.

tipo\_extracción: Se almacena de donde es extraída el agua, si por la industria o directa.

tipo\_actividad: Se almacenan las actividades donde es usada esa agua.

Alcantarillado (dim\_alcantarillado): Se almacena el sistema de alcantarillado.

Atributos asociados:

alcantarillado\_id: Es la llave primaria de la dimensión.

nombre: Se almacena el nombre de los distintos aspectos a pedir como son: extensión, localidades benéficas, número de sistema de tratamiento, volumen evacuado, entre otras.

unidad\_medida: Se almacenan las unidades de medidas.

Acueductos (dim\_acueductos): Se almacena el sistema de acueductos.

Atributos asociados:

acueductos\_id: Es la llave primaria de la dimensión.

nombre: Se almacena el nombre de los distintos aspectos a pedir como son: extensión de la red, número de plantas potabilizadoras, volumen de agua suministrada, entre otras.

unidad\_medida: Se almacenan las distintas unidades de medidas de los resultados que se mostrarán.

Distritos geográficos (dim\_distritos\_geográficos): Se almacenan los distritos de la información requerida.

Atributos asociados:

distritos\_id: Es la llave primaria de la dimensión.

nombre: Se almacena el nombre de cada distrito.

Instituciones de la flora (dim\_instituciones\_flora): Se almacena información de las instituciones.

Atributos asociados:

instituciones\_id: Es la llave primaria de la dimensión.

nombre: Se almacena el nombre de cada institución.

Categoría de la flora (dim\_categoría\_flora): Se almacena información sobre la flora amenazada.

Atributos asociados:

flora\_categoría\_id: Es la llave primaria de la dimensión.

categoría\_tipo: Se almacena el tipo de amenaza.

Categoría de la fauna (dim\_categoría\_fauna): Se almacena información sobre la fauna amenazada.

Atributos asociados:

fauna\_categoría: Es la llave primaria de la dimensión.

tipo\_categoría: Se almacena el tipo de amenaza ya sea si son: extintas, especie en peligro crítico, especies en peligro, especies vulnerables, entre otras.

Fauna Animal (dim\_fauna\_animal): Se almacena información de la fauna.

Atributos asociados:

fauna\_animal\_id: Es la llave primaria de la dimensión.

tipo: Se almacena si son vertebrados o invertebrados.

Áreas protegidas (dim\_áreas\_protegidas): Se almacena información sobre las áreas protegidas del país.

Atributos asociados:

áreas\_protegidas\_id: Es la llave primaria de la dimensión.

nombre: Se almacena el nombre de cada una de estas áreas.

Nombres comunes de la biota cubana (dim\_nombres\_comunes\_especies\_biota\_cubana): Se almacena información de los nombres comunes que se le llaman a las especies de la biota cubana.

Atributos asociados:

nombres\_comunes\_id: Es la llave primaria de la dimensión.

nombres: Se almacena el nombre de cada especie de la biota cubana.

Grupos taxonómicos (dim\_grupos\_taxonómicos): Se almacena información de las diferentes especies de la biota cubana.

Atributos asociados:

grupos\_taxonómicos\_id: Es la llave primaria de la dimensión.

nombre: Se almacena el nombre científico dado a estas especies de la biota cubana.

Cuencas hidrográficas (dim\_cuencas\_hidrográficas): Se almacena información de la Carga contaminante que existen en las cuencas.

Atributos asociados:

ch\_id: Es la llave primaria de la dimensión.

ch\_nombre: Se almacena información del nombre de las cuencas.

Especies de la biota cubana (dim\_especies\_de\_la\_biota\_cubana): Se almacena información de las especies de la biota cubana.

Atributos asociados:

especies\_id: Es la llave primaria de la dimensión.

especies\_descripción: Se almacena información sobre las especies como: las endémicas, conocidas e introducidas.

Hábitat de las especies (dim\_hábitat\_especies): Se almacena información sobre el ambiente en que se desarrollan las especies.

Atributos asociados:

ambiente\_id: Es la llave primaria de la dimensión.

tipo\_ambiente: Se almacena información sobre el ambiente en que se desarrollan.

Productos de producción de materias primas (dim\_productos\_producción\_materias\_primas): Se almacena información sobre los productos para la producción de materias primas.

Atributos asociados:

productos\_id: Es la llave primaria de la dimensión.

tipo: Se almacena el tipo de producto si es metal ferroso o no.

unidad\_medida: Se almacena la unidad de medida de cada producto.

Tratamiento y recolección de desechos (dim\_tratamiento\_recolección\_desechos): Se almacena información de cómo se comporta el tratamiento y la recolección de desechos sólidos.

Atributos asociados:

tratamiento\_recolección\_id: Es la llave primaria de la dimensión.

tipo: Se almacena información en el tipo de recolección y tratamiento.

unidad\_medida: Se almacena información de las unidades de medidas para cada tratamiento y recolección de desechos sólidos.

Significación de áreas protegidas (dim\_significación\_áreas\_protegidas): Se almacena información sobre la significación de las áreas protegidas.

Atributos asociados:

significación\_id: Es la llave primaria de la dimensión.

significación\_descripción: Se almacena información sobre la significación del área, si es de significación local o de significación nacional.

Categoría de manejo de las áreas protegidas (dim\_categoría\_manejo\_áreas\_protegidas): Se almacena información sobre donde las áreas protegidas estarán concentradas.

Atributos asociados:

categoría\_manejo\_id: Es la llave primaria de la dimensión.

tipo\_categ\_manejo: Se almacena información sobre el tipo de categoría como: parque nacional, reserva ecológica y reserva florística manejada.

Superficie (dim\_superficie): Se almacena información sobre la superficie de las áreas protegidas.

Atributos asociados:

dim\_superficie\_id: Es la llave primaria de la dimensión.

superficie\_tipo: Se almacena información sobre la superficie si será terrestre o marina.

Localidades geográficas (dim\_localidades\_geográficas): Se almacena información sobre las localidades donde ocurren los terremotos.

Atributos asociados:

localidades\_id: Es la llave primaria de la dimensión.

nombre: Se almacena información del nombre de las localidades.

Sectores ambientales (dim\_sectores\_ambientales): Se almacena información sobre gastos que se hacen en los diferentes sectores ambientales.

Atributos asociados:

sectores\_ambientales\_id: Es la llave primaria de la dimensión.

nombre: Se almacena información de los nombres de los diferentes sectores ambientales.

Actividad ambiental (dim\_actividad\_ambiental): Se almacena información sobre las inversiones por actividades ambientales.

Atributos asociados:

actividad\_ambiental\_id: Es la llave primaria de la dimensión.

nombre: Se almacena información de los nombres de las actividades ambientales.

