

**Universidad de las Ciencias Informáticas**

**Facultad 6**



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

**“Sistema de Información de Gobierno. Mercado de datos  
Inversiones”**

**Autoras**

Yelaine Hernández Rodríguez  
Enelys Pérez Hernández

**Tutoras**

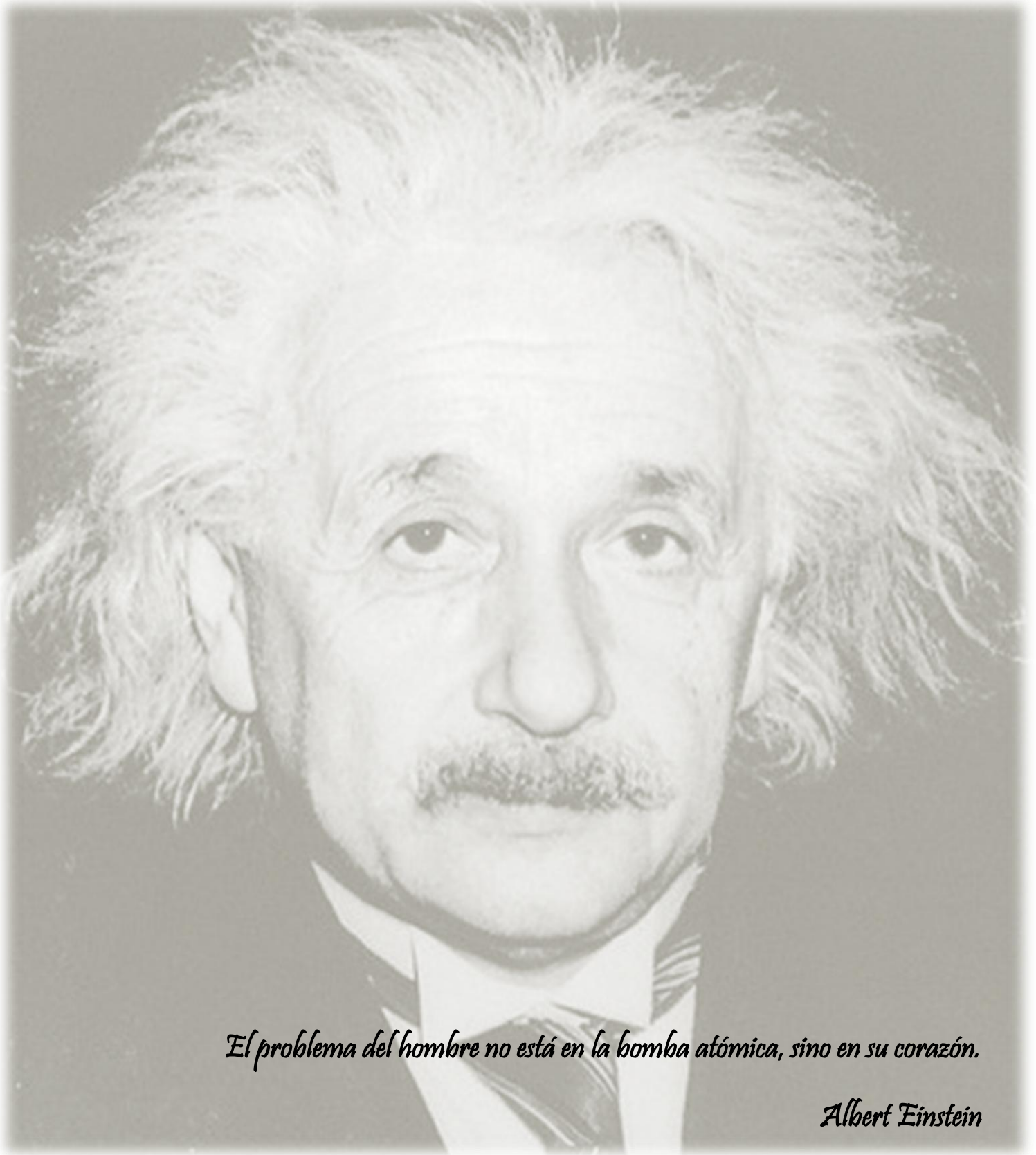
Ing. Yetel Castañeda Ramírez  
Ing. Yanelis Ramírez Hernández

**Co-Tutora**

Lic. Yanelis Benítez Fernández

**Junio, 2011**

**La Habana. Cuba**



*El problema del hombre no está en la bomba atómica, sino en su corazón.*

*Albert Einstein*

## DECLARACIÓN DE AUTORÍA

Declaramos ser las únicas autoras de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos de la misma, con carácter exclusivo.

Para que así conste firmamos la presente a los \_\_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

---

Enelys Pérez Hernández

Autora

---

Yelaine Hernández Rodríguez

Autora

---

Ing. Yetel Castañeda Ramírez

Tutora

---

Ing. Yanelis Ramírez Hernández

Tutora

---

Lic. Yanelis Benítez Fernández

Co-Tutora

## DATOS DE CONTACTO

**Tutora:** Ing. Yanelis Ramírez Hernández

**Especialidad de graduación:** Ingeniería en Ciencias Informáticas

**Categoría docente:** Instructor

**Categoría Científica:** -

**Años de experiencia en el tema:** 3

**Años de graduado:** 3

**Correo Electrónico:** yramirezh@uci.cu

**Tutora:** Ing. Yetel Castañeda Ramírez

**Especialidad de graduación:** Ingeniería en Ciencias Informáticas

**Categoría docente:** Instructor

**Categoría Científica:** -

**Años de experiencia en el tema:** 3

**Años de graduado:** 3

**Correo Electrónico:** ycastaneda@uci.cu

**Co-Tutora:** Lic. Yanelis Benítez Fernández

**Especialidad de graduación:** Lic. Mat. y Comp

**Categoría docente:** Asistente

**Categoría Científica:-**

**Años de experiencia en el tema:-**

**Años de graduado:** 10

**Correo Electrónico:** jhanelis@uci.cu

## AGRADECIMIENTOS

*A toda mi familia porque son mi mayor tesoro y sostén.*

*A mi mami por estar pendiente de cada paso que doy, por su cariño incondicional, su apoyo y dedicación.*

*A mi papi Vichi por llenarme de alegría incluso en los peores momentos, por su paciencia, ternura y porque siempre confió en mí.*

*A mi abuelita Eneida por hacerme sentir el centro de su mundo, por sus consejos y su amor.*

*A mi hermanita por hacer que no pierda nunca el alma de niña.*

*A mi padre por brindarme su apoyo y cariño desde la distancia.*

*A los amigos de toda la vida, a los que están lejos y a los que en poco tiempo han logrado colarse en mi corazón. A los que compartieron conmigo su comida, apartamento, alegrías y sentimientos. A todos ellos porque la realización de este trabajo se habría quedado en un sueño sin su ayuda.*

*A Yahi por haber estado ahí estos cinco años en los buenos y malos momentos, por su cariño y lealtad. Por aguantar mis malcriadeces y ayudarme a crecer como persona.*

*A mi compañera de tesis por aguantarme desde el primer día en la universidad y después de tanto tiempo compartir este triunfo conmigo.*

*A los muchachos y muchachas del laboratorio que siempre estuvieron ahí para aclarar mis dudas y soportar tantas molestias a cambio de una sonrisa.*

*A las dos Yanelis y a Salvador porque garantizaron la calidad de este trabajo.*

*En fin, a todos los que contribuyeron a mi formación como profesional, a los que hicieron que el camino fuera más fácil y siempre confiaron en que podía llegar.*

***Enelys***

*A mi mami y papi por darme la vida y enseñarme a vivirla.*

*A mi hermana por ser la mejor y darme a la sobrina más linda del mundo.*

*A Amauri por apoyarme y entenderme en los momentos difíciles.*

*A mi abuelito lindo por existir y siempre tenerme en su corazón.*

*A mis amigas más íntimas: Anllelis, Jenifer y mi morita Disleiny.*

*A Yulie por ser mi prima y amiga.*

*A mi dúo de tesis por soportarme, por haber sido mi enfermera cuando lo necesité y por ser mi amiga.*

*A quienes me enseñaron a estudiar y a ser independiente: Yanelis Benítez, Rolando y Luillis.*

*A mi tutora Yanelis Ramírez y a Salvador por todas las críticas constructivas.*

*En fin, a todos aquellos que me ayudaron poniendo su granito de arena en lo que hoy es mi triunfo.*

*Yelaine*

## DEDICATORIA

*A mi hermana, mami y papi.*

*A mi abuelito Ramón y a mis abuelas.*

*A mi sobrina, primos y primas para de ellos ser ejemplo.*

*A Amauri por apoyar mi comienzo y hoy después de tanto tiempo estar acompañándome.*

*En fin, a toda mi familia, amigas y amigos que siempre confiaron que podría lograrlo y me ayudaron a conseguirlo.*

*Yelaine*

*A mi mami que ha sido mi guía toda la vida.*

*A mi tata que es la hermana más linda del mundo.*

*A mi abuelita Eneida por hacer de mí la niña de sus ojos.*

*A la memoria de mi abuela Viviana por ser mi ejemplo.*

*A mi papi Vichi que es tan dueño de este triunfo como yo.*

*Enelys*

## **RESUMEN**

La presente investigación surge a partir de la necesidad de la Oficina Nacional de Estadísticas de actualizar el sistema de almacenamiento y gestión de la información que se utiliza en dicho centro. En esta entidad, la información almacenada ha ido incrementando con el paso del tiempo, esto trae consigo la necesidad de convertirla en conocimiento útil que contribuya a la toma de decisiones en el área de Inversiones del almacén de datos Sistema de Información de Gobierno, lo cual es el principal punto de enfoque en la presente investigación. Se desarrolló como solución un mercado de datos, el cual gestiona toda la información referente al sector de inversiones en el país. Para su desarrollo se mencionan las diferentes herramientas informáticas y metodologías a utilizar. Se refinó el análisis y diseño, se realizó la implementación del mercado de datos Inversiones así como las pruebas correspondientes con el objetivo de garantizar la calidad del producto y la aceptación del cliente.

**PALABRAS CLAVES:** Almacén de Datos, Herramientas, Inversiones, Mercado de Datos, Metodologías, Oficina Nacional de Estadísticas.



# ÍNDICE

INTRODUCCIÓN .....	1
CAPÍTULO I: FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN .....	5
Introducción .....	5
1.1 Tendencias en el mundo y en Cuba de sistemas que controlan las estadísticas de inversiones.....	5
1.2 Definiciones de Almacén de Datos y Mercado de Datos.....	6
1.2.1 Ventajas y desventajas de los Mercados de Datos.....	8
1.2.2 Modelo dimensional.....	9
1.3 Integración de datos .....	11
1.4 Inteligencia de negocios.....	12
1.4.1 Procesos OLTP y OLAP .....	13
1.5 Metodologías de desarrollo .....	14
1.6 Herramientas de desarrollo .....	17
1.6.1 Herramienta Case de Modelado.....	18
1.6.2 Sistema Gestor de Bases de Datos.....	18
1.6.3 Herramientas para la integración de datos.....	21
1.6.4 Herramientas para la Inteligencia de Negocios .....	22
1.7 Conclusiones .....	23
CAPÍTULO II: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS INVERSIONES .....	24
Introducción .....	24
2.1 Estudio preliminar del negocio.....	24
2.2 Necesidades de información .....	25
2.3 Requisitos de información.....	26
2.4 Reglas de Negocio.....	27
2.5 Requisitos funcionales.....	27
2.6 Requisitos no funcionales .....	28
2.7 Modelo de Casos de Uso del Sistema.....	29
2.7.1 Actores del sistema .....	29
2.7.2 Diagrama de Casos de Uso del Sistema .....	30
2.7.3 Casos de Uso de Información.....	30
2.7.4 Casos de Usos Funcionales.....	32
2.8 Identificación de dimensiones, hechos y medidas.....	35
2.8.1 Dimensiones .....	35
2.8.2 Hechos.....	37

2.8.3 Medidas .....	38
2.9 Desarrollo de la matriz BUS .....	38
2.10 Modelo de datos.....	39
2.11 Diseño de los procesos de integración de datos .....	39
2.12 Mapa de navegación .....	41
2.13 Políticas de respaldo y recuperación.....	42
2.14 Conclusiones .....	42
CAPÍTULO III: IMPLEMENTACIÓN DEL MERCADO DE DATOS INVERSIONES.....	43
Introducción .....	43
3.1 Usuarios y privilegios.....	43
3.2 Implementación del subsistema de integración de datos .....	45
3.2.1 Implementación del modelo de datos .....	45
3.2.2 Implementación de los flujos de transformación.....	46
3.2.3 Implementación del trabajo .....	49
3.3 Implementación del subsistema de visualización de datos .....	50
3.3.1 Cubos OLAP.....	50
3.3.2 Reportes candidatos.....	51
3.4 Conclusiones .....	53
CAPÍTULO IV: VALIDACIÓN DEL MERCADO DE DATOS INVERSIONES.....	54
Introducción .....	54
4.1 Calidad de software .....	54
4.2 Pruebas de calidad de software.....	55
4.2.1 Modelo V .....	55
4.2.2 Tipos de pruebas .....	56
4.2.3 Casos de prueba.....	57
4.2.4 Listas de chequeo.....	58
4.3 Resultado de las pruebas realizadas.....	59
4.4 Conclusiones .....	60
CONCLUSIONES GENERALES.....	61
RECOMENDACIONES.....	62
REFERENCIAS BIBLIOGRÁFICAS .....	63
BIBLIOGRAFÍA .....	65
ANEXOS .....	67
GLOSARIO DE TÉRMINOS .....	70

### **INTRODUCCIÓN**

La informática ha traído consigo un gran desarrollo tecnológico en el tratamiento automático de la información, donde la mayoría de los países han encontrado soluciones rápidas y rentables a sus problemas. Uno de ellos es el manejo de grandes volúmenes de información, donde esta ciencia se ha convertido en el principal apoyo para su control como por ejemplo, en las estadísticas.

La estadística es el lenguaje que permite comunicar información basada en datos cuantitativos, es de vital importancia para el buen control de los recursos y poder tomar decisiones inteligentes y acertadas en condiciones de incertidumbre. Abarca la recolección, presentación y caracterización de la información para ayudar tanto en el análisis como en la interpretación de los datos.

Mantener estadísticas precisas, actualizadas y comparables a nivel nacional, es significativo para los gobiernos, las instituciones públicas y el sector privado. Por esto, las Naciones Unidas se encargan de coordinar las actividades en materia de estadísticas y han establecido un órgano intergubernamental especializado en la materia: la División de Estadística de la Secretaría de la Organización de Naciones Unidas. La misma proporciona una amplia variedad de estadísticas y servicios para computadoras y usuarios de datos estadísticos en todo el mundo. En Cuba también existen instituciones encargadas de llevar el manejo de la información estadística.

La Oficina Nacional de Estadísticas (ONE), es el órgano rector de la estadística en Cuba, su objetivo es captar, analizar y difundir los datos recogidos a lo largo y ancho de todo el país para realizar una mejor toma de decisiones y medir los indicadores de interés. Tal organismo cuenta con una serie de modelos en los cuales recoge toda la información de los sectores de la economía y la sociedad. Los datos almacenados se encuentran en formatos de difícil acceso para su consulta y difusión por lo que esta entidad ha solicitado los servicios a la Universidad de las Ciencias Informáticas (UCI) para automatizar los procesos que allí se realizan.

La UCI es un centro formador de profesionales en la rama de Informática, donde se vincula el estudio con la producción y la investigación ligado a las necesidades sociales del país y de otros pueblos hermanos; abriéndose paso en la industria de software, favoreciendo la economía del país y su propia economía. Actualmente cuenta con más de 30 centros productivos que se destacan por sus resultados en las esferas de salud, educación, software libre, sistemas legales, realidad virtual, automatización, procesamiento de imágenes y señales, entre otras. Entre sus centros productivos se encuentra el Centro de Tecnologías de Gestión de Datos (DATEC). El mismo cuenta con cuatro líneas de trabajo fundamentales: Integración de

soluciones, PostgreSQL, Bioinformática y Almacenes de datos; esta última se encuentra desarrollando el almacén de datos llamado “Sistema de Información de Gobierno” (SIGOB).

Para un mejor control estadístico la ONE se encuentra dividida en áreas y una de ellas es Inversiones, en la cual se recoge la información de todas las actividades que se realizan de inversiones en los centros informantes (organismos, empresas, unidad presupuestada y otras unidades) existentes en las provincias de todo el país, ya sea del plan de ejecución física de las mismas como de los gastos en actividades de protección del medio ambiente.