Actividad consumidora (dim\_actividades): Se almacena información sobre las sustancias agotadoras de la capa de ozono por actividad consumidora.

Atributos asociados:

actividades\_id: Es la llave primaria de la dimensión.

nombre: Se almacena información de los nombres de las actividades consumidoras.

Actividades económicas (dim\_actividades\_económicas): Se almacena información sobre las inversiones por actividades económicas.

Atributos asociados:

actividades\_económicas\_id: Es la llave primaria de la dimensión.

nombre: Se almacena información de los nombres de las actividades económicas.

Clasificación genética de los suelos (dim\_clasificación\_genética\_suelos): Se almacena información sobre la genética de los suelos.

Atributos asociados:

clasificación\_suelos\_id: Es la llave primaria de la dimensión.

tipo\_suelos: Se almacenan los tipos de suelos que existen en Cuba.

Clasificación agroproductiva de los suelos (dim\_clasificación\_agroproductica\_suelos): Se almacena información sobre la productividad de los suelos.

Atributos asociados:

clasif\_agroproductiva\_id: Es la llave primaria de la dimensión.

tipo\_productividad\_suelos: Se almacena el tipo de producción de los suelos si es muy productiva, poco productiva.

Factores limitantes edáficos (dim\_factores\_limitantes\_edáficos): Se almacena información sobre los principales factores limitantes edáficos.

Atributos asociados:

fact\_limit\_edáficos\_id: Es la llave primaria de la dimensión.

nombre: Se almacena el nombre de los diferentes factores.

Causas de pérdidas económicas (dim\_causas\_pérdidas\_económicas): Se almacena información sobre las diferentes causas de pérdidas económicas debido a las afectaciones por huracanes.

Atributos asociados:

causas\_id: Es la llave primaria de la dimensión.

tipo\_causas: Se almacena información de los distintos tipos de causas de pérdidas económicas.

Huracanes (dim\_huracanes): Se almacena información de los diferentes huracanes azotados a Cuba

Atributos asociados:

huracanes\_id: Es la llave primaria de la dimensión.

nombre: Se almacena información de los nombres de los huracanes azotados en Cuba.

Gases de efecto invernadero (dim\_gases\_efecto\_invernadero): Se almacena información de los diferentes gases de efecto invernadero.

Atributos asociados:

gases\_id: Es la llave primaria de la dimensión.

gases\_tipo: Se almacena información de los tipos de gases de efecto invernadero.

Sustancias controladas (dim\_sustancias\_controladas): Se almacena información de las distintas sustancias controladas de la capa de ozono.

Atributos asociados:

sust\_controladas\_id: Es la llave primaria de la dimensión.

nombre: Se almacena información de los nombres de cada una de las sustancias.

Las dimensiones poseen como característica principal la definición de las jerarquías entre sus atributos, inmediatamente de haber definido las dimensiones candidatas se localizó una de ellas, a continuación se especificará: Dimensión temporal: Año -> Mes -> Día

### 2.8.3 Medidas

Las medidas numéricas definidas se almacenarán en las tablas de hechos. A continuación aparecerán los extraídos de cada temática.

1. Temperaturas mínimas.
2. Temperaturas máximas.
3. Cantidad de días con lluvia.
4. Nubosidad media.
5. Humedad Relativa.
6. Dirección del viento.
7. Rapidez del viento.
8. Cantidad de lluvia media.
9. PH de la lluvia.
10. Variación del consumo de sustancias agotadoras.
11. Consumo de sustancias.
12. Cantidad de huracanes.
13. Cantidad de frente frío.
14. Oferta de energía renovable.
15. Dispositivos como combustible.
16. Biomasa como combustible.
17. Emisiones brutas de gases.
18. Valor promedio de la concentración de SO<sub>2</sub>.
19. Valor promedio de la concentración de NO<sub>2</sub>.
20. Precipitaciones.
21. Recurso de agua dulce renovable.
22. Cantidad de población con cobertura.
23. Sistemas de acueducto y alcantarillado.
24. Variación de extracción de agua dulce.
25. Participación de destinos.

- |   |   |
|---|---|
| 26. Extracción de agua por destinos.          | 42. Cantidad de incendios forestales.           |
| 27. Cantidad de endemismo vegetal.            | 43. Variación de incendios forestales.          |
| 28. Cantidad de endemismo vegetal.            | 44. Superficie dañada.                          |
| 29. Cantidad de especies.                     | 45. Zona núcleo.                                |
| 30. Carga dispuesta.                          | 46. Zona de amortiguamiento.                    |
| 31. Variación de carga.                       | 47. Zona de transición.                         |
| 32. Cantidad de tratamiento y recolección.    | 48. Carga dispuesta.                            |
| 33. Cantidad de especies de flora.            | 49. Carga reducida.                             |
| 34. Especies floras en por ciento.            | 50. Gasto de inversión.                         |
| 35. Volumen de desechos sólidos.              | 51. Magnitud.                                   |
| 36. Producción de materias primas recicladas. | 52. Profundidad.                                |
| 37. Por ciento de materias primas.            | 53. Intensidad.                                 |
| 38. Cantidad de especies de fauna.            | 54. Cantidad de inversiones del Medio Ambiente. |
| 39. Cantidad de áreas protegidas.             | 55. Cantidad de suelos.                         |
| 40. Cantidad de viviendas dañadas.            | 56. Productividad.                              |
| 41. Cantidad de pérdidas económicas.          | 57. Área agrícola afectada en por ciento.       |
| 58. Área agrícola afectada.                   |   |

Luego de haber definido las dimensiones, medidas y los hechos, se procede a la estructuración del modelo o los modelos dimensionales. Se puede resaltar que por las necesidades actuales del negocio existe un modelo relacionado con el Indicador Medio Ambiente los cuales agrupan las dimensiones y las medidas puntualizadas.

La solución propuesta (ver anexo 1) se desarrolló basada en el modelo tipo estrella, donde existirá 33 tablas centrales asociadas a 45 dimensiones y 58 medidas. Cada de una de estas dimensiones posee una llave primaria que es la encargada de mantener la integridad referencial entre ellas y los hechos.

### 2.8.4 Matriz Bus

Seguidamente de realizar el modelo lógico del trabajo correspondiente al Indicador ambiental, se perpetrará la matriz BUS que demostrará a través de una tabla las relaciones entre los hechos y las dimensiones, como el diseño de la solución es muy grande solo se mostrará una pequeña parte de la matriz dimensional. Para observar completamente estas relaciones dirigirse a la tabla de Excel: [Matriz BUS.xls](#)

Hechos	Dimensiones										
	Año	Mes	DPA	Temporal	Categ_ Huracanes	Regiones	Periodo	Estaciones	Sustancias	Clasif_ genética	Intensid_ frente frío
Huracanes	x	x			x	x	x				
Clima	x		x					x			
Concentración de gases	x							x			
Factor PAO	x								x		
Genética de los suelos	x		x							x	
Lluvias	x	x	x								
Temperaturas			x	x			x		x		
Frentes de frío		x				x	x				x

### 2.9 Política de respaldo y recuperación

No es ninguna novedad el valor que tiene la información y los datos para las empresas. Lo que resulta increíble es la falta de precauciones que se suele tener al confiar al núcleo de los negocios al sistema de almacenamiento de lo que en la mayoría de los casos resulta ser una computadora pobremente armada tanto del punto de vista de hardware como de software. Si el monitor, la memoria e incluso la CPU

(Unidad Central de Procesamiento) del computador dejan de funcionar, simplemente se reemplaza y no hay mayores dificultades. Pero si falla el disco duro, el daño puede ser irreversible y existirá pérdida de los datos, por esta razón, se debe respaldar la información más importante.

Si esto llegara a suceder a la ONE las pérdidas económicas podría ser cuantiosas por lo que se propuso aplicar los respaldos de información en caso de que una falla o error se presente. La política de respaldo que se emplea en la solución se mide por el siguiente punto:

Una vez al mes se realizarán backups (salvas) de la información total que posee la BD del Indicador ambiental, así lo tiene definido actualmente la organización, garantizando que exista una copia exacta de la información que está vigente en el servidor. Se hará mensualmente debido a que la ONE tiene información valiosa de los Indicadores porque muestran datos de interés nacional. En caso de que ocurra un desastre natural estará disponible un servidor de respaldo para salvar todas las estadísticas.

### **Conclusiones**

Todos estos espacios desarrollados posibilitaron dar cumplimiento al segundo objetivo. Se definió el proceso del negocio para conocer el comportamiento del flujo de información del Indicador Medio Ambiente y el tema de análisis se centralizó en este Indicador. Se realizó un largo estudio de las 54 series ambientales para extraer todas las informaciones que ayudarán a desarrollar el mercado de datos, por lo que se extrajeron 6 reglas del negocio, se establecieron 16 requisitos de información y las variables de entrada y salida de cada tabla de Excel, además se establecieron 3 requisitos funcionales y 8 requisitos no funcionales; algunos de estos son: software, hardware, eficiencia, entre otros. Para entender el funcionamiento del sistema se realizaron los diagramas y las descripciones de casos de usos, además se determinaron los roles y las acciones que debe hacer al interactuar con el sistema. Luego de realizar el proceso de análisis de todas las series ambientales se realizó el modelo dimensional; para ello se identificaron 58 medidas, 33 tablas de hechos, 45 dimensiones y se realizó la matriz BUS para relacionarlos. Para la culminación se efectuó todos los artefactos que se generaron del análisis y diseño para el Centro de Tecnologías de Almacenamiento de Datos. La ejecución de este significativo capítulo valdrá de base para el próximo donde se presenta la implementación y prueba del sistema a desarrollar.

## **CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA**

### **Introducción**

Desde un inicio las bases de datos se convirtieron en una herramienta fundamental de control y manejo de las operaciones comerciales. Fue así como en unos pocos años en grandes empresas y negocios existía un considerable número de información almacenada en diferentes fuentes de datos y estas ya habían alcanzado un tamaño considerablemente grande. Lo ideal es unificar las diferentes fuentes de información de las cuales se disponen, en un único lugar, al que solo se le incorpora información relevante, sobre la base de una estructura organizada, integrada, lógica, dinámica y de fácil explotación. La respuesta sería realizar un almacén de datos el cual constituye uno de los soportes fundamentales para el proceso de toma de decisiones gerenciales; de ahí la importancia de que la información guardada sea confiable y con calidad. Para el desarrollo del almacén de datos de la Oficina Nacional de Estadística se realizará la implementación y prueba del mercado de datos del Indicador Medio Ambiente donde el presente capítulo se centralizará en esta senda.

### **3.1 Modelo de datos físico**

Un modelo de datos es un lenguaje utilizado para la descripción de una base de datos. Por lo general, un modelo de datos permite describir las estructuras de datos de la base (el tipo de los datos que incluye la base y la forma en que se relacionan), las restricciones de integridad (las condiciones que los datos deben cumplir para reflejar correctamente la realidad deseada) y las operaciones de manipulación de los datos (agregado, borrado, modificación y recuperación de los datos de la base). En un enfoque más amplio, un modelo de datos permite describir los elementos que intervienen en una realidad o en un problema dado y la forma en que se relacionan dichos elementos entre sí.

La clasificación de los modelos de datos se realiza de acuerdo al nivel de abstracción. Los modelos de datos conceptuales son aquellos que describen las estructuras de datos y restricciones de integridad. Se utilizan durante la etapa de análisis de un problema dado y están orientados a representar los elementos que intervienen y sus relaciones. Los modelos de datos lógicos se centran en las operaciones y se implementan en algún manejador de base de datos. Por último, los modelos de datos físicos, que son estructuras de datos a bajo nivel implementadas dentro del propio manejador. (Definición , 2008)

En otro término se dice que el modelo de datos físico describe como se almacenan los datos en la computadora, representando información tal como las estructuras de registro, el ordenamiento de los

registros y las rutas de acceso. (Gabriel Requielme, 2008) Los modelos físicos se crean a partir de los modelos lógicos, el cual no es más que el modelo dimensional elaborado en el capítulo anterior. A continuación se muestra cómo es el proceso de la implementación del mercado de datos por lo que se manifestará la estructura de la base de datos, en el cual se describirán los objetos que lo conforman como son: esquemas, tablas, secuencias e índices; se instaurarán los usuarios de la base de datos con los correspondientes privilegios y por último se mostrará la carga de cada uno de los nomencladores.

### 3.1.1 Estructuras de datos

Una estructura de datos es una forma de instaurar un conjunto de datos elementales con el objetivo de facilitar su manipulación. Los datos almacenados están ordenados por tablas que cada una pertenece a un esquema. Esto posibilita una adecuada organización y entendimiento de los mismos, a continuación se abordarán estos aspectos además de las secuencias e índices.

#### 3.1.1.1 Esquemas y tablas

El esquema y las tablas permiten obtener una clara idea general de la base de datos, analizar y profundizar los contenidos básicos. La investigación cuenta con 78 tablas identificadas, para una buena organización de las mismas se mostrará la manera en que están agrupadas por esquemas (dimensiones o hechos). Ver anexo 3 para ver los restantes resultados.