Existen algunas deficiencias en la gestión de los datos, aunque las estadísticas recogidas en esta área se tienen en formato Excel, los archivos solo pueden ser consultados por especialistas de la información. Los mismos suelen ser a menudo especialistas de la informática, con alto conocimiento del negocio ya que algunas de las fuentes están codificadas y se necesita ser conocedor de la terminología usada para comprender su contenido. También se generan ficheros anuales con los cuales se hace muy difícil la obtención rápida de información, la cual va acumulándose año tras año y es cada vez más complicado analizarla. Además, los datos no están integrados y atenta contra la calidad de estos, lo que implica que pueden existir referencias a la misma información que usan diferente codificación, o diferente cantidad de caracteres, esto ocurre porque existen múltiples versiones de los mismos datos. Por otro lado actualmente solo existen aplicaciones que procesan información, algunas son muy antiguas y para usarlas hay que simular el entorno MS-DOS, estas son las que pueden exportar los datos al formato DBF.

Por los elementos antes expuestos surge el siguiente **problema a resolver**: ¿Cómo contribuir a la toma de decisiones en el área de Inversiones del Sistema de Información de Gobierno?

La investigación tiene como **objeto de estudio** los almacenes de datos enmarcado en el **campo de acción** mercado de datos para el área de Inversiones del Sistema de Información de Gobierno.

Para dar solución a la problemática antes mencionada se define como **objetivo general**: desarrollar el mercado de datos Inversiones del Sistema de Información de Gobierno que contribuya a la toma de decisiones, el cual se desglosa en los siguientes **objetivos específicos**:

- Refinar el análisis y diseño del mercado de datos del área Inversiones.
- Implementar el mercado de datos del área Inversiones.
- Validar el mercado de datos del área Inversiones.

Para lograr los objetivos específicos planteados se definieron las siguientes **tareas de la investigación**:

- Caracterización de las metodologías, herramientas y tecnologías a utilizar en el desarrollo del Mercado de Datos.
- Refinamiento de requisitos.
- Refinamiento de la descripción de los casos de uso del mercado de datos.
- Refinamiento de los hechos, las medidas y las dimensiones del mercado de datos.
- Refinamiento del modelo de datos.
- Refinamiento de la arquitectura del mercado de datos.
- Diseño del subsistema de integración.
- Diseño del subsistema de visualización.
- Diseño de los casos de pruebas.
- Implementación del subsistema de integración.
- Implementación del subsistema de visualización.
- Aplicación de las listas de chequeo.
- Aplicación de los casos de pruebas.

Estructura capitular:

### **Capítulo I:** Fundamentación teórica de la investigación

En el presente capítulo se abordan las tendencias en el mundo de sistemas que controlan las estadísticas de inversiones, así como las tecnologías de almacenamiento y sus características principales. También se explican los tipos de modelos dimensionales que existen y algunos conceptos relacionados al tema, necesarios para entender el modelo multidimensional utilizado. Del mismo modo se tratan conceptos necesarios para comprender la integración de datos y la inteligencia de negocio. Además se explica la metodología seleccionada por el centro y se documentan las herramientas de desarrollo a utilizar.

### **Capítulo II:** Análisis y diseño del mercado de datos Inversiones

Para el desarrollo del Mercado de Datos Inversiones es de vital importancia realizar un buen estudio preliminar del negocio para captar las necesidades del cliente y que se logre un buen desarrollo de toma de decisiones. En este capítulo se documenta todo lo referente al estudio del negocio, así como también se exponen las necesidades de información, las reglas del negocio, el levantamiento de los requisitos, la descripción de los casos de uso, se identifican las dimensiones, hechos y medidas con que cuenta el negocio, se desarrolla la matriz BUS. Se realiza el diseño de los procesos de integración de datos y se diseña el mapa de navegación. Estos datos guiarán posteriormente a los desarrolladores a lograr un producto con la calidad requerida.

### **Capítulo III:** Implementación del mercado de datos Inversiones

En este capítulo se documenta todo lo referente a la implementación del Mercado de Datos. Se definen los usuarios y privilegios de los mismos así como los niveles de acceso al sistema del Mercado de Datos de cada uno de los roles. Se implementa el subsistema de integración de datos, la estructura de los datos, esquemas y tablas de la Base de Datos. Se realiza la implementación del subsistema de visualización de datos donde se diseñan de los cubos OLAP, se implementan los reportes candidatos y se desarrolla la estructura de navegación.

### **Capítulo IV:** Validación del mercado de datos Inversiones

En este capítulo se realizan las pruebas y validaciones al Mercado de Datos Inversiones para garantizar la calidad, eficiencia y obtención de los resultados esperados. Se describen los distintos tipos de pruebas que fueron aplicadas y sus resultados, así como la validación de la solución propuesta.

## **CAPÍTULO I: FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN**

### **Introducción**

En el presente capítulo se abordan las tendencias en el mundo de sistemas que controlan las estadísticas de inversiones, así como las tecnologías de almacenamiento y sus características principales. También se explican los tipos de modelos dimensionales que existen y algunos conceptos relacionados al tema, necesarios para entender y poder optar por un modelo. Del mismo modo se tratan conceptos necesarios para comprender la integración de datos y la inteligencia de negocio. Además se explica la metodología seleccionada por el centro y se documentan las herramientas de desarrollo a utilizar.

### **1.1 Tendencias en el mundo y en Cuba de sistemas que controlan las estadísticas de inversiones**

La estadística es comúnmente considerada como una colección de hechos numéricos expresados en términos de una relación manejable, y que han sido recopilados a partir de otros datos numéricos. Murria R. Spiegel, (1991) dice: "La estadística estudia los métodos científicos para recoger, organizar, resumir y analizar datos, así como para sacar conclusiones válidas y tomar decisiones razonables basadas en tal análisis". Pero las estadísticas suelen ser dejadas como algo que no es útil, aunque por el contrario, ellas reflejan tendencias y permiten tomar decisiones, existen tendencias de sistemas que controlan las estadísticas, a continuación se exponen ejemplos de los mismos:

- En la Secretaría General de la Comunidad Andina (SGCAN) en Lima, Perú se llevan a cabo reuniones de expertos gubernamentales en estadísticas sobre inversión extranjera directa de la comunidad Andina. En las mismas se desarrolla un plan de acción para la armonización de las estadísticas sobre inversión extranjera directa dejando dispuestos los objetivos específicos, resultados esperados, actividades del plan de acción y también un cronograma de plan de acción. **(1)**
- Proyecto de Facilitación del Tratado de Libre Comercio México y la Unión Europea (PROTLCUEM), en él se realiza un estudio comparativo sobre los sistemas de estadísticas de inversión en México y la Unión Europea. El punto esencial es el impacto positivo y negativo de la inversión extranjera sobre el empleo y desarrollo regional en México así como en la transferencia de tecnología. **(2)**
- Sistema Nacional de Inversiones (SNI) de Chile, es un sistema que establece las bases y condiciones sobre las cuales se desarrolla el proceso de inversión pública en el país, para impulsar aquellas iniciativas de inversión más rentables para la sociedad, según los lineamientos de la política de Gobierno. El SNI debe abarcar en su integralidad el proceso de inversión a través del cual las iniciativas van evolucionando desde su identificación hasta que entran en operación. **(3)**

- En Cuba se encuentra la ONE encargada de integrar el trabajo estadístico estatal del país en un sistema eficiente, destinado a brindar la información requerida para el control del plan de la economía, el análisis del comportamiento de la economía y las demás necesidades del país. Referente a las inversiones se muestra una visión general del proceso inversionista del sector civil en el país.

En el sitio oficial de la ONE se ofrece la ejecución física de inversiones por componentes, de la economía nacional. Los datos que se presentan son reportados a través de los modelos del Sistema de Información Estadístico Nacional (SIEN), a salida de empresa, unidad presupuestada y otras unidades contempladas en dicho sistema. El registro económico de los datos sobre las inversiones está establecido en todas las entidades que acometen inversiones, ya sea por concepto de construcción civil y montaje de equipos, como por la adquisición de equipos y suministros con los mismos fines, así como por incurrir en otros gastos. (4)

La estadística de inversiones, cuyos proyectos son previamente aprobados por el órgano de planificación correspondiente, abarca todo el proceso inversionista en su conjunto, aunque por cuestiones de orden práctico, las cifras estadísticas se registran a partir del momento en que se incurre en el primer gasto, independientemente de su pago y con cargo a cada proyecto de inversión definido.

### **1.2 Definiciones de Almacén de Datos y Mercado de Datos**

Con el transcurso de los años y el aumento de la complejidad en las tecnologías también la información que se almacenaba comenzó a aumentar a tal punto que en ocasiones resultaba demasiado complicado analizarla. Con la aparición de los conceptos de Almacén de Datos (AD) y su utilización desde finales de la década de los 80 se abrió una puerta para proporcionar respuesta a esta situación. En la actualidad se puede afirmar que los avances alcanzados confirman que ya es una tecnología madura, estable y soluciona la problemática presentada, lo que no significa que no continúe en constante evolución.

En el área de AD se reconocen como figuras principales a: Bill Inmon y Ralph Kimball. Estos dos científicos, han proporcionado las definiciones y los principios de diseño que la mayoría de los profesionales utilizan hoy en día. Aunque sus guías no sean seguidas exactamente, es común hacer referencia a la definición de AD de Inmon y a las reglas de diseño de Kimball. A continuación se enuncian dos de ellas:

**Bill Inmon** define un AD en términos de las características del repositorio de datos: "Un AD es una colección de datos orientados por temas, integrados, variables en el tiempo y no volátil para el apoyo de la toma de decisiones". (5)



## Capítulo I: Fundamentación teórica de la investigación

---

*Orientado a temas:* sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales.

*Variante en el tiempo:* el tiempo es parte implícita de la información contenida en un AD. La información almacenada en el AD sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el AD se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

*No volátil:* el almacén de información de un AD existe para ser leído, y no modificado. La información es por tanto permanente, significando la actualización del AD la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

*Integrado:* los datos almacenados en el AD deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. (5)

**Ralph Kimball** define un AD como: "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". También fue Kimball quien determinó que un AD no era más que: "la unión de todos los mercados de datos de una entidad". (6)

Teniendo conocimiento de los principales conceptos de AD se puede explicar que un Mercado de Datos (MD) es un pequeño AD diseñado para una unidad de negocio, el cual se centra únicamente en los requerimientos de los usuarios asociados con dicha unidad de negocio, por tanto, estos son más fáciles de entender y navegar. A continuación se anuncia la definición dada por Inmon:

**Inmon** define un MD como: "una colección de temas organizados en áreas para la toma de decisiones necesarias en un departamento. Finanzas tiene su propio, marketing el suyo y ventas el suyo. Cada departamento tiene su propio MD según sus propias necesidades. El diseño del MD se realiza siguiendo una estructura óptima para las necesidades de los usuarios del departamento. El cual contiene solo el mínimo de información histórica y el grado de granularidad necesario para el departamento". (5)

Los MD presentan las siguientes **características**:

- El diseño del MD se realiza siguiendo una estructura consistente para las necesidades de los usuarios.
- Contiene solo el mínimo de información histórica.
- Tiene el grado de granularidad necesario.
- Supone costes adicionales en hardware, software y accesos de red.
- Se centran en los requisitos de los usuarios asociados a un departamento o área de negocio concretos.

- Son sencillos a la hora de utilizarlos y comprender sus datos.

Entonces se resume de los conceptos abordados anteriormente que un MD es creado con el fin de satisfacer las necesidades específicas de un grupo de usuarios, para brindarle a las entidades una mejor vista de la información con la que cuentan y de este modo dar soporte al proceso de toma de decisiones de la dirección.

### **1.2.1 Ventajas y desventajas de los Mercados de Datos**

Los MD aportan facilidad e inmediatez en el manejo de la información. Las **ventajas** más significativas son:

- Menores cantidades de datos implican que se procesan más rápido las cargas de datos como las consultas. Las peticiones pueden acotarse al área o red que sirve esos datos, sin afectar al resto de los usuarios. La aplicación cliente, que pide la consulta es independiente del servidor que la procesa y del servidor de bases de datos que almacenan la información.
- Los costos que implica la construcción de un MD son mucho menores a los de la implementación de un AD.
- El enorme retorno de la inversión es evidencia de la gran ventaja competitiva que acompaña a esta tecnología. La ventaja competitiva se consigue por permitir a los responsables de la toma de decisiones tácticas y estratégicas el acceso a datos que pueden mostrar información que antes era no disponible, desconocida o sin explotar, en clientes, tendencias y demandas.
- Incremento de la productividad de los responsables de la toma de decisiones. La transformación de los datos orientados a las aplicaciones en información orientada a la toma de decisiones permite a dichos responsables realizar un análisis más preciso y consistente.
- Permite analizar los datos del negocio desde la perspectiva de su evolución en el tiempo, predecir tendencias de evolución del negocio, identificar nuevas oportunidades de negocio y tomar decisiones estratégicas y reducir los costos materiales y humanos en la toma de decisiones. (7)

Las **desventajas** más sobresalientes asociadas con el desarrollo y gestión de un MD son las siguientes:

- No permite el manejo de grandes volúmenes de información por lo que muchas veces se debe recurrir a un conjunto de MD para cubrir todas las necesidades de información de la empresa.
- Baja estimación de los recursos necesarios para la carga de datos.
- Problemas ocultos en los sistemas fuente.
- Alto mantenimiento.
- Proyectos de larga duración.

- Complejidad de integración. (7)

## 1.2.2 Modelo dimensional

A diferencia de los clásicos sistemas de bases de datos que presentan sus estructuras diseñadas mediante el modelo Entidad-Relación, los MD se diseñan mediante un modelo dimensional. Poseen la misma información que el diagrama Entidad-Relación pero la organiza de forma diferente para garantizar la velocidad y eficiencia en la recuperación de la misma. Para el modelo dimensional se utilizan diversos esquemas como el esquema estrella, el esquema copo de nieve (snowflake) y el esquema constelación mediante el cual va a ser representado el modelo dimensional de la solución propuesta.

**Esquema constelación:** se puede decir que es el esquema más complejo comparado con los demás, pues contiene múltiples tablas de hechos que comparten tablas de dimensiones que se visualizan como una constelación de hechos. (8)

Para un mejor entendimiento del esquema explicado anteriormente se abordan los conceptos de cubo, dimensión, medida, granularidad y jerarquía por estar fuertemente asociados al modelo dimensional:

- **Cubo:** un modelo de datos multidimensional soporta el manejo de una vasta cantidad de datos empresariales y temporales. De esta forma surge la instancia del modelo multidimensional, también conocido como cubo o hipercubo. (9)

- **Dimensión:** las dimensiones son un concepto esencial de bases de datos multidimensionales. Las mismas son usadas para seleccionar y agregar datos a un cierto nivel deseado de detalle. Cada instancia de una dimensión o valor de una dimensión, corresponde a un nivel particular. Viendo los datos dentro de un cubo se tiene la ventaja de que se puede manejar cualquier número de dimensiones. Sin embargo, usualmente un cubo tiene entre cuatro y 12 dimensiones. Generalmente, un cubo soporta una vista de dos o tres dimensiones simultáneamente. (9)

- **Medida:** la medida o hecho es un dato numérico que representa una actividad específica de un negocio, mientras que una dimensión representa una perspectiva de los datos. Una medida contiene una propiedad numérica y una fórmula. Existen tres clases de medidas:

- **Medidas aditivas:** pueden ser combinadas a lo largo de cualquier dimensión.
- **Medidas semiaditivas:** pueden ser combinadas a lo largo de una o más dimensiones.
- **Medidas no aditivas:** no pueden ser combinadas a lo largo de ninguna dimensión. (9)

- **Granularidad:** es el nivel más bajo de información que será almacenado en la tabla de hechos e indica el grado de detalle asociado a un hecho particular. La granularidad depende directamente del número de dimensiones que se asocian a la tabla de hechos. El primer paso para diseñar una tabla de hechos es determinar la granularidad. El gran factor y el más decisivo de la granularidad es el tiempo, ya que mientras menor sea el intervalo de tiempo, mayor será el grado de detalle obtenido. Se deben considerar otros factores como la carga del procesador, espacio de almacenamiento y satisfacer a cabalidad los requerimientos del cliente. (9)

- **Jerarquía:** es un conjunto de atributos descriptivos que se rigen por un orden preestablecido. Una jerarquía implica una organización de niveles dentro de una dimensión, con cada nivel representando el total agregado de los datos del nivel inferior. Además, definen como los datos son sumados desde los niveles más bajos hasta los más altos. Una jerarquía puede contener todos los valores que se encuentran en una dimensión, pero no es necesario. (9)

El modelo usado en el MD Inversiones es el esquema constelación, el cual es un modelo multidimensional. Analizando que el sistema de constelación es más complejo que el esquema copo de nieve y estrella, pues tiene muchas tablas de hechos que comparten tablas de dimensiones lo cual se adecúa al negocio de la entidad. En la siguiente imagen se observa un ejemplo de dicho esquema.