**Tabla:** Hech\_Huracanes **Esquema:** hechos

PK	FK	Nombre	Tipo de datos	No null
✓	✓	categ_huracanes_id	integer	✓
✓	✓	período_id	integer	✓
✓	✓	regiones_id	integer	✓
✓	✓	anno_id	integer	✓
✓	✓	mes_id	integer	✓
		cant_huracanes_dif_int_azotado_cuba	char(10)	

**Tabla:** dim\_actividad\_ambiental **Esquema:** dimensiones

PK	FK	Nombre	Tipo de datos	No null
✓		actividad_ambiental_id	serial	✓
		nombre	varchar(80)	

**Tabla:** dim\_actividades\_consumidoras **Esquema:** dimensiones

PK	FK	Nombre	Tipo de datos	No null
✓		actividades_id	serial	✓
		nombre	varchar(80)	

**Tabla:** dim\_actividades\_del\_uso\_gases **Esquema:** dimensiones

PK	FK	Nombre	Tipo de datos	No null
✓		actividades_id	serial	✓
		actividades_tipo	varchar(80)	

**Tabla:** dim\_actividades\_económicas **Esquema:** dimensiones

PK	FK	Nombre	Tipo de datos	No null
✓		actividades_económicas_id	serial	✓
		nombre	varchar(80)	

**Tabla:** dim\_acueducto **Esquema:** dimensiones

PK	FK	Nombre	Tipo de datos	No null
----	----	--------	---------------	---------

✓		acueducto_id	serial	✓
		nombre	varchar(80)	
		unidad_medida	varchar(10)	

**Tabla:** dim\_alcantarillado **Esquema:** dimensiones

PK	FK	Nombre	Tipo de datos	No null
✓		alcantarillado_id	serial	✓
		nombre	varchar(80)	
		unidad_medida	varchar(10)	

**Tabla:** dim\_anno **Esquema:** dimensiones

PK	FK	Nombre	Tipo de datos	No null
✓		anno_id	serial	✓
		anno_número	integer	

### 3.1.1.2 Restricciones y secuencias

Una restricción es una condición que obliga el cumplimiento de ciertas condiciones en la base de datos. Algunas no son determinadas por los usuarios, sino que son inherentemente definidas por el mismo gestor de BD. Algunas restricciones las puede definir el usuario, por ejemplo, usar un campo con valores enteros entre 1 y 10. Las restricciones proveen un método de implementar reglas en la base de datos y restringen los datos que pueden ser almacenados en las tablas. Y las secuencias son atributos que se incrementan secuencialmente durante el ingreso de los datos. En el presente trabajo se definieron aproximadamente 150 llaves primarias y 98 llaves foráneas. Ver anexo 4 para ver las secuencias generadas.

Algunas de las llaves primarias son:

Llaves primarias	Tabla	Esquema	Campo
PK115	hech_Áreas_protegidas	hechos	dim_superficie, significación_id, categoría_manejo_id, anno_id, dpa_id
PK103	Hech_Huracanes	hechos	categ_huracanes_id, período_id, regiones_id, anno_id, mes_id
PK119	hech_Biota_Cubana	hechos	grupos_taxonómicos_id, ambiente_id, especies_id, nombres_comunes_id, anno_id
PK30	hech_Cobertura_Agua_Potable	hechos	cobertura_agua_potable_saneamiento_id, zona_id, anno_id, dpa_id
PK106	hech_Frente_frío	hechos	intensidades_id, período_id, regiones_id, mes_id

Algunas de las llaves foráneas son:

Llaves foráneas	Tabla	Esquema	Campo
Refdim_categoría_huracanes102	Hech_Huracanes	hechos	categ_huracanes_id
Refdim_acueducto131	hech_Acueducto_Alcantarillado	hechos	acueducto_id
Refdim_huracanes67	hech_Afectaciones_Huracanes	hechos	huracanes_id
Refdim_anno208	hech_Afectaciones_Incendios_Forestales	hechos	anno_id
Refdim_dpa151	hech_Afectaciones_Incendios_Forestales	hechos	dpa_id

### 3.1.1.3 Índices

El índice de una base de datos es una estructura de datos que mejora la velocidad de las operaciones, permitiendo un rápido acceso a los registros de una tabla. Al aumentar drásticamente la velocidad de acceso, se suelen usar sobre aquellos campos sobre los cuales se hagan frecuentes búsquedas.

Los índices de una base de datos al igual que los índices de los libros sirven para encontrar más rápido aquello que se busca, por lo tanto se puede decir que sirven para agilizar las consultas a las tablas. Próximamente se mostrarán algunos de los 175 índices que generó automáticamente el gestor, el resto se encontrará en el anexo 5.

Índice	Tabla	Esquema	Tipo	Llave primaria
PK103	Hech_Huracanes	hechos	btree	✓
Ref104104	Hech_Huracanes	hechos	btree	
Ref105105	Hech_Huracanes	hechos	btree	
Ref125181	Hech_Huracanes	hechos	btree	
Ref128182	Hech_Huracanes	hechos	btree	
Ref18102	Hech_Huracanes	hechos	btree	
PK89	dim_actividad_ambiental	dimensiones	btree	✓
PK26	dim_actividades_consumidoras	dimensiones	btree	✓
PK109	dim_actividades_del_uso_gases	dimensiones	btree	✓
PK88	dim_actividades_económicas	dimensiones	btree	✓

### 3.1.2 Usuarios y privilegios

En esta sección se describirá los usuarios del equipo y los privilegios que debe tener cada uno de ellos. Antes de abordar el tema es necesario aclarar que los roles identificados en el capítulo anterior no son los mismos, los que se mostrarán recientemente son los usuarios de la base de datos.

#### Usuarios y roles

Los usuarios y privilegios se crean con el objetivo de garantizar una mejor protección y seguridad de la base de datos; a continuación se especificará cada uno de ellos:

- *Administrador*: Tiene acceso a la administración y configuración de la BD en su totalidad.
- *Analista*: Consultar la información de la BD para analizarlos.
- *Especialista de ETL*: Su función se basa en la realización del proceso de ETL.

## Privilegios

Los privilegios asignados a los usuarios del sistema se basan en el rol que cada uno desempeña. De esta manera el usuario solo tendrá acceso a los datos si se encuentra autenticado en la base de datos. Los mismos tienen los siguientes permisos:

- Tiene derecho de lectura y actualización de la BD: *administrador*.
- Solo tiene permiso de lectura de los datos almacenados en el almacén: *analista*.
- Tiene los privilegios de lectura y escritura de los datos almacenados en el almacén: *especialista de ETL*.