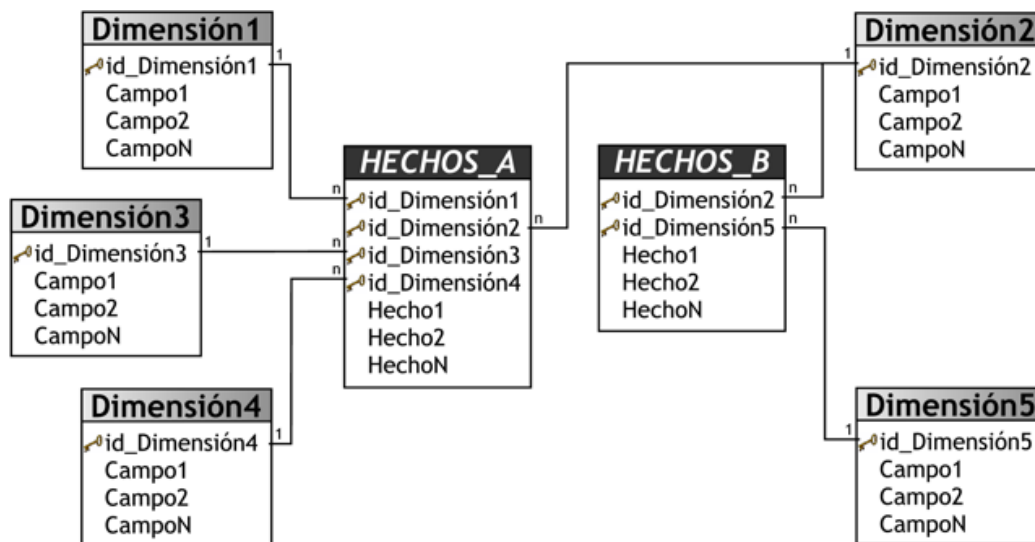


Figura 1: Esquema constelación

Las bases de datos multidimensionales representan los datos como matrices n-dimensionales denominadas hipercubos. Los miembros de una dimensión se agrupan de forma jerárquica. Las dimensiones actúan como índices para identificar valores dentro de la matriz multidimensional. Cada celda

de un hipercubo contiene datos agregados que relacionan los elementos de las dimensiones. En la siguiente imagen se observa un ejemplo de hipercubo:

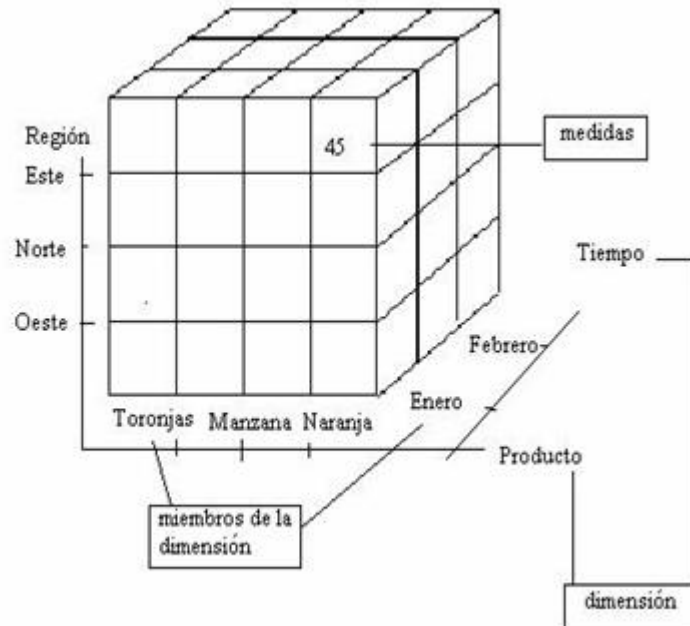


Figura 2: Vista multidimensional usando un cubo

En una base de datos multidimensional la estructura contiene información sobre los elementos de datos ya que las perspectivas están embebidas en la estructura como dimensiones. (7)

El modelo multidimensional es una técnica de diseño lógico que busca presentar los datos en un estándar, que permita una recuperación adecuada de estos. El modelo multidimensional es un modelo adecuado que provee un camino viable para agregar hechos a lo largo de múltiples atributos y dimensiones. Los datos son almacenados como hechos y dimensiones de datos relacional. (10)

### 1.3 Integración de datos

El éxito de un MD se debe sobre todo al buen estudio del negocio a partir de los conocimientos extraídos de las fuentes externas de datos. La integración de los datos es el proceso donde las organizaciones extraen datos desde múltiples fuentes. Estos son limpiados, asegurando la calidad y consistencia de los mismos, se homogenizan los datos de sistemas divergentes para que puedan ser utilizados de una forma conjunta y finalmente se generan los datos en el formato apropiado para ser utilizados posteriormente; luego son cargados en el sistema de almacenamiento para ser analizados. Este proceso tiene gran

importancia, ya que de no realizarse se podrían obtener datos incorrectos que afectarían la toma de decisiones de la organización.

Las principales técnicas de integración de datos utilizadas en los sistemas MD son ETL (Extract-Transform-Load) y ELT (Extract-Load-Transform). Para el MD asociado a este trabajo se seleccionó la técnica ETL ya que esta consiste en la extracción, transformación y posteriormente la carga de los datos no siendo así con la técnica ELT donde ocurre la extracción y carga de los datos antes de la transformación de los mismos, por lo que la técnica escogida es la que se adecúa a las metodologías de desarrollo que se utilizarán en el proceso de construcción del MD.

La técnica de integración de datos ETL consta de 3 procesos fundamentales:

- *Extracción*: es el proceso de extracción de la información de distintas fuentes que pueden ser tanto internas como externas.
- *Transformación*: después de haberse realizado el proceso de extracción los datos son provenientes de diversas fuentes y estos pueden tener errores, ser incoherentes, inconsistentes, estar incompletos y no estar homogenizados. Debido a esto, se realiza un proceso de limpieza donde se busca que los datos sean completos, precisos, interpretables y accesibles. Luego se realiza la integración de los datos con el propósito de identificar las fuentes de datos más fiables y eliminar los problemas de redundancia. Por último se realiza la transformación de los datos para estandarizar los códigos, eliminar los duplicados, usar conversiones y combinaciones para generar nuevos campos.
- *Carga*: es el último proceso donde los datos ya limpios y transformados son cargados en el sistema de destino, ya sea en un MD o AD.

Otra parte importante de los procesos de integración de datos es la identificación de las fuentes, donde se debe realizar un estudio del negocio en el cual se identifiquen las fuentes involucradas en el mismo y se determinen sus particularidades, para que permitan clasificar, planificar y desarrollar los mecanismos de integración.

### 1.4 Inteligencia de negocios

Inteligencia de negocios (Business Intelligence) se refiere a las tecnologías utilizadas para usar de forma efectiva la información con el fin de mejorar las operaciones realizadas en el negocio de las organizaciones. Un buen sistema de inteligencia de negocios debe ofrecerles a los usuarios del negocio acceso a la información importante que necesitan para realizar sus tareas diarias y principalmente que los ayude en el proceso de toma de decisiones.

Inteligencia de negocios es un concepto que integra como solución el almacenamiento y procesamiento de enormes cantidades de datos e información para transformarla en conocimiento y decisiones en tiempo real a través de una fácil explotación. (11)

Algunas de sus principales **características** son:

- **Accesibilidad a la información:** los datos son la fuente principal de este concepto. Lo primero que deben garantizar este tipo de herramientas y técnicas será el acceso de los usuarios a los datos con independencia de la procedencia de estos.
- **Apoyo en la toma de decisiones:** se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular sólo aquellos datos que les interesen.
- **Orientación al usuario final:** se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas. (11)

### 1.4.1 Procesos OLTP y OLAP

Un elemento de vital importancia para el desarrollo de los procesos de inteligencia de negocio es el diseño de los cubos multidimensionales, los cuales pueden realizarse mediante las tecnologías OLAP y OLTP.

**Procesos OLTP:** OLTP (Procesamiento de Transacciones en Línea) permite tener respuestas rápidas en las transacciones que se realizan en una organización. Sin embargo, tiene la desventaja de que no son eficientes en el manejo de grandes volúmenes de información ni en consultas de mucha complejidad. Estos sistemas soportan las actualizaciones ya que los datos están normalizados y se han eliminado las redundancias pero los usuarios deben analizar los datos registro por registro lo que toma una gran cantidad de tiempo si se trata de gran cantidad de información. Por lo antes mencionado este sistema no resulta óptimo para los sistemas de ayuda a la toma de decisiones.

**Procesos OLAP:** la tecnología de Procesamiento Analítico en Línea (OLAP) permite el uso más eficiente de los AD en línea, proporcionando respuestas rápidas a consultas analíticas complejas e iterativas utilizadas generalmente para sistemas de ayuda a la toma de decisiones. Presenta los datos de manera natural, de forma que los usuarios puedan entender mejor la información.

#### Principales características de OLAP:

- **Rápido:** proporciona información al usuario a una velocidad constante (principalmente en cinco segundos o menos).

- **Análisis:** realiza análisis estadísticos y numéricos básicos de los datos.
- **Compartida:** permite compartir los datos potencialmente confidenciales a través de una gran cantidad de usuarios, implementando para esto los requerimientos de seguridad necesarios.
- **Multidimensional:** permite ver la información en determinadas vistas o dimensiones.
- **Información:** accede a todos los datos necesarios, donde quiera que estos residan, y mientras no esté limitada por el volumen. (12)

### Existen diferentes tipos de OLAP:

- **ROLAP (Procesamiento Analítico en Línea Relacional):** organización física que se implementa sobre tecnología relacional, pero que tiene algunas facilidades para mejorar el rendimiento. Accede directamente a los datos del AD y soporta técnicas de optimización de accesos para acelerar las consultas, por ejemplo, partición de datos a nivel de aplicación, soporte a la desnormalización y joins múltiples. Esta organización tiene como desventaja que es más lenta que las demás estrategias de almacenamiento. (13)
- **MOLAP (Procesamiento Analítico en Línea Multidimensional):** usa las bases de datos multidimensionales bajo la premisa que OLAP está mejor implantado almacenando los datos multidimensionalmente. Tiene el mejor tiempo de respuesta, excelente rendimiento y compresión de datos. Tiene la desventaja que es una solución particular, para soluciones con volúmenes de información y cantidad de dimensiones más bien modestos. (13)
- **HOLAP (Procesamiento Analítico en Línea Híbrido):** combina las arquitecturas ROLAP y MOLAP brindando así lo mejor de las dos: un desempeño superior y gran escalabilidad. Mantiene los registros de detalle en la base de datos relacional y las agregaciones en un almacén MOLAP separado. Es usado generalmente para cubos que requieren rápida respuesta con una gran cantidad de datos. (13)

El tipo de OLAP que se utiliza en el desarrollo de este trabajo es ROLAP debido a las características que presenta, las cuales fueron mencionadas anteriormente.

### 1.5 Metodologías de desarrollo

La metodología es la encargada del estudio de los métodos, los cuales brindan los procedimientos para lograr los objetivos. Las metodologías son parte del proceso de investigación que posibilita la sistematización de los métodos y de las técnicas necesarias para llevarlas a cabo. En la actualidad existen muchas metodologías para el desarrollo de software y concretamente algunas de estas inciden en el



proceso de desarrollo de los MD, a continuación se analizan las dos tendencias de desarrollo más significativas que son las metodologías definidas por William H. Inmon y Ralph Kimball:

**Metodología de Kimball:** se basa en dividir el mundo de Inteligencia de Negocios entre los hechos y las dimensiones, esta es eficaz y conduce a una solución completa en un corto período de tiempo. Además, tiene abundante documentación y se puede encontrar una respuesta a casi todas las preguntas que se puedan tener. Entre sus características principales están que su arquitectura es ascendente (bottom-up), que plantea que se debe crear por cada departamento un conjunto de MD independientes, orientados a los temas que estén relacionados con él. (14)

**Metodología de Inmon:** la visión de Inmon se basa en un enfoque descendente (top-down), es decir, de necesidades generales a tareas específicas, propone construir primero el AD y a partir de este los MD, plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Al ser construido descendentemente, los MD se nutren del AD 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 la base para los MD, por lo que no se crean los MD directamente desde los Sistemas de Procesamiento de Transacciones en Línea (OLTP) sino a través de un área de ensayo. (14)

Analizando entonces estas dos tendencias de desarrollo se puede ver claramente que el enfoque que brinda Kimball parte de lo específico a lo general, donde comienza con elementos simples, encaminando la solución hacia arriba para buscar una mayor complejidad, es decir se selecciona un subconjunto del problema, para luego ir hacia la solución global.

Mientras que Inmon materializa un enfoque que en la práctica es muy difícil ya que en muchas ocasiones la organización no se encuentra muy al tanto de la tecnología y no tiene la suficiente experiencia en el desarrollo de aplicaciones. Este enfoque conlleva a una poca consulta por parte de los usuarios finales sobre el MD ya que su implementación es planificada principalmente en la cabeza de la organización.

No se puede decir que una de estas dos tendencias es mejor que la otra ya que los beneficios de cada metodología depende del ambiente de la empresa en que se vayan a implantar. Basados en estas propuestas se han desarrollado un conjunto de metodologías que no siguen obligatoriamente una específica sino que realizan una selección de lo mejor de cada una y definen su propia metodología. Como ejemplo se presenta el modelo para el desarrollo de soluciones de Almacenes de Datos e Inteligencia de Negocio (DW&BI) creada por DATEC adaptada a las necesidades del centro la cual utiliza

una adaptación de la tesis de Doctorado de Leopoldo Zenaido y como base la metodología Ciclo de vida Kimball por los siguientes elementos:

- Utiliza el concepto de dimensiones y hechos, lo cual facilita el proceso de desarrollo y lo hace más eficaz.
- Propone ir construyendo el AD a través de la construcción de los MD departamentales, lo que constituye una buena estrategia y coincide con la división lógica de las empresas, entidades y organismos.
- Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles.

Es importante mencionar que la metodología creada por DATEC es una propuesta resistente y adaptable a cambios, que dentro del ciclo de vida presenta una serie de flujos de trabajos que serán mencionados a continuación:

- **Estudio preliminar o Planeación:** se realiza el estudio de la entidad cliente, la planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y otras actividades.
- **Requerimientos:** se realiza en dos direcciones, una, mediante la identificación de las necesidades de información y reglas del negocio; y la otra con un levantamiento detallado de las fuentes de datos a integrar. Después se procede a la definición de los requerimientos.
- **Arquitectura y Diseño:** se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga y se define la arquitectura de información que regirá el desarrollo de la solución.
- **Implementación:** se diseña físicamente el repositorio de datos, se crean las estructuras de almacenamiento, el área temporal de almacenamiento, se ejecutan las reglas de extracción, transformación y carga, se configuran e implementan las herramientas de inteligencia de negocio para la obtención de los elementos (reportes y gráficos) que se acordaron con el cliente final.
- **Prueba:** se realizan las pruebas al sistema desde las pruebas de unidad hasta las de aceptación con el cliente final.
- **Despliegue:** se realiza un despliegue piloto en el cual se configuran los servidores, se instalan las herramientas y se carga una muestra de los datos para demostrar que el sistema funciona. Posterior a la aceptación del cliente se realiza la carga de los datos, capacitaciones y transferencia tecnológica.
- **Soporte y Mantenimiento:** tras la implantación de la solución se brindan los servicios de soporte en línea, vía telefónica, web u otras según el contrato firmado y las condiciones de soporte establecidas.

# Capítulo I: Fundamentación teórica de la investigación

- **Gestión y Administración del proyecto:** a lo largo del ciclo de vida se realizan actividades de control, gestión y chequeo del desarrollo, los gastos, las utilidades, los recursos y demás actividades por parte del Grupo de dirección del proyecto.

Como se ha descrito, en cada flujo intervienen grupos específicos, cada uno con actividades y responsabilidades concretas siendo los encargados de todo el trabajo y estos grupos específicos son:

- Grupo de Análisis
- Grupo de Almacenes de Datos
- Grupo de ETL
- Grupo de BI
- Grupo de Dirección

En la siguiente imagen se observa la representación de los grupos en cada flujo de trabajo.

Grupos/ Flujos	Estudio Preliminar	Requerimientos	Arquitectura y Diseño	Implementación	Prueba	Despliegue	Soporte y Mantenimiento
Análisis	Participa	Responsable	Participa	No Participa	Responsable	No Participa	No Participa
Almacén	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
ETL	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
BI	Participa	No Participa	Responsable	Responsable	Participa	Responsable	Participa
Dirección	Responsable	Responsable	Participa	Responsable	Responsable	Responsable	Participa



**Leyenda:**  
 **Responsable**  
 **Participa**  
**No Participa**

Figura 3: Relación de los Grupos con los flujos de trabajo

## 1.6 Herramientas de desarrollo

Con los avances tecnológicos hoy en día existen muchas herramientas de desarrollo para la implementación de un MD. El MD Inversiones representa una de las áreas departamentales que forma parte del AD Sistema de Información de Gobierno, las herramientas a utilizar se documentan a continuación.

## 1.6.1 Herramienta Case de Modelado

Las herramientas CASE representan una forma que permite modelar los procesos de negocios de las empresas y desarrollar los sistemas de información gerenciales. A continuación se mencionan algunas de las características de la herramienta Visual Paradigm, la cual es la utilizada en este trabajo. (15)

**Visual Paradigm** en su versión 6.4 es una herramienta de modelado UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Está dotada de una buena cantidad de productos o módulos para facilitar el trabajo durante la confección de un software, lo cual garantiza la calidad del producto final.

**Entre sus características más importantes están:**

- Soporta aplicaciones web.
- Posee generación de código para Java y exportación como HTML.
- Fácil de instalar y actualizar.
- Tiene compatibilidad entre ediciones.
- Permite realizar diagramas de procesos de negocio: proceso, decisión, actor de negocio, documento.
- Tiene modelado colaborativo con CVS y Subversion.
- Tiene opción de ingeniería inversa - código a modelo, código a diagrama.
- Posee generación de código - modelo a código, diagrama a código.
- Presenta editor de detalles de casos de uso, entorno todo en uno para la especificación de los detalles de los casos de uso, incluyendo la especificación del modelo general y de las descripciones de los casos de usos.
- Posee generación de bases de datos. Transformación de diagramas de entidad-relación en tablas de base de datos.
- Distribución automática de diagramas - reorganización de las figuras y conectores de los diagramas UML. (15)

Esta herramienta no es de código abierto pero la universidad paga su licencia por lo que puede ser utilizada.

## 1.6.2 Sistema Gestor de Bases de Datos

Los Sistemas Gestores de Bases Datos (SGBD) son aplicaciones que permiten a los usuarios definir, crear y mantener Bases de Datos (BD), proporcionando un acceso controlado a la misma, es decir un

SGBD es un software de propósito general que facilita el proceso de definir, construir y manipular BD para diversas aplicaciones. Pueden ser de propósito general o específico. (16)

Presenta 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 BD** 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. (16)

El gestor de bases de datos seleccionado es PostgreSQL versión 8.4.

**PostgreSQL** versión 8.4 es un SGBD, que no pertenece a ninguna compañía, sino que es dirigido por una comunidad de desarrolladores, que se hace llamar “PostgreSQL Global Development Group” y organizaciones comerciales que están a cargo de su desarrollo. Tiene más de 15 años de desarrollo activo y se ha ganado la reputación de ser confiable y mantener la integridad de los datos. Existen compañías que aseguran haberlo empleado durante varios años y con altas tasas de actividad, sin experimentar problemas de ningún tipo. El PostgreSQL se ejecuta en la mayoría de los sistemas operativos más utilizados en el mundo, incluido Linux, versiones de UNIX y, por supuesto, Microsoft Windows XP o superior.

Las **características** más relevantes de este gestor son:

- Soporta distintos tipos de datos: además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, elementos gráficos, datos sobre redes (MAC, IP), cadenas de bits. También permite la creación de tipos propios.

- Incorpora funciones de diversa índole: manejo de fechas, geométricas, orientadas a operaciones con redes.
- Permite la declaración de funciones propias, así como la definición de disparadores.
- Soporta el uso de índices, reglas y vistas.
- Incluye herencia entre tablas (aunque no entre objetos, ya que no existen), por lo que a este gestor de bases de datos se le incluye entre los gestores objeto-relacionales.
- Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos. (17)

Presenta numerosas **ventajas** como son:

- **Instalación ilimitada:** PostgreSQL no tiene costo asociado a la licencia del software. Esto tiene varias ventajas adicionales como son modelos de negocios más rentables con instalaciones a gran escala, no existe la posibilidad de ser auditado para verificar cumplimiento de licencia en ningún momento, flexibilidad para hacer investigación y desarrollo sin necesidad de incurrir en costos adicionales de licenciamiento.
- **Ahorros considerables en costos de operación:** PostgreSQL fue diseñado y creado para tener un mantenimiento y ajuste mucho menor que los productos de los proveedores comerciales, conservando todas las características, estabilidad y rendimiento. Sus programas de entrenamiento son reconocidamente mucho más costo-efectivos, manejables y prácticos en el mundo real que aquellos de los principales proveedores comerciales.
- **Estabilidad y confiabilidad legendarias:** en contraste a muchos sistemas de bases de datos comerciales, es extremadamente común que compañías reporten que PostgreSQL nunca ha presentado caídas en varios años de operación de alta actividad.
- **Extensible:** el código fuente está disponible para todos sin costo. Si se necesita extender o personalizar PostgreSQL de alguna manera, se puede hacer con un mínimo esfuerzo, sin costos adicionales. Esto es complementado por la comunidad de profesionales de PostgreSQL.
- **Multiplataforma:** PostgreSQL está disponible en casi cualquier Unix (34 plataformas en la última versión estable), y una versión nativa de Windows está actualmente en estado beta de pruebas.
- **Diseñado para ambientes de alto volumen:** PostgreSQL usa una estrategia de almacenamiento de filas llamada MVCC para conseguir mejor respuesta en ambientes de grandes volúmenes. Los principales proveedores de sistemas de bases de datos comerciales usan también esta tecnología, por las mismas razones. (18)

Una de las ventajas más importantes que tiene esta herramienta es que es multiplataforma, de código abierto y no presenta problemas con las licencias, lo que es de vital importancia ya que la universidad se encuentra en migración hacia el software libre y la independencia tecnológica.

### 1.6.3 Herramientas para la integración de datos

A continuación se aborda sobre las principales características y ventajas que poseen las herramientas que fueron previamente seleccionadas para utilizar en la implementación de la integración de datos del MD Inversiones.

**El DataCleaner** en su versión 1.5.3 es una aplicación de código abierto para el perfilado, la validación y comparación de datos. Estas actividades ayudan a administrar y supervisar la calidad de los datos con el fin de garantizar que la información es útil y aplicable a su situación de negocio. Es la alternativa gratuita al software de gestión de datos maestros, metodologías, almacenamiento de datos, proyectos de investigación estadística y la preparación para el proceso de extracción, transformación y carga. (19)

El DataCleaner puede acceder y analizar prácticamente cualquier AD, incluyendo:

- Bases de datos como Oracle, Microsoft SQL Server, PostgreSQL, MySQL, OpenOffice (ODB)
- Los archivos separados por comas y separados por tabuladores (.csv / .tsv)
- Hojas de cálculo Excel (.xls)
- Archivos XML

**Pentaho Data Integration (Kettle)** en su versión 4.0.1 es una herramienta de integración de datos, la cual incluye un conjunto de herramientas que proporcionan una extracción de gran alcance, transformación y carga utilizando un enfoque innovador, orientado a los metadatos. Uno de sus objetivos es que el proceso sea fácil de generar, mantener y desplegar. (20)

Además de ser de código abierto y sin costos de licencia, las **características** básicas de esta herramienta son:

- Entorno gráfico de desarrollo.
- Uso de tecnologías estándar: Java, XML, JavaScript.
- Fácil de instalar y configurar.
- Multiplataforma: Windows, Macintosh, Linux.
- Basado en dos tipos de objetos: transformaciones (colección de pasos en un proceso de extracción, transformación y carga) y trabajos (colección de transformaciones).
- Incluye cuatro herramientas:

- Spoon: utilizado para diseñar transformaciones ETL usando el entorno gráfico.
- PAN: se usa para ejecutar transformaciones diseñadas con spoon.
- CHEF: específicamente para crear trabajos.
- Kitchen: específicamente para ejecutar trabajos.

### 1.6.4 Herramientas para la Inteligencia de Negocios

Las herramientas de Inteligencia de Negocios permiten la posibilidad de ofrecerles a los usuarios finales la visualización de todo el volumen de información que se tenga almacenada proporcionando un nivel de uso, capacidad, carga y robustez compatibles con disímiles necesidades críticas. En este caso las herramientas utilizadas se abordan a continuación.

**El Pentaho Business Intelligence Suite Enterprise Edition (Pentaho BI)** versión 3.6 es una herramienta que proporciona al usuario final una simplicidad y una escalabilidad mejorada. Pentaho BI es una plataforma de inteligencia de negocios orientada a soluciones y centrada en procesos, que incluye todos los principales componentes requeridos para implementar soluciones basadas en procesos y ha sido concebida desde el principio para estar basada en procesos. Esta herramienta es la encargada de orquestar todos los procesos de inteligencia de negocio. Sobre esta plataforma se definen las áreas de interés que poseen los usuarios para la preparación de reportes, consultas dinámicas y la realización del análisis OLAP. (21)

En el plano de código abierto la herramienta más significativa es Mondrian, la cual es una de las aplicaciones más importantes de la Suite Pentaho BI y la utilizada en el departamento “Almacenes de Datos”.

**Mondrian** en su versión 3.0.4 es un servidor OLAP de código abierto que gestiona la comunicación entre una aplicación OLAP y la base de datos con los datos fuente. Es desarrollado en Java/Servlets/JSPs, permite ser instalado en servidores de aplicaciones como JBoss. Entre sus principales características se encuentra la facilidad para el análisis de grandes volúmenes de información que se encuentren almacenados en bases de datos que soporten JDBC. Mondrian soporta el lenguaje Microsoft's Multidimensional Expressions (DMX). También soporta los APIs: Java OLAP (JOLAP) y XML para el análisis de aplicaciones programadas. Permite crear cubos de información, dichos cubos se componen de archivos XML y en ellos se definen las dimensiones y las conexiones de los datos. Los archivos XML por lo general son complejos de realizar manualmente por lo que es común utilizar herramientas gráficas para realizar la edición de estos. (21)



Los esquemas Mondrian se alojan en un archivo XML. Este esquema puede ser editado de forma manual, o utilizando las herramientas Workbench o Cube Designer de Mondrian, en el caso del departamento se utiliza el Workbench.

**El Workbench** en su versión 3.0.3 es una herramienta para el desarrollo de esquemas multidimensionales en XML Workbench desarrollada en Java. Este programa recientemente publicado (2007) entrega todas las facilidades para poder realizar el modelo lógico del cubo OLAP al cual se le realizarán las consultas. Este programa se conecta directamente con la base de datos para así poder diseñar los cubos OLAP que se requieren para que el usuario final pueda visualizar los indicadores. Luego, el archivo generado se utiliza para definir la estructura de cubo en Jasper. (21)

Como servidor web en la presente investigación se utiliza el **Apache Jakarta Tomcat** en su versión 5.5. El mismo es un servidor de código abierto, un contenedor de aplicaciones Web basadas en Java que fue creado para ejecutar *Servlets* y *Java Server Page*, (JSP por sus siglas en inglés), de aplicaciones web. Existe en el marco del subproyecto Apache Jakarta, donde es apoyado y reforzado por un grupo de voluntarios de la comunidad de código abierto de Java. Tomcat puede funcionar como servidor web por sí mismo. En sus inicios existió la percepción de que el uso de Tomcat de forma autónoma era sólo recomendable para entornos de desarrollo y entornos con requisitos mínimos de velocidad y gestión de transacciones. Hoy en día ya no existe esa percepción y Tomcat es usado como servidor web autónomo en entornos con alto nivel de tráfico y alta disponibilidad. Dado que Tomcat fue escrito en Java, funciona en cualquier sistema operativo que disponga de la máquina virtual Java. (22)

### **1.7 Conclusiones**

Este capítulo abarca una perspectiva general del proceso de desarrollo de un MD y se investigaron temas sobre los AD y MD, como son sus ventajas y características, dejando de manera clara que un MD es la solución para este trabajo. Se documenta la metodología a utilizar: el modelo para el desarrollo de soluciones de Almacenes de Datos e Inteligencia de Negocio (DW&BI), desarrollada por DATEC. El diseño de la solución se hace mediante el modelo constelación y se utiliza el proceso de almacenamiento ROLAP. Se desarrolla el sistema mediante las herramientas: Visual Paradigm 6.4, PostgreSQL 8.4, DataCleaner 1.5.3, Pentaho Data Integration (Kettle) 4.0.1, Pentaho Business Intelligence Suite Enterprise Edition (Pentaho BI) 3.6, Mondrian 3.0.4, Workbench 3.0.3 y Apache Jakarta Tomcat 5.5.

### **CAPÍTULO II: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS INVERSIONES**

#### **Introducción**

Para el desarrollo del MD Inversiones es de vital importancia realizar un buen estudio preliminar del negocio para captar las necesidades del cliente y que se logre un buen desarrollo de toma de decisiones. En este capítulo se documenta todo lo referente al estudio del negocio, así como también se exponen las necesidades de información, las reglas del negocio, el levantamiento de los requisitos, la descripción de los casos de uso, se identifican las dimensiones, hechos y medidas con que cuenta el negocio, se desarrolla la matriz BUS. Se realiza el diseño de los procesos de integración de datos y se diseña el mapa de navegación. Estos datos guiarán posteriormente a los desarrolladores a lograr un producto con la calidad requerida.

#### **2.1 Estudio preliminar del negocio**

La ONE tiene como objetivo general 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 y las demás necesidades del país en información estadística. En el SEN se incluyen los órganos, organismos, instituciones y entidades que elaboran estadísticas en Cuba y está integrado por tres subsistemas:

- SIEN: Sistema de Información Estadística Nacional. Está dirigido a la elaboración de estadísticas destinadas a satisfacer los requerimientos informativos de los más altos niveles del Estado y el Gobierno además de informar a los organismos internacionales. La ONE controla la información relacionada con la población y las entidades, siendo la productora a través de su red.
- SIEC: Sistema de Información Estadística Complementaria. Está dirigido a la elaboración de estadísticas destinadas a satisfacer los requerimientos informativos de los organismos a los efectos del control administrativo de sus entidades. La ONE revisa y ejerce su control metodológico, aunque no es la productora a través de su red. Controla la información de las entidades subordinadas y no subordinadas.
- SIET: Sistema de Información Estadística Territorial. Dirigido a la elaboración de estadísticas destinadas a satisfacer los requerimientos informativos del Gobierno en el territorio y otros actores locales.

## *Capítulo II: Análisis y diseño del Mercado de Datos Inversiones*

---

A través de SEN, la ONE con sus 15 delegaciones territoriales SIET y 168 oficinas municipales en todo el país que conforman el SIEC. La información una vez que es captada directamente de los centros informantes es digitalizada y se desarrollan centros de estudio, de población y desarrollo. (4)

Específicamente en el área de inversiones se muestra una visión general del proceso inversionista del sector civil en el país. En los reportes se muestra una ejecución física de inversiones por componentes, de la economía nacional. Los datos que se presentan son reportados, a través de los modelos del SIEN 5702 y 1006, a salida de empresa, unidades presupuestadas y otras entidades contempladas en dichos sistemas.

El registro económico de los datos sobre las inversiones está establecido en todas las entidades que acometen inversiones, ya sea por concepto de construcción civil y montaje de equipos, como por la adquisición de equipos y suministros con los mismos fines, así como por incurrir en otros gastos vinculado directamente con los objetivos inversionistas.

Las estadísticas de inversiones, cuyos proyectos son previamente aprobados por el órgano de planificación correspondiente, abarca todo el proceso inversionista en su conjunto, aunque por cuestiones de orden práctico, las cifras estadísticas se registran a partir del momento en que se incurre en el primer gasto, independientemente de su pago y con cargo a cada proyecto de inversión definido. (4)

### **2.2 Necesidades de información**

Las necesidades de información responden a la información que necesita manejar el cliente con el producto. De las necesidades de información se obtienen los requisitos de información, los cuales son información que debe estar disponible para su consulta, los mismos resultan la entrada fundamental para el posterior proceso de inteligencia de negocio y los posibles reportes candidatos. Después de realizar un análisis minucioso del negocio se definen las siguientes necesidades existentes en la ONE en la temática de inversiones las cuales corresponden a los modelos 1006 y 5702:

#### **Modelo 1006: gastos de inversión en actividades de protección del medio ambiente.**

- Obtener gastos de inversión en actividades de protección del medio ambiente por entidad, organismo, DPA, NAE, indicadores de ámbito ambiental en un tiempo dado.