### 3.1.3 Carga de nomencladores

Los nomencladores son todas las dimensiones que fueron encontrados después de haber profundizado cada una de las series estadísticas. Es importante precisar que solo se efectuará la carga de las dimensiones, por lo que esta sección describirá los valores de cada una de estas. Seguidamente se mostrará algunos nomencladores, el resto se encontrará en el anexo 6.

Nomencladores	Valores
nombre_actividad_ambiental	Protección del aire y el clima, Reducción del ruido y las vibraciones, Gestión de las aguas.
nombre_actividades_consumidoras	Aerosoles, Refrigeración, Solventes.
actividades_tipo	Energía, Procesos Industriales, Uso de Solventes.
nombre_actividades_económicas	Agricultura, Ganadería, Caza y Silvicultura, Explotación de Minas y Canteras, Industria Manufacturera.
nombre_acueducto	Extensión de la red, Localidades beneficiadas, Número de plantas potabilizadoras.
nombre_alcantarillado	Extensión, Localidades beneficiadas, Número de sistemas de tratamiento.
anno_número	1985, 1998, 2009.
nombre_áreas_protegidas	Reservas de la Biósfera, Guanahacabibes, Sierra del Rosario.
nombre_biomasa	Bagazo de caña, Leña, Serrín de madera.
tipo_categoria_fauna	Extintas, En Peligro Crítico, En Peligro.

categoria_tipo_flora	Extintas, Especies vulnerables, Casi amenazadas.
categ_huracanes_tipo	SS1 (118-153 km/hora), SS2 (154-177 km/h), SS3 (178 - 209 km/hora)
tipo_categ_manejo	Parque nacional, Reserva ecológica, Reserva florística manejada.
tipo_causas	Gastos en medidas preventivas, Costo de reposición de vivienda, Instalaciones.
tipo_productividad_suelos	Muy productivos, Productivos, Poco productivos.
tipo_genetica_suelos	Alíticos, Ferríticos, Ferríticos.
tipo_agua_potable	Población con acceso a agua potable, Conexión domiciliaria, Servicio público.
ch_nombre	Cuyaguaje, Ariguanabo, Almendares – Vento.
dispositivos_nombre	Molinos de viento, Digestores plantas de biogás, Plantas de biogás.

## 3.2 Guía de implantación

La implantación de un sistema informático es la necesidad que tiene un software para su instalación. Para darle un mantenimiento al trabajo en proposición es necesario crear una guía de implantación que analice los detalles técnicos y organizativos. En este epígrafe se tiene en cuenta los requerimientos no funcionales que son los utilizados para el desarrollo de la aplicación y se muestra los pasos a seguir para instalar adecuadamente la base de datos del Indicador ambiental.

### 3.2.1 Requerimientos

Los requerimientos del sistema fueron definidos en el capítulo anterior donde se identificaron los requisitos funcionales que será: realizar el proceso de ETL para cada una de las series de los Indicadores del Medio Ambiente y los requisitos no funcionales son las cualidades que debe tener el producto; los identificados fueron los de hardware, software, usabilidad, fiabilidad, rendimiento, eficiencia, portabilidad y restricciones en el diseño y la implementación. Estos últimos requisitos mencionados son los que se necesitan para la realización de la aplicación.

### 3.2.2 Secuencia de pasos

Para instalar la base de datos es necesario seguir un grupo de pasos que a continuación se precisan:

1. Primeramente debe estar instalado el gestor de base de datos PostgreSQL 8.4.

2. Debe estar instalada la herramienta de administración de base de datos definida, el PgAdmin3 1.10.
3. Se debe crear una nueva base de datos utilizando la herramienta de administración de bases de datos.
4. Luego se crean los usuarios: administrador, analista y especialista de ETL para que obtengan sus permisos definidos anteriormente.
5. Se carga el script DDL\_Medio\_Ambiente.SQL para crear la estructura física del almacén.
6. Teniendo la estructura de la BD están creadas las condiciones para correr el script DML\_Medio\_Ambiente.SQL quedando así cargados los nomencladores en el almacén.
7. Al estar la BD creada estructuralmente con los nomencladores cargados se debe copiar la información que se encuentra en el fichero DCL\_Medio\_Ambiente y correrlo, así los usuarios creados tendrán los permisos definidos en el almacén.

### **3.3 Validación y pruebas**

La validación y las pruebas, estas actividades en conjuntas se realizan para garantizar que el producto cumpla con las exigencias propuestas por el cliente y que la solución tenga la calidad requerida. Para la obtención de un mercado de datos eficiente y aceptado por el usuario es imprescindible realizar las listas de chequeos a los artefactos realizados en la fase de análisis y diseño, validar los requisitos por el cliente y realizar las pruebas de implantación. A continuación se describen los procedimientos realizados por cada uno de los citados anteriormente.

#### **3.3.1 Listas de chequeo análisis**

Las listas de chequeos son documentos que evalúan el comportamiento de los artefactos que se generaron en el análisis. Estas se realizan con el propósito de que los mismos cumplan con el formato predefinido y que la información recopilada tenga una estructura organizada relacionada con el tema de análisis del Indicador Medio Ambiente. Las listas que se realizaron al análisis se encuentran guardadas en DATEC, las mismas son:

- Lista de chequeo Especificación de requisitos.
- Lista de chequeo Especificación de las áreas de la organización.
- Lista de chequeo Herramienta para la recolección y análisis de la información.

Luego de realizar cada una de las listas de chequeo mencionadas se debe resaltar que los resultados fueron satisfactorios. La calificación por el evaluador fue de B, el documento no contiene errores ortográficos, cumple con el estándar de la plantilla entregada por el centro y la información es asequible y entendible para los usuarios que desean utilizarlos.

### **3.3.2 Validación de requisitos por el cliente**

Se realizó un encuentro para validar los requisitos del mercado de datos del Indicador Medio Ambiente y así verificar la aceptación de la solución, en el cual estuvo presente la compañera Elena Leonila Fernández García: representante de la ONE en la Universidad de las Ciencias en Informáticas, conjuntamente con Asnioby Hernández López: dirigente de la línea de almacenes de datos de DATEC que se incorporó en la revisión, los mismos observaron y analizaron detalladamente el diseño presentado, quedando satisfechos con el resultado obtenido del trabajo propuesto.

### **3.3.3 Lista de chequeo de diseño**

Las listas de chequeo son importantes para lograr un buen diseño de la solución. Este tiene como objetivo principal que los artefactos elaborados centralizados en el Indicador ambiental sean informativos y cumplan con un formato organizado. La lista de chequeo de diseño al igual que las mencionadas anteriormente se almacenó en DATEC para así dejar constancia el trabajo realizado. A continuación se nombrará la que se generó en esta fase:

- Lista de chequeo Modelo de datos.