#### **Modelo 5702: control del plan de la ejecución física de las inversiones.**

- Obtener el control del plan de la ejecución física de las inversiones por entidad, organismo, DPA, NAE, indicadores de ejecución física en un tiempo dado.

### **2.3 Requisitos de información**

Luego de realizar un análisis de las necesidades de información existentes en la ONE con respecto a las inversiones se definieron los requisitos de información (RI), a continuación se pueden observar algunos de los RI definidos para el modelo 1006, para mayor profundización consultar el artefacto “Especificación de Requisitos” del expediente de proyecto:

#### **Modelo 1006**

- RI1.** Obtener total de gastos de inversión en actividades de protección del medio ambiente, por ámbito ambiental en actividades de protección del aire y el clima, por entidad, organismo, DPA, NAE en un tiempo dado.
- RI2.** Obtener total de gastos de inversión en actividades de protección del medio ambiente, por ámbito ambiental en actividades de reducción del ruido y las vibraciones, por entidad, organismo, DPA, NAE en un tiempo dado.
- RI3.** Obtener total de gastos de inversión en actividades de protección del medio ambiente, por ámbito ambiental en actividades de gestión de las aguas, por entidad, organismo, DPA, NAE en un tiempo dado.
- RI4.** Obtener total de gastos de inversión en actividades de protección del medio ambiente, por ámbito ambiental en actividades de residuos, por entidad, organismo, DPA, NAE en un tiempo dado.
- RI5.** Obtener total de gastos de inversión en actividades de protección del medio ambiente, por ámbito ambiental en actividades de protección y rehabilitación de los suelos, por entidad, organismo, DPA, NAE en un tiempo dado.
- RI6.** Obtener total de gastos de inversión en actividades de protección del medio ambiente, por ámbito ambiental en actividades de protección de la biodiversidad y los paisajes, por entidad, organismo, DPA, NAE en un tiempo dado.
- RI7.** Obtener total de gastos de inversión en actividades de protección del medio ambiente, por ámbito ambiental en actividades de protección contra las radiaciones, por entidad, organismo, DPA, NAE en un tiempo dado.
- RI8.** Obtener total de gastos de inversión en actividades de protección del medio ambiente, por ámbito ambiental en actividades de investigación y desarrollo, por entidad, organismo, DPA, NAE en un tiempo dado.

**RI9.** Obtener total de gastos de inversión en actividades de protección del medio ambiente, por ámbito ambiental en otras actividades de protección del medio ambiente, por entidad, organismo, DPA, NAE en un tiempo dado.

### **2.4 Reglas de Negocio**

Las reglas de negocio no son más que condiciones que deben satisfacerse o políticas que deben cumplirse, por lo que éstas regulan el negocio. Luego de realizar un análisis minucioso del negocio se identificaron las reglas de negocio, a continuación se muestran algunas de las reglas identificadas en el modelo 1006, para mayor profundización consultar el artefacto “Reglas de negocio y transformación” del expediente de proyecto:

#### **Modelo 1006**

**RN1.** Total = Protección del aire y el clima + Reducción del ruido y las vibraciones + Gestión de las aguas + Residuos + Protección y rehabilitación de los suelos + Protección de la biodiversidad y los paisajes + Protección contra las radiaciones + Investigación y desarrollo+ Otras actividades de protección del medio ambiente.

**RN2.** Protección del aire y el clima = Prevención de la contaminación atmosférica por modificación de procesos + Tratamiento de los gases de escape y el aire de ventilación + Empleo de fuentes renovables de energía y ahorro de energía + Sustitución y eliminación de sustancias agotadoras de la capa de ozono + Equipos e instalaciones para el control, medición y eliminación de la contaminación atmosférica + Otras actividades.

**RN3.** Protección del aire y el clima  $\leq$  Total.

**RN4.** Reducción del ruido y las vibraciones = Modificaciones preventivas en el origen + Construcción de dispositivos antirruído y antivibraciones + Equipos e instalaciones para la medición y el control del ruido y las vibraciones + Otras actividades.

**RN5.** Reducción del ruido y las vibraciones  $\leq$  Total.

### **2.5 Requisitos funcionales**

Los requisitos funcionales (RF) son condiciones o capacidades que el sistema debe cumplir dando respuesta a los requisitos de información. Los mismos definen el comportamiento interno del software como la manipulación de datos y otras funcionalidades específicas que indican cómo se desarrollarán en la práctica los casos de uso. Estos son identificados a partir de conocer las necesidades de los usuarios y

de las reglas de negocio identificadas. A continuación se enuncian varios de los requisitos funcionales que fueron identificados, para mayor profundización de los mismos consultar el artefacto “Especificación de Requisitos” del expediente de proyecto:

**RF1.** Extraer datos

**RF2.** Realizar transformación y carga de los datos

**RF3.** Eliminar usuario

**RF4.** Agregar usuario

**RF5.** Eliminar rol

**RF6.** Agregar rol

**RF7.** Eliminar reporte

**RF8.** Agregar reporte

**RF9.** Modificar reporte

**RF10.** Autenticar usuario

### **2.6 Requisitos no funcionales**

Los requisitos no funcionales (RNF) son los que hacen referencia a las características que debe tener el sistema como pueden ser la fiabilidad, rendimiento, entre otros. Los mismos son los encargados de limitar las condiciones en las que el producto debe funcionar de forma ideal para el cliente. A continuación se enuncian algunos de los RNF más importantes, para mayor profundización de estos consultar el artefacto “Especificación de Requisitos” del expediente de proyecto:

#### **Requisitos de Software**

**RNF36.** Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema.

Las configuraciones de software de las máquinas clientes deben tener al menos los siguientes elementos:

- Sistema Operativo Windows 98 o superior o cualquiera de las distribuciones de Linux.
- Firefox 2.0 o superior.
- Java Virtual Machine 6.0 y Schema Workbench 3.2.1 en caso de que un usuario capacitado requiera la construcción de esquemas multidimensionales para el diseño de nuevos reportes.

#### **Requisitos de Hardware**

**RNF37.** Proporcionar características mínimas de hardware a las estaciones de trabajo

Las características técnicas mínimas de hardware deben ser las siguientes:

- Procesador de doble núcleo a 2.8 GHz.
- 1 GB de RAM, DDR2 667 MHz.
- 160 GB de disco duro, SATA 5400 RPM.

**RNF38.** Proporcionar características mínimas de hardware a los servidores.

Los servidores deben contar con los siguientes requerimientos de hardware para lograr una explotación aceptable del sistema:

- Servidores Blade de gama media.
- 16 GB RAM.
- 72 o 145 GB de Disco Duro.
- 1 Microprocesador Core2Duo.

### 2.7 Modelo de Casos de Uso del Sistema

Los casos de uso del sistema son los encargados de que el usuario pueda obtener la información que desea, se utilizan para representar los requisitos del sistema a implementar y como interactúa cada usuario con el sistema. A continuación se mencionan los actores que fueron identificados en el MD y se muestra el Diagrama de Casos de Uso del Sistema, para mayor detalle de la especificación de cada uno de los mismos se puede consultar el artefacto “Modelo de Casos de Uso del Sistema” del expediente de proyecto.

#### 2.7.1 Actores del sistema

- **Especialista:** es el usuario que utiliza el sistema para buscar información relacionada con los indicadores de inversiones y puede autenticarse.
- **Administrador de ETL:** es el usuario encargado de realizar el proceso de extracción, transformación y carga de los datos en el MD.
- **Administrador del Sistema:** es el usuario que además de las tareas del especialista también administra usuarios, administra rol y administra reportes.

2.7.2 Diagrama de Casos de Uso del Sistema

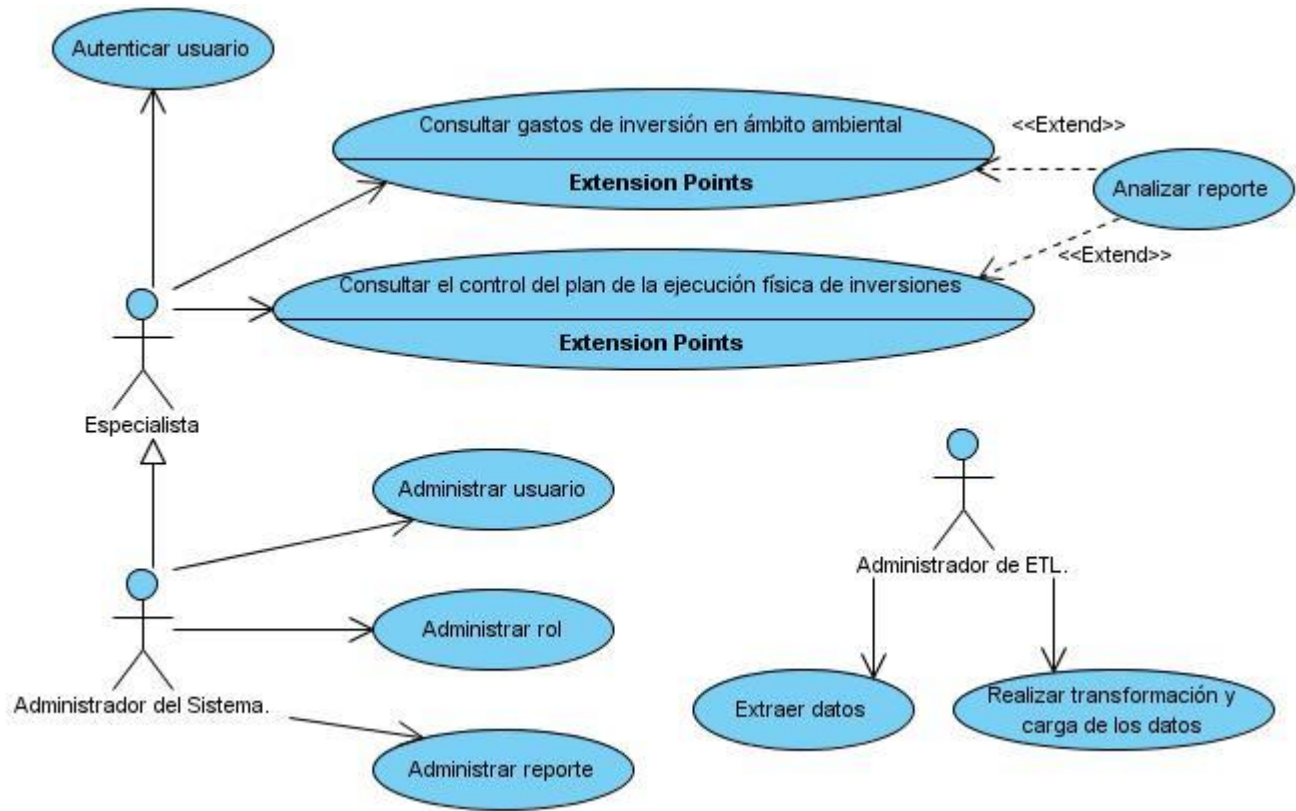


Figura 4: Diagrama de Casos de Uso del Sistema

2.7.3 Casos de Uso de Información

Los casos de uso de información (CUI) son aquellos que responden a los requisitos de información, lo cual es la información que desea obtener o consultar el cliente. Los identificados para la realización del MD son: consultar gastos de inversión en ámbito ambiental y consultar el control del plan de la ejecución física de las inversiones. A continuación se muestra la descripción de uno de los CUI, para mayor detalle del resto de CUI consultar el artefacto “Modelo de Casos de Uso del Sistema” del expediente de proyecto:

- CUI1: Consultar gastos de inversión en ámbito ambiental.

<b>Caso de Uso:</b>	Consultar gastos de inversión en ámbito ambiental
<b>Tipo:</b>	Información
<b>Actores:</b>	Especialista, Administrador del sistema



## *Capítulo II: Análisis y diseño del Mercado de Datos Inversiones*

<b>Resumen:</b>	El caso de uso inicia cuando el especialista o el administrador del sistema desean consultar un reporte de Gastos de inversión en actividades de protección del medio ambiente. Se consulta el reporte deseado. Finaliza cuando se muestra el reporte.	
<b>Precondiciones:</b>	Debe haber completitud en el almacén de datos y los datos deben estar cargados.	
<b>Referencias</b>	RI1, RI2, RI3 RI4, RI5, RI6, RI7, RI8, RI9, RI10, RI11	
<b>Prioridad</b>	Crítico	
<b>Flujo Normal de Eventos</b>		
<b>Acción del Actor</b>	<b>Respuesta del Sistema</b>	
1. El actor entra al sistema.	2. El sistema muestra el A.A.G SIGOB.	
3. El actor despliega el A.A.G SIGOB.	4. El sistema muestra las A.A existentes en el A.A.G SIGOB.	
5. El actor despliega el A.A Inversiones.	6. El sistema muestra los LT existentes en el A.A Inversiones.	
7. El actor selecciona el LT Gastos de inversión en actividades de protección del medio ambiente.	8. El sistema muestra los reportes existentes en el LT Gastos de inversión en actividades de protección del medio ambiente.	
9. El actor selecciona el reporte que desea visualizar.	10. El sistema carga y muestra el reporte seleccionado en un área de trabajo (Ver CUF Analizar reporte).	
<b>Opciones de reportes de Gastos de inversión en actividades de protección del medio ambiente.</b>		
<b>Perspectivas de análisis</b>	<b>Posibles resultados</b>	
	<b>Medidas</b>	<b>Periodicidad</b>
Dimensiones disponibles relacionadas con el caso de uso Consultar gastos de inversión en actividades de protección del medio ambiente.	➤ Gastos total	Trimestral

- Entidad
- Organismo
- DPA
- NAE
- Trimestre
- Indicadores
- UM

**Prototipo de Interfaz**

Organismo	Entidad	Indicadores	UM	Medidas
Todos	Todos	Medio ambiente	MP	501.214
		Total	MP	384.229,5
		Protección del aire y el clima	MP	31.286,4
		Reducción del ruido y las vibraciones	MP	76,1
		Gestión de las aguas	MP	208.683,5
		Residuos	MP	17.997
		Protección y rehabilitación de los suelos	MP	14.674,3
		Protección de la biodiversidad y los paisajes	MP	242.743,2
		Protección contra las radiaciones	MP	
		Investigación y desarrollo	MP	
		Otras actividades de protección del medio ambiente	MP	
		Cuencas	MP	112.766,6

**Poscondiciones**

Disponibilidad de opciones (cruces de variables) de reportes relacionados con el L.T Gastos de inversión en actividades de protección del medio ambiente.

Tabla 1: CUI1: Controlar gastos de inversión en ámbito ambiental.

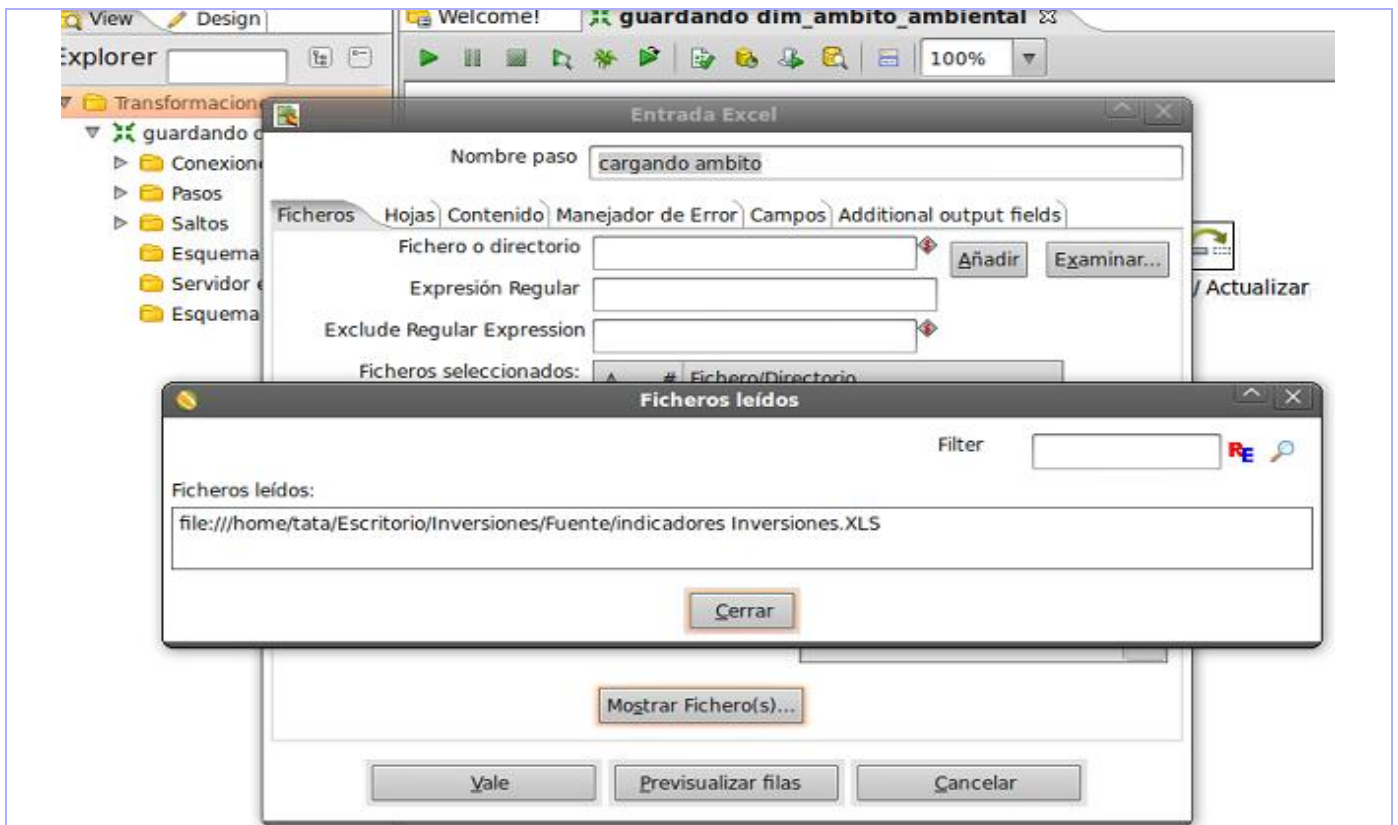
**2.7.4 Casos de Usos Funcionales**

Los casos de uso funcionales son aquellos que responden a los requisitos funcionales los cuales son condiciones que el sistema debe cumplir. A continuación se puede observar la descripción del caso de uso Extraer datos, para mayor detalle se puede consultar el artefacto “Modelo de Caso de Uso del Sistema” del expediente de proyecto:

- CUF1: Extraer datos.