La lista de chequeo fue observada por algunos analistas, obteniendo la máxima calificación: B. La misma cumple con las expectativas recomendadas, no tienes errores ortográficos y la información está muy organizada. Este artefacto proporciona una información importante para las personas por eso es tan imprescindible su revisión a través de la lista de chequeo.

### **3.3.4 Pruebas de implantación**

Para verificar la calidad del proceso de implementación se realizó un modelo de casos de prueba que facilita la identificación de los resultados eficientes.

<b>Casos de Prueba</b>	<b>Pre condición</b>	<b>Pos condición</b>	<b>Resultado esperado</b>
------------------------	----------------------	----------------------	---------------------------

Crear base de datos	Instalado el gestor de base de datos PostgreSQL 8.4	Creada la base de datos.	La base de datos creada sobre el gestor de base de datos.
Roles y permisos	Base de datos creada, e instalada la herramienta de administración.	Creada la base de datos con sus correspondientes roles y permisos.	Creado los roles y sus respectivos permisos.
Cargar los Nomencladores	Base de datos creada con sus roles y permisos.	Base de datos lista para su utilización.	Campos llenos de la base de datos.

**Tabla 2:** Pruebas de implantación

## Conclusiones

El presente capítulo da la terminación de una jornada de trabajo y el cumplimiento del último objetivo específico planteando. Para la implementación, se realizó el modelo físico a partir del modelo lógico que se efectuó en la fase de diseño. Primeramente se cargó el script del modelo dimensional, partiendo de este, se realizó la carga de cada uno de los nomencladores para obtener el DML (Lenguaje de manipulación de datos), luego se establecieron los usuarios de la base de datos con sus correspondiente privilegios con el objetivo de crear el DCL (Lenguaje de acceso de datos) y para obtener la estructura de la base de datos se generó el DDL (Lenguaje de definición de datos) así de esta forma, se tiene una visión general de cómo están estructurados los datos del Indicador Medio Ambiente mostrándolo a través de las tablas, secuencias y los índices. Para la validación, se efectuaron las pruebas a los requisitos propuestos por el cliente de la ONE y el diseño presentado cumplió con todas las necesidades de la especialista. Además se realizaron los casos de pruebas de implantación y las listas de chequeo de los artefactos del análisis y diseño obteniendo una alta evaluación en la revisión de los mismos. De esta manera la sección manifestó los pasos detallados de la implementación y prueba de los Indicadores ambientales, obteniendo como resultado: el mercado de datos estadístico ambiental para la Oficina Nacional de Estadística.

## **CONCLUSIONES GENERALES**

La realización de este Trabajo Diploma fue de gran aporte para los autores por los conocimientos y resultados obtenidos que fueron los esperados. Después de desarrollar cada aspecto de la investigación se obtuvieron las siguientes conclusiones:

- Se realizó un estudio del estado del arte sobre los sistemas de bases de datos similares, vinculados a los procesos de la gestión de los datos estadísticos de los Indicadores ambientales, donde se definieron cuáles aspectos eran los menos apropiados para aplicar a la solución definida en el trabajo. Para el desarrollo del mercado de datos con la calidad y rapidez requerida, se utilizó la metodología de Kimball y las herramientas del software libre PostgreSQL 8.4, PgAdmin3 1.10 y Visual Paradigm 6.4.
- Se observaron cada una de las series estadísticas correspondientes al Indicador Medio Ambiente para extraer los principales temas a analizar y luego utilizarlos en el diseño. Se pudo concluir que fue una tarea trabajosa pero no imposible porque marcó el paso más importante para la elaboración del modelo dimensional. Luego de obtener la información de la fase de análisis se realizó eficientemente el diseño del Indicador Medio Ambiente.
- El modelo lógico elaborado se transformó al modelo físico, se realizó la carga de cada uno de los nomencladores obteniendo una adecuada implementación de la base de datos y una organización de la información de los Indicadores del Medio Ambiente.
- Se realizaron las pruebas a la solución desarrollada, en la cual se observó que el mercado de datos da respuesta a todos los requisitos propuestos por el cliente, además se obtuvo una alta calificación en las listas de chequeos realizadas en el análisis y diseño.

El trabajo dio cumplimiento al objetivo general expuesto en el presente trabajo, obteniendo satisfactoriamente el mercado de datos del Indicador Medio Ambiente que será de mucha ayuda para el almacén de datos de la Oficina Nacional de Estadística ya que contribuye a una excelente gestión y control de la información relacionada con las estadísticas ambientales.

## **RECOMENDACIONES**

Durante el avance de los objetivos trazados en el trabajo, surgieron un grupo de ideas que se recomienda para el desarrollo futuro del sistema.

- Realizar el proceso de extracción, transformación y la carga de los datos ambientales.
- Efectuar la capa de visualización para representar las estadísticas de los Indicadores del Medio Ambiente.
- Extender el uso de la aplicación a otras instituciones que recopilen y organicen los datos estadísticos de los Indicadores ambientales.

## REFERENCIA BIBLIOGRÁFICA

1. **An. Luis Walter Reynoso Socio Fundador y Aldo Marcelo Socio Gerente. 2005.** SQL max connections. *SQL max connections*. [En línea] 2001. [Citado el: 3 de febrero de 2010.] <http://www.sqlmax.com/dataw1.asp>.
2. **María Evelia Casales Cabrera. 2009.** Maestría en Ciencias e Ingeniería de Computación. [En línea] 2009. [Citado el: 15 de febrero de 2010.] <http://hp.fciencias.unam.mx/~alg/bd/dwh.pdf> .
3. **Ana Buigues. 2010.** Temas etiquetados como: 'Data Warehouse' . *Data Warehouse y Data Warehousing*. [En línea] 2010. [Citado el: 14 de febrero de 2010.] <http://anabuigues.com/tag/data-warehouse/>.
4. **Juan Luis Martínez Albuerne. 2008** . Trabajo fin de carrera . *Diseño del catálogo de metadatos para la automatización del proceso de carga de un data webhouse* . [En línea] 2008 . [Citado el: 2 de febrero de 2010.] [http://oa.upm.es/1024/1/PFC\\_JUAN\\_LUIS\\_MARTINEZ\\_ALBURNE.pdf](http://oa.upm.es/1024/1/PFC_JUAN_LUIS_MARTINEZ_ALBURNE.pdf) .
5. **Leandro. 2009.** Lea en binario. *Preparación para los datos*. [En línea] 2009. [Citado el: 3 de febrero de 2010.] <http://leaenbinario.blogspot.com/2009/10/kdd-preparacion-de-los-datos.html> .
6. **María de los Ángeles Ibarra. 2005** . [En línea] 2005 . [Citado el: 17 de febrero de 2010.] <http://exa.unne.edu.ar/depar/Áreas/informatica/SistemasOperativos/OLAP.pdf> .
7. **Paqui Valle Pérez. 2009** . mailxmail . *Informática en la administración pública*. [En línea] 2009 . [Citado el: 3 de febrero de 2010.] <http://www.mailxmail.com/curso-informatica-administracion-publica-3/bases-datos-concepto-caracteristicas-funcionalidades>.
8. **Guillermo Espinosa. 2008.** C.F.T SOEDUC ACONCAGUA. *C.F.T SOEDUC ACONCAGUA*. [En línea] 2008. [Citado el: 6 de febrero de 2010.] [www.soeduc.cl/apuntes/basededatos.doc](http://www.soeduc.cl/apuntes/basededatos.doc).
9. **CEPALSTAT. 2009.** CEPAL. *Comisión económica para américa latina y el caribe*. [En línea] 2009. [Citado el: 5 de febrero de 2010.] <http://www.eclac.org/estadisticas/>.
10. **Editorial McGraw-Hill. 2008.** Mailxmail. *Mailxmail*. [En línea] 2008. [Citado el: 7 de febrero de 2010.] <http://www.mailxmail.com/curso-sistemas-bases-datos/sgbd-sistemas-gestores-bases-datos-introduccion>.
11. **Dr. Marco Antonio Cruz-Chávez. 2009** . MySQL. [En línea] 2009 . [Citado el: 16 de febrero de 2010.] <http://www.uaem.mx/posgrado/mcruz/cursos/miic/MySQL.pdf> .