## *Capítulo II: Análisis y diseño del Mercado de Datos Inversiones*

<b>Caso de Uso:</b>	Extraer datos	
<b>Tipo:</b>	Funcional	
<b>Actores:</b>	Administrador de ETL	
<b>Resumen:</b>	El caso de uso inicia cuando el administrador de ETL necesita adicionar nuevos datos al mercado de datos. Se realiza la extracción de datos de la fuente. Finaliza cuando los datos ya se han extraídos.	
<b>Precondiciones:</b>	Las fuentes de los datos tienen que estar disponibles.	
<b>Referencias</b>	RF1	
<b>Prioridad</b>	Crítico	
<b>Flujo Normal de Eventos</b>		
	<b>Acción del Actor</b>	<b>Respuesta del Sistema</b>
	1. Interactúa con la herramienta para realizar la extracción de los datos.	2. El sistema levanta el área de trabajo y le da la posibilidad al administrador de ETL de cargar la transformación necesaria.
	3. Carga la transformación a ejecutar.	
	4. Configura los parámetros de entrada de la transformación.	5. El sistema almacena los datos en el área temporal.
	6. El administrador de ETL previsualiza los datos.	7. El sistema muestra los datos existentes en el área temporal.
	8. Presiona el botón aceptar y cierra la configuración de los parámetros de entrada.	
<b>Prototipo de Interfaz</b>		



### Flujos Alternos

#### Acción del Actor

#### Respuesta del Sistema

6.1 El sistema muestra un mensaje de error.

6.2 Regresar al paso 4 del flujo normal de eventos.

### Prototipo de Interfaz

**Poscondiciones** Los datos quedan disponibles para transformar.

Tabla 2: CUF1: Extraer Datos.

## 2.8 Identificación de dimensiones, hechos y medidas

Luego de realizar el análisis de los modelos existentes en el área de inversiones, se identificaron las dimensiones, hechos y medidas que se exponen a continuación.

### 2.8.1 Dimensiones

1. dim\_entidad: dimensión que contiene las entidades que integran los organismos.



Figura 5: Dimensión entidad

2. dim\_organismo: dimensión que contiene todos los organismos existentes en el país.

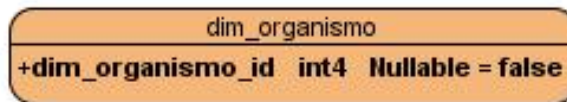


Figura 6: Dimensión organismo

3. `dim_dpa`: dimensión que contiene todas las provincias y municipios que conforman el país.



Figura 7: Dimensión DPA

4. `dim_nae`: dimensión que está compuesta por sección, división y clase.



Figura 8: Dimensión NAE

5. `dim_temporal_trimestre`: dimensión de tiempo que contiene todos los trimestres.

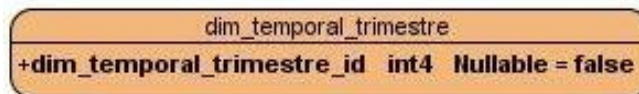


Figura 9: Dimensión temporal trimestre

6. `dim_temporal_anno`: dimensión de tiempo que contiene todos los años.



Figura 10: Dimensión temporal anual

7. `dim_um`: dimensión que especifica la unidad de medida.

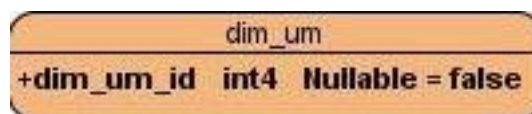


Figura 11: Dimensión unidad de medida

8. dim\_indicadores\_ambito\_ambiental: dimensión que contiene todos los indicadores con los gastos de inversión en actividades de protección del medio ambiente.

dim_indicadores_ambito_ambiental		
<b>+dim_indicadores_ambito_ambiental_id</b>	<b>int4</b>	<b>Nullable = false</b>
indicador_nombre	varchar(50)	Nullable = false
indicador_codigo	varchar(11)	Nullable = false
indicador_descripcion	varchar(150)	Nullable = false
tematica_nombre	varchar(30)	Nullable = false
subtematica_nombre	varchar(50)	Nullable = false
indicador_agregable	varchar(1)	Nullable = false
seguridad_codigo	varchar(1)	Nullable = false
seguridad_nombre	varchar(10)	Nullable = false

Figura 12: Dimensión indicadores de ámbito ambiental

9. dim\_indicadores\_ejecucion\_fisica: dimensión que contiene todos los indicadores que poseen el control del plan de ejecución física de las inversiones.

dim_indicadores_ejecucion_fisica		
<b>+dim_indicadores_ejecucion_fisica_id</b>	<b>int4</b>	<b>Nullable = false</b>
indicador_nombre	varchar(50)	Nullable = false
indicador_descripcion	varchar(150)	Nullable = false
indicador_codigo	varchar(11)	Nullable = false
tematica_nombre	varchar(30)	Nullable = false
subtematica_nombre	varchar(50)	Nullable = false
indicador_agregable	varchar(1)	Nullable = false
seguridad_codigo	varchar(1)	Nullable = false
seguridad_nombre	varchar(10)	Nullable = false

Figura 13: Dimensión indicadores de ejecución física

### 2.8.2 Hechos

hech\_ambito\_ambiental: en esta tabla se encuentra toda la información referente a los gastos de inversión para las actividades protección del medio ambiente.

hech_ambito_ambiental		
<b>+#dim_entidad_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_um_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_organismo_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_dpa_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_nae_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_temporal_trimestre_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_indicadores_ambito_ambiental_id</b>	<b>int4</b>	<b>Nullable = false</b>
gastos_total	float4	Nullable = false

Figura 14: Hecho ámbito ambiental

## Capítulo II: Análisis y diseño del Mercado de Datos Inversiones

hech\_ejecucion\_fisica: en esta tabla se encuentra toda la información referente al control del plan de la ejecución física de las inversiones.

hech_ejecucion_fisica		
<b>+#dim_organismo_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_dpa_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_nae_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_indicadores_ejecucion_fisica_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_temporal_anno_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_entidad_id</b>	<b>int4</b>	<b>Nullable = false</b>
<b>+#dim_um_id</b>	<b>int4</b>	<b>Nullable = false</b>
plan_total	float4	Nullable = false
real_acumulado	float4	Nullable = false
real_acumulado_cuc	float4	Nullable = false
plan_total_cuc	float4	Nullable = false

Figura 15: Hecho ejecución física

### 2.8.3 Medidas

En la siguiente tabla se relacionan las medidas identificadas en el MD y el hecho al que corresponden.

Medida	Hecho al que pertenece
gastos_total	hech_ambito_ambiental
plan_total	hech_ejecucion_fisica
real_acumulado	hech_ejecucion_fisica
plan_total_cuc	hech_ejecucion_fisica
real_acumulado_cuc	hech_ejecucion_fisica

Tabla 3: Medidas del MD Inversiones

### 2.9 Desarrollo de la matriz BUS

En la siguiente tabla se muestra la matriz BUS la cual representa la relación existente entre los hechos y dimensiones del MD Inversiones.

Hechos	Dimensiones								
	1	2	3	4	5	6	7	8	9
hech_ambito_ambiental	X	X	X	X	X		X	X	
hech_ejecucion_fisica	X	X	X	X		X	X		X



Tabla 4: Matriz BUS del MD Inversiones

## 2.10 Modelo de datos

A continuación se muestra el modelo de datos, obtenido a partir de la identificación de dimensiones, hechos y medidas anteriormente expuestos, extraídos a partir de los modelos existentes en el área de inversiones.

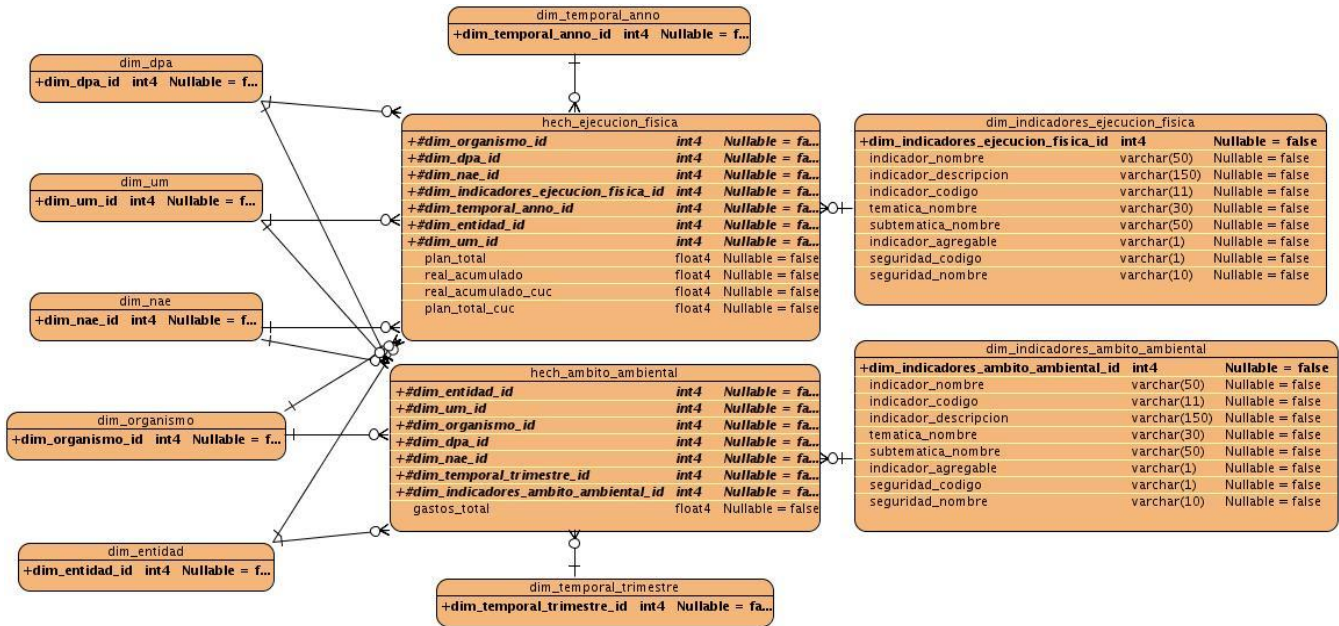


Figura 16: Modelo de datos

## 2.11 Diseño de los procesos de integración de datos

En el diseño de los procesos de integración de datos se realizaron dos actividades importantes para la posterior implementación del subsistema de integración de datos, el diseño de las transformaciones de los procesos de ETL y el perfilado de datos. Las mismas serán abordadas a continuación.

### Diseño de las transformaciones de los procesos de ETL

El diseño de las transformaciones de los procesos de ETL describe los pasos que se realizan para realizar la carga de los hechos y las dimensiones en la base de datos. En la figura siguiente se muestra un ejemplo del flujo de actividades diseñadas para poblar una dimensión.

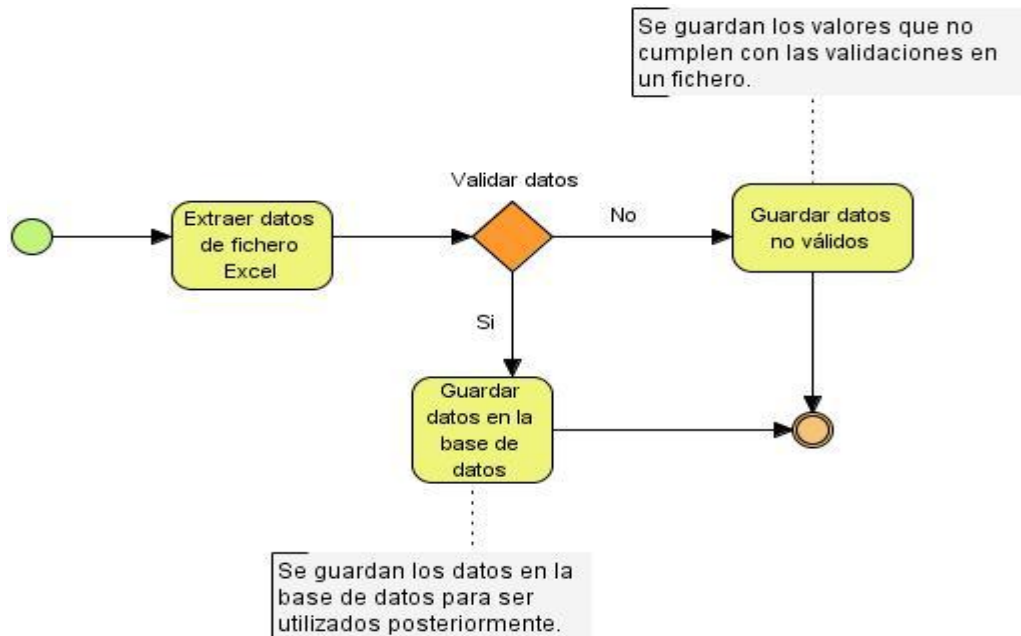


Figura 17: Diseño de la transformación de la dimensión indicadores de ámbito ambiental

### Perfilado de datos

El perfilado de los datos permite lograr un mejor entendimiento de los mismos y verificar la existencia de valores nulos, distintos, duplicados, entre otros; permitiendo definir nuevas reglas del negocio que posteriormente pasan a ser las reglas de transformación para llevar a cabo los procesos de ETL. Para mayor detalle consultar el artefacto “Perfil de Datos”.

A continuación se muestra el resultado del desarrollo del perfilado de datos para el modelo 1006, referente a los gastos de inversión en actividades de protección del medio ambiente. En la tabla resumen se observa el estado en que se encuentra la fuente de datos reconociendo atributos, tipo de datos, total de valores, distintos, nulos, únicos, duplicados, mínimo de caracteres y máximos de caracteres. Dichos resultados fueron obtenidos luego de realizar el perfilado con la herramienta establecida en el Capítulo I Epígrafe 1.6.3. En la Figura 19 se muestra un resultado directo de la utilización de la herramienta.

Tabla	Atributo	Tipo de dato	Total de valores	Distintos	Nulos	Únicos	Duplicados	Min	Max
Modelo 1006	Empresa	String	820	167	0	0	167	5	5
	Organismo	String	820	20	0	0	20	3	3
	Provincia	String	820	15	0	0	15	2	2
	DPA	String	820	89	0	0	89	4	4
	NAE	String	820	35	0	0	35	6	6
	FIL	String	820	49	0	6	55	8	8
	CO1	Número	820	227	0	132	359	7	1

Figura 18: Resultado final del perfilado para el modelo 1006

Standard measures							
	ORG	CO1	DPA	FIL	NAE	PRO	
Row count	820	820	820	820	820	820	820
Null values	7	7	7	7	7	7	7
Empty values	0	0	0	0	0	0	0
Highest value	324	958,3	9901	00000195	189600	99	
Lowest value	105	0	0101	00000001	010211	01	

Figura 19: Análisis de todas las tuplas, modelo 1006

### 2.12 Mapa de navegación

En el MD, que da solución al problema de la investigación planteado, se almacena la información referente a las inversiones que se maneja en la ONE para convertirla en datos que ayuden en la toma de decisiones. En relación con esto se identifica que la información va estar estructurada en el Área de Análisis General SIGOB (A.A.G SIGOB), y dentro de ésta los reportes del MD van a estar localizados en el Área de Análisis Inversiones (A.A Inversiones). Dichos reportes dentro del Área de Análisis van a estar distribuidos en dos Libros de Trabajos (L.T). A continuación se detallan los elementos que componen la estructura de navegación que tendrá la información para su consulta.