12. **Franco Catrin. 2009.** Fayer Wayer. *Fayer Wayer*. [En línea] 2009. [Citado el: 9 de febrero de 2010.] <http://www.fayerwayer.com/2009/07/postgresql-84-mantiene-distancia-con-sus-rivales/>.
13. **Marco A Cruz-Chávez. 2009** . Universidad Autónoma del Estado de Morelos . Centro de Investigación en Ingeniería y Ciencias Aplicadas . [En línea] 2009 . [Citado el: 6 de febrero de 2010.] <http://www.uaem.mx/posgrado/mcruz/cursos/miic/oracle.pdf> .
14. **Carlos D. González . 2010** . Curso de base de datos PostgreSQL, SQL avanzado y PHP. Curso de base de datos PostgreSQL, SQL avanzado y PHP. [En línea] 2010 . [Citado el: 8 de febrero de 2010.] <http://www.usabilidadweb.com.ar/postgre.php> .
15. **Ralfm . 2007** . www.linux-es.org. Introducción a PostgreSQL- Instalación e inicialización . [En línea] 2007 . [Citado el: 7 de febrero de 2010.] <http://www.linux-es.org/node/536> .
16. **Ubuntu. 2008.** guia-ubuntu. PgAdmin\_III. [En línea] 2008. [Citado el: 9 de febrero de 2010.] [http://www.guia-ubuntu.org/index.php?title=PgAdmin\\_III](http://www.guia-ubuntu.org/index.php?title=PgAdmin_III).
17. **GUI . 2009** . Guía de la Comunidad para las herramientas GUI de PostgreSQL . PgAdminIII. [En línea] 2009 . [Citado el: 9 de febrero de 2010.] [http://wiki.postgresql.org/wiki/Gu%C3%ADa\\_de\\_la\\_Comunidad\\_para\\_las\\_herramientas\\_GUI\\_de\\_PostgreSQL](http://wiki.postgresql.org/wiki/Gu%C3%ADa_de_la_Comunidad_para_las_herramientas_GUI_de_PostgreSQL).
18. **Ana Mercedes Caceres. 2006.** Facultad de ingeniería escuela de computación análisis y diseño de sistemas ciclo II/06 . *Herramienta CASE* . [En línea] 2006. [Citado el: 7 de febrero de 2010.] <http://www.galeon.com/rcruz0423/docs/case.pdf> .
19. **Equipo Danysoft. 2006** . Danysoft . *Modelado de bases de datos*. [En línea] 2006 . [Citado el: 10 de febrero de 2010.] <http://info.danysoft.com/free/model2.pdf>.
20. **Embarcadero T. 2009** . Embarcadero Technologies. *ER/Studio*. [En línea] 2009 . [Citado el: 11 de febrero de 2010.] [http://www.soluciones-ag.com/pdf\\_productos/Spanish\\_ER-Studio\\_Datasheet\\_2009.pdf](http://www.soluciones-ag.com/pdf_productos/Spanish_ER-Studio_Datasheet_2009.pdf) .
21. **Descargas de software. 2007** . Sitios de descargas de software . *Visual Paradigm*. [En línea] 2007 . [Citado el: 10 de febrero de 2010.] [http://www.freedownloadmanager.org/es/downloads/Paradigma\\_Visual\\_para\\_UML\\_%28M%C3%8D%29\\_14720\\_p/](http://www.freedownloadmanager.org/es/downloads/Paradigma_Visual_para_UML_%28M%C3%8D%29_14720_p/).

22. **Blog de Eduardo León. 2007** . *Visual Paradigm*. [En línea] 2007 . [Citado el: 12 de febrero de 2010.] <http://slion2000.blogspot.com/2007/04/visual-paradigm-una-herramienta-de-lo.html> .
23. **Luis Giraldo y Yuliana Zapata. 2005**. Herramientas de desarrollo de ingeniería de SW para Linux. *Visual Paradigm*. [En línea] 2005. [Citado el: 10 de febrero de 2010.] [http://hugolopez.phi.com.co/docs/download/file=Giraldo-Zapata-Herramientas%20de%20ISW.pdf,\\_id=17](http://hugolopez.phi.com.co/docs/download/file=Giraldo-Zapata-Herramientas%20de%20ISW.pdf,_id=17).
24. **Curto. 2008**. J. CIF vs MD. *Dos enfoques clásicos en el diseño de la arquitectura de una Almacén de datos*. [En línea] 2008. [Citado el: 10 de febrero de 2010.] <http://bi-businessintelligence.blogspot.com/2009/01/cif-vs-md-dos-enfoques-clasicos-en-el.html>.
25. **Carlos Minay. 2009** . Scribd . *Modelos de bases de datos* . [En línea] 2009 . [Citado el: 12 de febrero de 2010.] <http://www.scribd.com/doc/17170125/Modelos-de-Bases-de-Datos> .
26. **Guillermo, Storti. 2007**. Entidad Relación. [En línea] 2007. [Citado el: 8 de febrero de 2010.] [http://www.belgrano.esc.edu.ar/matestudio/carpeta\\_de\\_access\\_introduccion.pdf](http://www.belgrano.esc.edu.ar/matestudio/carpeta_de_access_introduccion.pdf).
27. **Sara Álvarez. 2007**. Desarrollo Web. [En línea] 2007. [Citado el: 9 de febrero de 2010.] <http://www.desarrolloweb.com/articulos/modelo-entidad-relacion.html> .
28. **Hobbs, Lilian. 2005**. *Oracle Database 10g Data Warehousing*. EUA : ELSEVIER Digital Press, 2005.
29. **Hazbleydi C. Verástegui. 2007** . Modelos Dimensionales. [En línea] 2007 . [Citado el: 10 de febrero de 2010.] [http://www.dbsystem.com/pls/portal/docs/PAGE/SITIOWWWDB/ARTICULOS/MODELADO%20DIMENSIONAL%20DE%20DATOS\\_V2.PDF](http://www.dbsystem.com/pls/portal/docs/PAGE/SITIOWWWDB/ARTICULOS/MODELADO%20DIMENSIONAL%20DE%20DATOS_V2.PDF)
30. **ONE. (1958-2008)**. *Oficina Nacional de Estadísticas*. [En línea] 1958-2008. [Citado el: 22 de marzo de 2010] <http://www.one.cu/publicaciones/50aniversario/medio%20ambiente/public%20completa.pdf>
31. **JACOBSON, Ivar y BOOCH, Grady, RUMBAUGH, James. 2000**. *El Proceso Unificado de Desarrollo de Software*. s.l. : Addison Wesley, 2000. Capítulos 7, 8 páginas 125-163, 187-202.
32. **Mafla. 2005**. *Especificaciones Generales Requeridas Para La Implementación Del Sistema De Seguridad y Alta Disponibilidad*. Ecuador: : Aduana del Ecuador., 2005.

33. **Josep Curto. 2007** . Information Management Data Warehousing . *Data Warehouse y Datamart*. [En línea] 2007. [Citado el: 1 de febrero de 2010.]  
<http://informationmanagement.wordpress.com/2007/10/07/data-warehousing-data-warehouse-y-datamart/>
34. **Definición . 2008** . Modelo de datos . [En línea] 2008 . [Citado el: 2 de mayo de 2010.]  
<http://definicion.de/modelo-de-datos/> .
35. **Gabriel Requelme, Anabel Ruíz, David Torres. 2008** . Slideshare . *Modelos de Datos y Modelado Conceptual*. [En línea] 2008 . [Citado el: 2 de mayo de 2010.]

## BIBLIOGRAFÍA

1. Comisión económica para América Latina y el Caribe. Disponible en: <http://www.eclac.org/>
2. Europa. Disponible en: <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>
3. INE de España. Disponible en: <http://www.ine.es>
4. Oficina Nacional de Estadísticas. Disponible en: <http://www.one.cu>
5. Naciones Unidas. Disponible en: <http://data.un.org/>
6. Entorno Virtual de Aprendizaje. Disponible en: <http://eva.uci.cu/>
7. Biblioteca. Disponible en: <http://biblioteca.uci.cu/>
8. Roberto Hernández, Carlos y Pilar. 1998. Metodología de la investigación. Disponible en: <http://bibliodoc.uci.cu/pdf/9701018990.pdf>
9. Roberto Solana, Rubén García, Raúl Herranz. 2008. Erwin data modeler. Disponible en: <http://74.125.93.132/search?q=cache:etfNYOVFE7gJ:kybele.escet.urjc.es/documentos/HC/Exposiciones/Erwin.pdf+herramienta+Erwin&cd=5&hl=es&ct=clnk&gl=cu>
10. gopacBI. 2007. Inteligencia del Negocio. Disponible en: <http://www.gopac.com.mx/v3/gopacbi/quees.asp>.
11. RUMBAUGH, James, JACOBSON, Ivar y BOOCH, Grady. 2000. *El lenguaje unificado de modelado*. s.l. : Addison Wesley, 2000.
12. Liudmila Padrón Torres. 2009. Almacenes de datos. Disponible en: <http://www.zonaeconomica.com/informatica/almacenesdedatos>
13. Respinosa Milla. 2010. Kimball vs Inmon. Disponible en: <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimensional/>
14. Software Greenhouse. 2010. Data Mart. Disponible en: <http://www.swgreenhouse.com/Productos/Hi-Spins/DataMart.html>
15. Rodríguez, Alexei Domínguez. 2009. Diseño e implementación de la Base de Datos del Sistema de planificación del Entrenamiento deportivo de Judo Femenino. La Habana : s.n., 2009.

## GLOSARIO DE TÉRMINOS

**ANSI** o Instituto Americano Nacional de Estándares: Es una organización que supervisa el desarrollo de estándares para productos, servicios, procesos y sistemas en los Estados Unidos. Ha creado diversos estándares, entre los que se puede citar ASCII.

**Atributo:** Es cada una de las cualidades, propiedades o características de un elemento.

**CPU:** Es el centro del ordenador y sin ella no podría funcionar. También es el conjunto de las piezas que forman la caja del ordenador.

**CITMA:** Ministerio de Ciencia, Tecnología y Medio Ambiente.

**Escalabilidad:** La capacidad del sistema informático de cambiar su tamaño o configuración para adaptarse a las circunstancias cambiantes.

**FN:** En la teoría de bases de datos relacionales, las formas normales (FN) proporcionan los criterios para determinar el grado de vulnerabilidad de una tabla a inconsistencias y anomalías lógicas. Mientras sea más alta la forma normal aplicable a una tabla, es menos vulnerable a inconsistencias y anomalías.

**GPL:** Licencia Pública General de GNU o más conocida por su acrónimo del inglés GNU GPL, es una licencia creada por la Fundación del Software Libre y está orientada principalmente a proteger la libre distribución, modificación y uso de software.

**Microsoft** (acrónimo de Microcomputer Software): Es una empresa de Estados Unidos, fundada por Bill Gates y Paul Allen. Dueña y productora de los sistemas operativos: Microsoft DOS y Microsoft Windows, que se utilizan en la mayoría de las computadoras del planeta.

**Open Source** (código abierto): Práctica de desarrollo de software que promueve el acceso al código fuente de los sistemas computacionales.

**PC** o computadora personal: Se basa en un microprocesador, diseñada para ser utilizada por una sola persona a la vez y equipada para cumplir tareas comunes.

**PHP** (Pre-procesador de Hipertexto): Es un lenguaje de programación interpretado, diseñado originalmente para la creación de páginas web dinámicas.

**Rol:** Papel desempeñado por las personas en la sociedad.

**Rollback:** Es el proceso automático que se ejecuta durante una transacción, si ocurre un error la base de datos regresa al estado en que se encontraba desde la última transacción.