**9. Área de Análisis General:** es el área más abarcadora donde se agrupan los temas de análisis, en el caso específico de la solución propuesta constituye el AD SIGOB.

- **Área de Análisis:** es la forma de agrupación de la información según su propósito, en el caso específico de la solución propuesta constituye las secciones en las que está dividido el AD.
- **Libro de Trabajo:** constituye las categorías en los que se agrupan los reportes dentro del área de análisis.
- **Reportes:** es la unidad básica utilizada para mostrar la información almacenada.

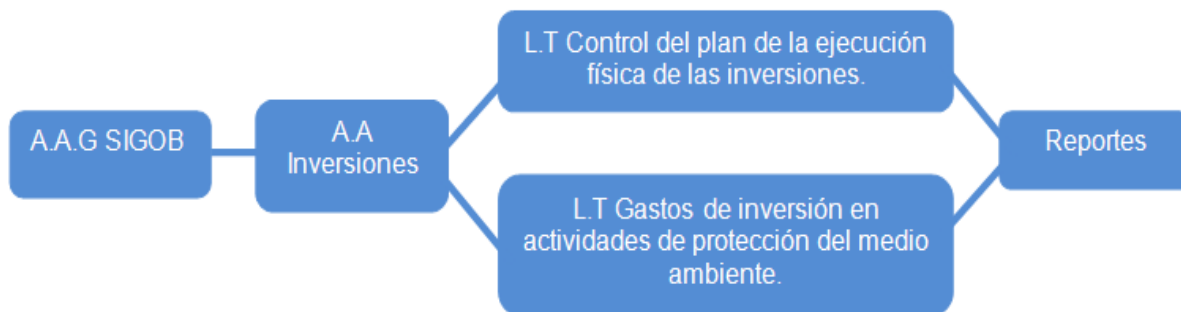


Figura 20: Mapa de navegación

### 2.13 Políticas de respaldo y recuperación

Las políticas de respaldo y recuperación son las medidas y precauciones que se toman en caso de que ocurra algún fallo en el sistema; las que utiliza la solución es sencilla pero a la vez sólida, por ello se miden tres puntos esenciales:

**1. Periodicidad de las salvas:** las salvas se efectúan mensualmente de la información total que posea la base de datos, así lo tiene definido actualmente la organización, certificando en todo momento que exista una copia estricta de la información que está presente en el servidor.

**2. Tablas involucradas:** las tablas que se involucran en la realización son hech\_ambito\_ambiental y hech\_ejecucion\_fisica.

**3. Salvadas (Backups) existentes:** actualmente no existen backups en esta área.

**3.1. Periodicidad de reemplazo de los backups:** se realizan los reemplazos de backups cada un año.

**3.2. Periódicas de las pruebas a los backups:** el estado de los backups se chequea mensualmente, mediante pruebas de integración.

### 2.14 Conclusiones

En este capítulo se realizó el estudio del negocio, donde se obtuvieron las necesidades del cliente y posteriormente se identificaron 16 requisitos de información, los cuales son de gran importancia para la calidad del MD. También fueron identificados 27 requisitos funcionales y 43 requisitos no funcionales. Se definieron 30 reglas de negocio así como el modelo de datos; también se identificaron nueve dimensiones, dos hechos y cinco medidas del MD. Se identificaron además siete casos de uso funcionales y dos casos de uso de información. Se obtuvo la matriz BUS, el mapa de navegación y se definieron las políticas de respaldo y recuperación.

### CAPÍTULO III: IMPLEMENTACIÓN DEL MERCADO DE DATOS INVERSIONES.

#### Introducción

En este capítulo se documenta todo lo referente a la implementación del MD. Se definen los usuarios y privilegios de los mismos así como los niveles de acceso al sistema del MD de cada uno de los roles. Se implementa el subsistema de integración de datos, la estructura de los datos, esquemas y tablas de la BD. Se realiza la implementación del subsistema de visualización de datos donde se diseñan de los cubos OLAP, se implementan los reportes candidatos y se desarrolla la estructura de navegación.

#### 3.1 Usuarios y privilegios

Las políticas de seguridad empleadas en el MD están basadas principalmente en los niveles de acceso al sistema. Esta se maneja mayoritariamente por los roles y permisos que poseen los usuarios que interactúan con el sistema.

El sistema permitirá la transmisión de información por canales cifrados usando el protocolo HTTPS. Las contraseñas de los usuarios no deben ser almacenadas en texto plano, esta información debe ser privada y específica de cada uno de ellos, de manera que nadie pueda reemplazar la identidad de un usuario en el sistema.

#### Roles y permisos

Roles	Permisos
Especialista	Opera con los esquemas y las tablas, posee permisos de lectura, en este caso, la acción llamada SELECT en la BD.
Administrador del sistema	Controla la BD, su configuración y administración, posee todos los permisos en la BD.
Administrador de ETL	Visualiza y modifica los datos de la BD, por lo que posee permisos de lectura y escritura, las acciones nombradas SELECT, INSERT, UPDATE, DELETE en la BD.

Tabla 5: Roles y permisos

## Capítulo III: Implementación del Mercado de Datos Inversiones

---

De acuerdo a la estructura de la aplicación se especifica el nivel de acceso al sistema del AD de cada uno de los roles:

Elemento de la aplicación	Roles con acceso
A.A.G SIGOB	<ul style="list-style-type: none"><li>• Administrador del Sistema</li><li>• Especialista</li></ul>
A.A Inversiones	<ul style="list-style-type: none"><li>• Administrador del Sistema</li><li>• Especialista</li><li>• Administrador de ETL</li></ul>

Tabla 6: Nivel de acceso al sistema del AD

Además de los niveles de acceso a la aplicación, antes mencionados, la plataforma Pentaho BI lleva incluida su propia seguridad de acceso a los datos, la cual se basa en una infraestructura proporcionada por el sistema de seguridad Acegi, el cual es un framework de seguridad de Java. Este se divide en cuatro áreas principales, las cuales son:

- Seguridad de acceso a datos de objetos: incluye usuarios, contraseñas, autorizaciones permitidas, recursos web y protección a datos.
- Autenticación: tiene que ver con el procesamiento de información interactiva de inicio de sesión (ejemplo: nombre de usuario y contraseña) comparándola con la información recuperada del AD de seguridad.
- Autorización de recursos web (URL): brinda protección a las URL para responder a cada usuario si pueden o no acceder a una determinada página. Esto es decidido por el administrador de recursos web, el cual le brinda a cada usuario autenticado un permiso de seguridad, delimitando las páginas a las que tiene acceso y a las que no.
- Autorización a objetos del dominio: en el sistema los únicos objetos del dominio protegidos por la plataforma son los objetos de repositorio otorgados al usuario autenticado. Es responsabilidad de los objetos del dominio autorizar las operaciones solicitadas por este. (21)

### **3.2 Implementación del subsistema de integración de datos**

El subsistema de integración de datos permite la carga de los registros fuentes verificando la no existencia de valores nulos, para posteriormente hacer la búsqueda en la BD de las dimensiones relacionadas con los hechos y consecutivamente cargar los mismos.

#### **3.2.1 Implementación del modelo de datos**

##### **Estructura de los datos**

La estructura de los datos es una colección de los mismos donde su organización se caracteriza por las funciones de acceso que se usan para acceder y almacenar a elementos individuales de datos; es una forma de organizar un conjunto de datos con el objetivo de facilitar la manipulación de los mismos. Por debajo de su estructura hay un nivel donde se sitúan los discos, particiones, archivos y espacios de tablas. La buena manipulación y utilización de dichos elementos son de gran importancia para el éxito y la calidad de la solución.

##### **Esquemas**

Los esquemas en la BD son una forma de organizar los datos en la misma, estos pueden tener funciones, operadores y tipos de datos. Los usuarios pueden acceder a ellos siempre que tengan los permisos necesarios.

Para el desarrollo del sistema propuesto se definieron dos esquemas:

- dimensiones: esquema que contiene las tablas de las dimensiones que son comunes con la BD general del AD SIGOB.
- mart\_inversiones: esquema que contiene todas las tablas de hechos y dimensiones propias del MD Inversiones.

##### **Tablas**

La BD de la solución propuesta cuenta con once tablas en total, de ellas nueve dimensiones y dos hechos distribuidas en los esquemas anteriormente mencionados.

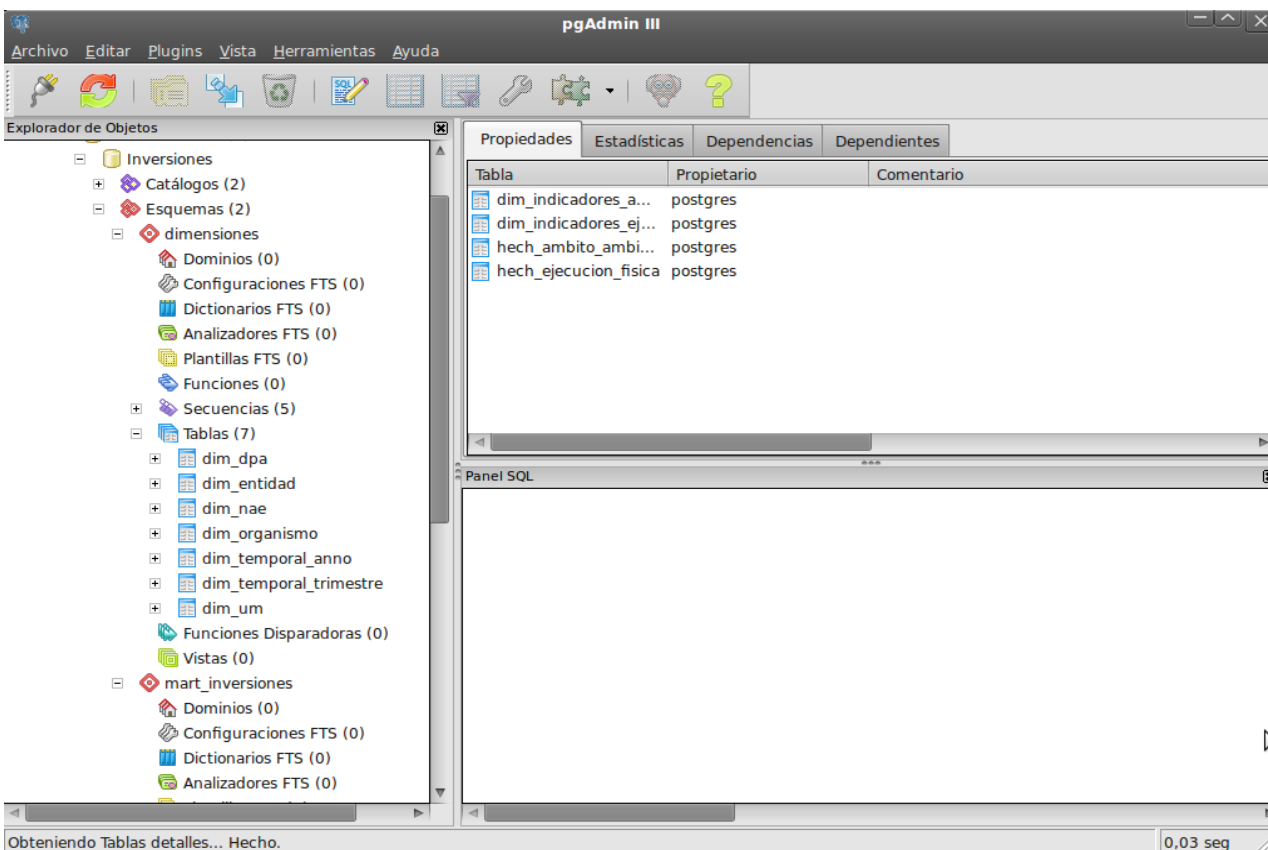


Figura 21: Estructura física de la BD

### 3.2.2 Implementación de los flujos de transformación

Después de finalizada la extracción de los datos, el sistema se encuentra listo para la etapa de transformación. Las transformaciones son el elemento más importante en el transcurso de la implementación del subsistema de integración de datos, las mismas están compuestas por pasos y estos a su vez se encuentran unidos a través de saltos, dichos pasos son los elementos más pequeños dentro de las transformaciones. Una vez concluidas las transformaciones, los datos estarán listos para ser cargados en la BD.

Para implementar el subsistema de integración de datos del MD Inversiones se realizan cuatro transformaciones, dos de ellas para cargar los indicadores y dos para cargar los hechos, quedando de la siguiente manera:

- **guardando\_dim\_ambito\_ambiental:** esta transformación extrae desde un fichero Excel los indicadores establecidos para el modelo 1006, referente a gastos de inversión en actividades de



protección del medio ambiente, para la tabla correspondiente. Realizando una previa validación para que los datos no se carguen nulos.

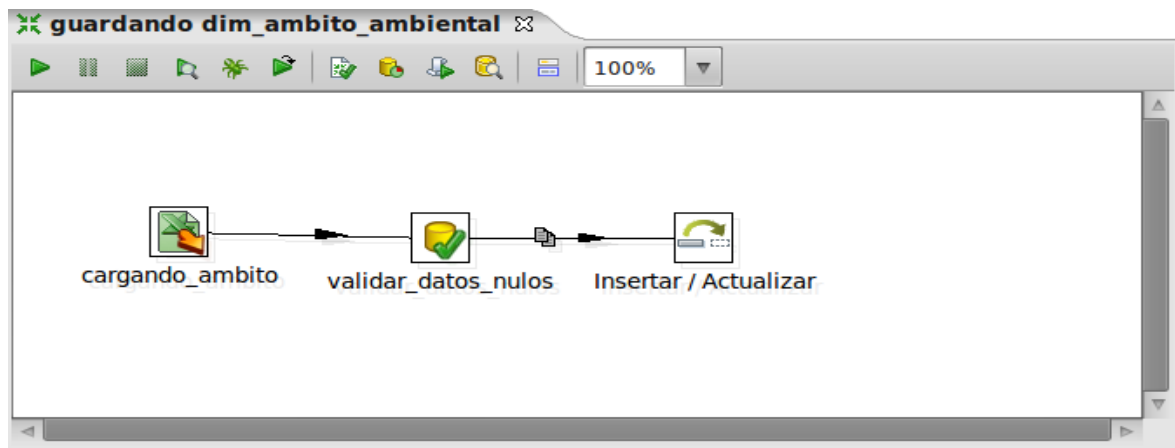


Figura 22: Transformación de la dimensión indicadores de ámbito ambiental

- **guardando\_dim\_ejecucion\_fisica:** esta transformación extrae desde un fichero Excel los indicadores establecidos para el modelo 5702, referente al control del plan de la ejecución física de las inversiones, para la tabla correspondiente. Realizando una previa validación para que los datos no se carguen nulos.



Figura 23: Transformación de la dimensión indicadores de ejecución física

- **cargando\_hecho\_ambito\_ambiental:** la carga del hecho ámbito ambiental se inicia con la entrada de un fichero Excel, posteriormente se valida que no existan valores nulos, luego se pasa a verificar la existencia de cada dimensión para no tener valores nulos por la no existencia de una dimensión relacionada con dicho hecho. Luego se pasa a ordenar las filas, agruparlas y finalmente pueden ser insertadas o actualizadas en la BD.

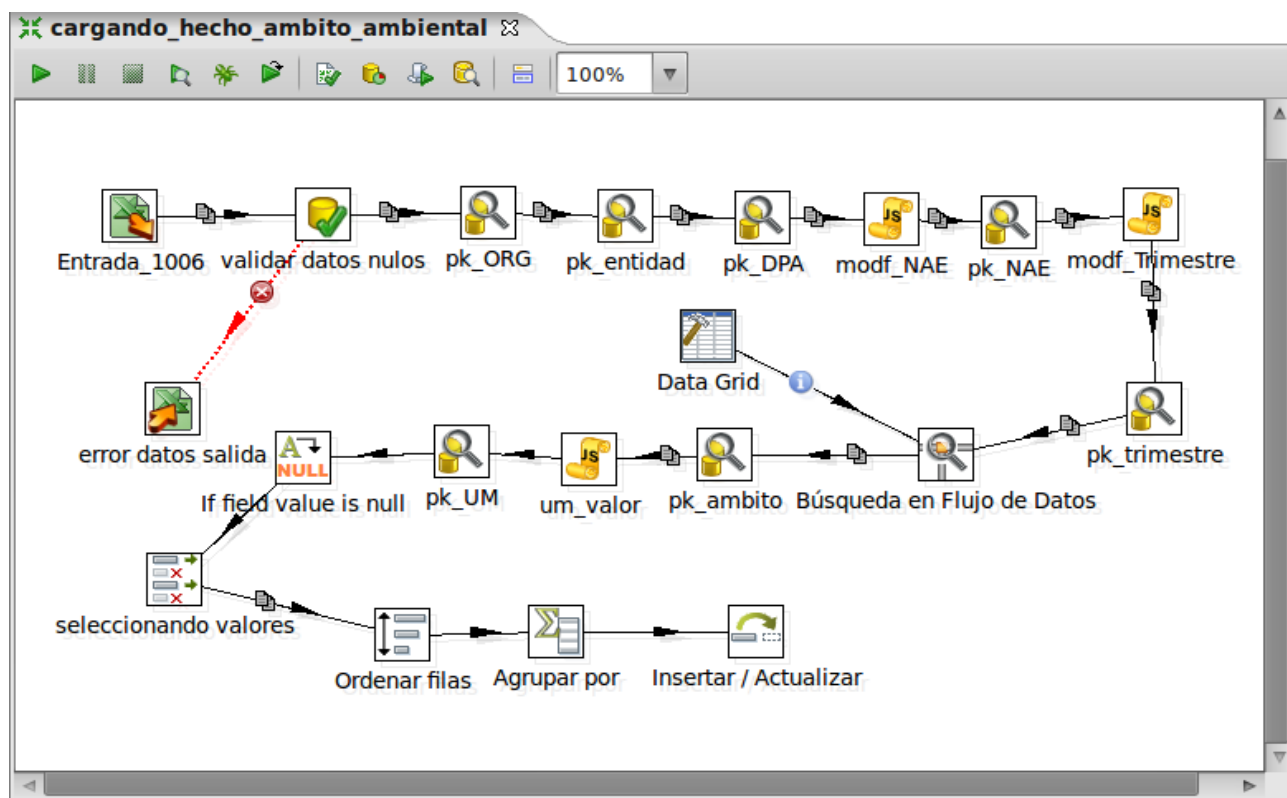


Figura 24: Transformación del hecho ámbito ambiental

- **cargando\_hecho\_ejecucion\_fisica**: la carga del hecho ejecución física se realiza con varias entradas de ficheros Excel, donde se tiene un archivo por cada indicador existente el modelo 5702. Posteriormente se valida que no existan valores nulos, luego se pasa a verificar la existencia de cada dimensión para no tener valores nulos por la no existencia de una dimensión relacionada con dicho hecho. Luego se pasa a ordenar las filas, agruparlas y finalmente pueden ser insertadas o actualizadas en la BD.

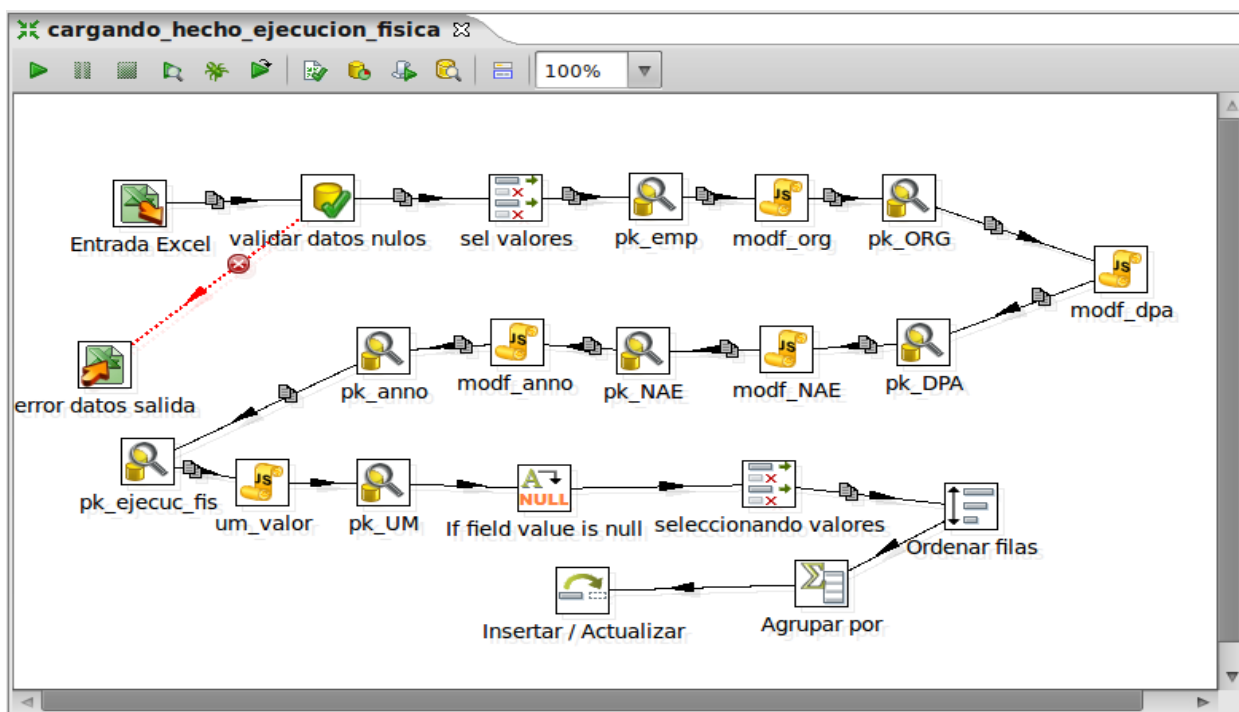


Figura 25: Transformación del hecho ejecución física

Las transformaciones realizadas para posibilitar la carga de los hechos garantizan que los datos cargados del registro fuente, en este caso archivos Excel, no pasen a la carga de hechos con valores nulos y que los id de las dimensiones sean válidos. Se realizan las búsquedas en la BD de las dimensiones relacionadas con el hecho y finalmente se carga el hecho.

### 3.2.3 Implementación del trabajo

Un trabajo o job se utiliza para realizar una acción determinada. En este caso es usado para realizar pasos específicos que son distintos a los disponibles en las transformaciones, exactamente para ejecutar varias transformaciones que fueron diseñadas. Mediante los trabajos se define el horario y frecuencia de la carga, así como el orden en que van a ser ejecutadas las transformaciones para poder realizar exitosamente la carga de los datos. En el caso de la solución propuesta el trabajo es programado para ejecutarse anualmente.

En el trabajo diseñado para ejecutar las transformaciones del área de inversiones primeramente se chequea la conexión, en caso de no estar disponible se presentará una notificación de error si no se insertan o actualizan las dimensiones ámbito ambiental y ejecución física. Posteriormente se ejecuta la transformación destinada a cargar los datos del modelo 1006 y luego se ejecuta la carga de los datos

referentes al modelo 5702 distribuidos en un fichero Excel por indicador. Finalmente se obtiene una notificación de confirmación de la culminación de la carga.

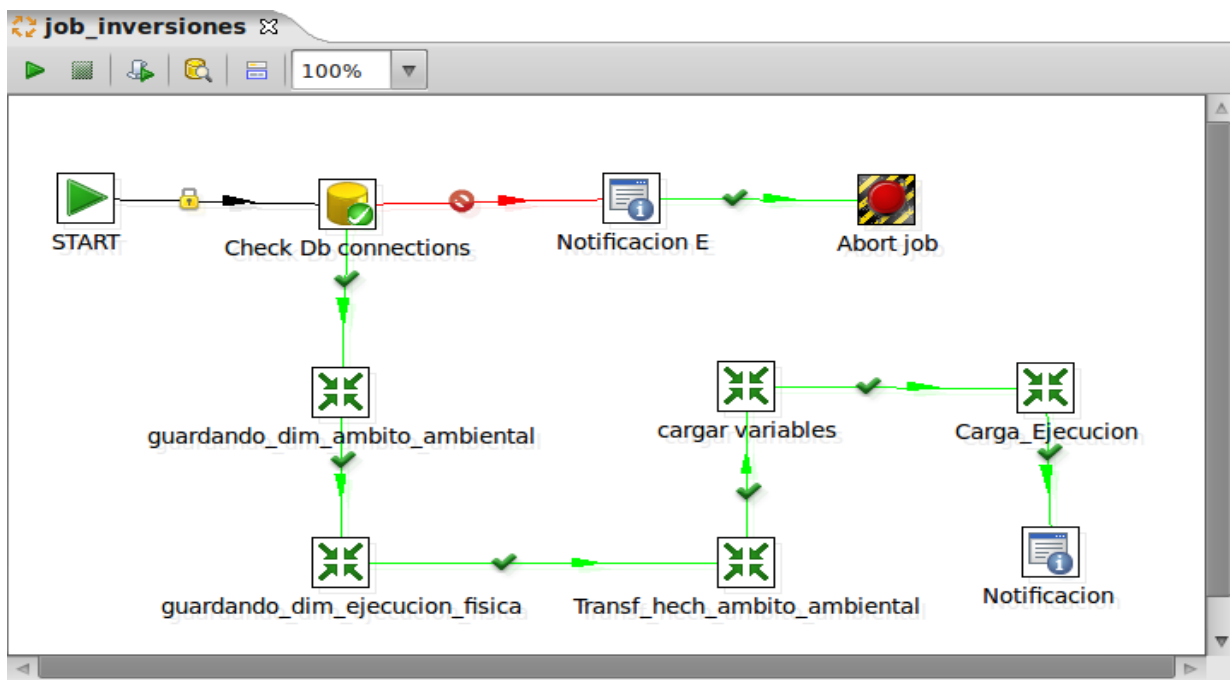


Figura 26: Implementación del trabajo

### 3.3 Implementación del subsistema de visualización de datos

La implementación del subsistema de visualización de datos, es el proceso de analizar los datos acumulados en una entidad y extraer conocimiento de ellos, incluyendo bases de datos, información de la cadena de suministro, ventas personales y cualquier actividad de marketing o fuente de información relevante para la empresa. Esto con el objetivo de ayudar en la toma de decisiones del negocio incrementando la efectividad de la entidad. (23)

Para lograr una efectiva presentación de la capa de visualización uno de los elementos fundamentales que se deben desarrollar son los cubos OLAP, donde se organiza la información de forma estructurada.

#### 3.3.1 Cubos OLAP

Luego de haberse desarrollado las transformaciones necesarias en los hechos del MD Inversiones, se realiza el diseño de los cubos OLAP. En la solución propuesta se desarrollaron dos cubos multidimensionales, los cuales son: hech\_ambito\_ambiental y hech\_ejecucion\_fisica. Estos fueron desarrollados mediante la utilización de la herramienta Pentaho Schema Workbench anteriormente documentada en el Capítulo I. Los cubos OLAP fueron definidos de acuerdo a la estructura de las tablas

de hechos del MD, logrando una buena organización estructural de los datos. A continuación se muestra el diseño de los cubos OLAP:



Figura 27: Cubo hech\_ambito\_ambiental

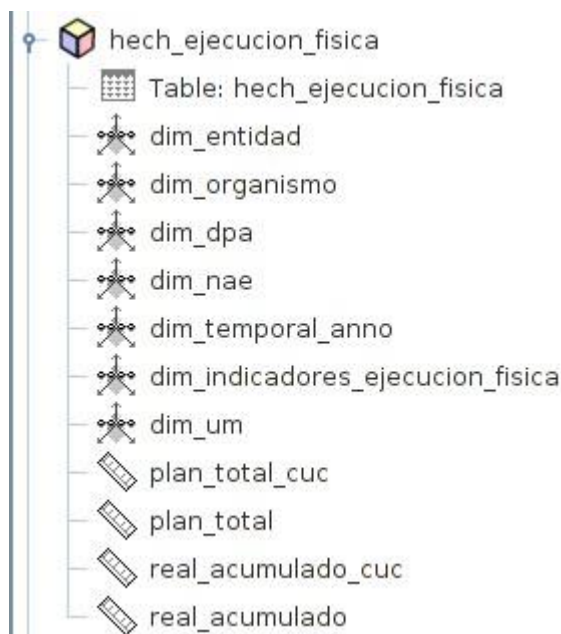


Figura 28: Cubo hech\_ejecucion\_fisica

### 3.3.2 Reportes candidatos

Los reportes candidatos son la información que el cliente desea que se muestre como finalidad del producto. Los mismos fueron identificados luego de realizar un análisis de los modelos 1006 y 5702, donde se recoge toda la información referente al área de inversiones de la ONE. Estos fueron ubicados de acuerdo a los elementos de la estructura de navegación anteriormente diseñada, la cual quedó de la siguiente forma:

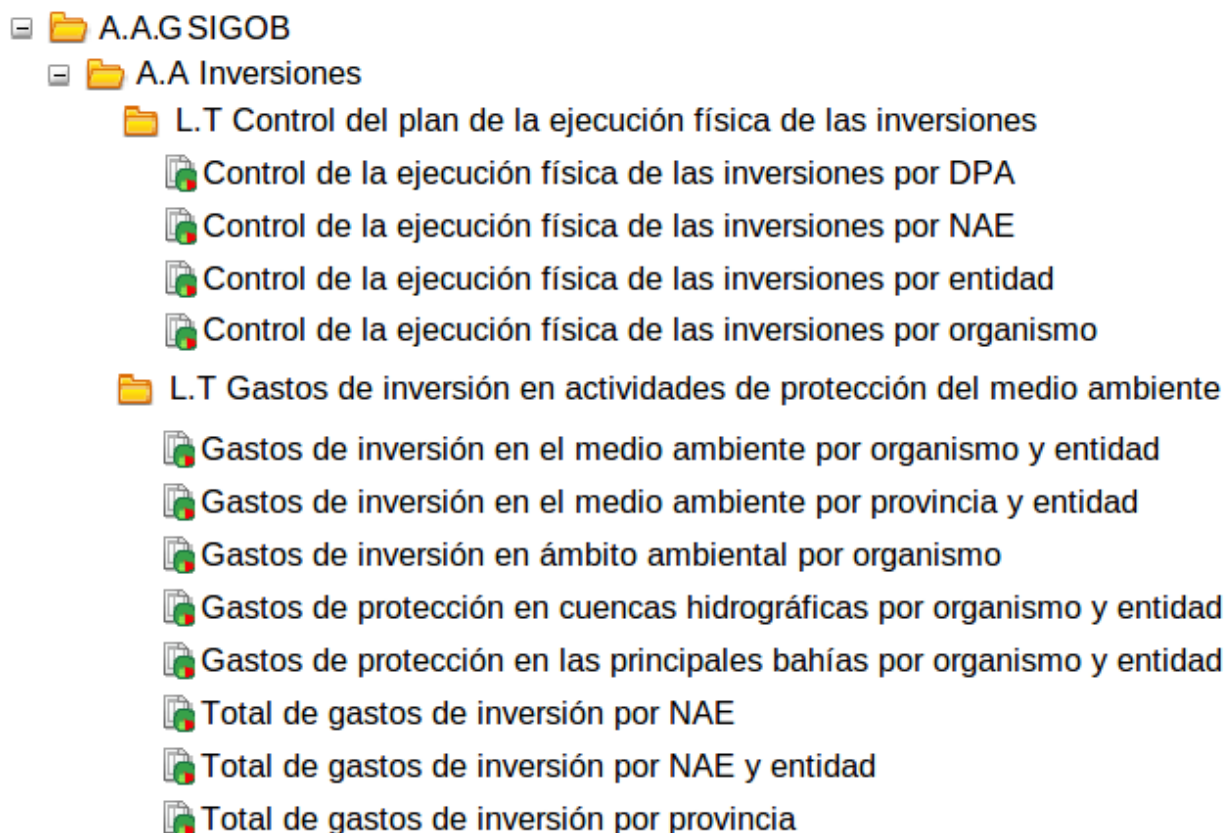


Figura 29: Estructura de navegación

A continuación se mencionan los reportes candidatos que fueron implementados, distribuidos en dos Libros de Trabajo. Para mayor profundización los mismos se encuentran detallados en los artefactos “Reportes candidatos” y “Arquitectura de información”.

### **L.T Gastos de inversión en actividades de protección del medio ambiente**

- Gastos de inversión en el medio ambiente por organismo y entidad
- Gastos de inversión en el medio ambiente por provincia y entidad
- Gastos de inversión en ámbito ambiental por organismo
- Gastos de protección en cuencas hidrográficas por organismo y entidad
- Gastos de protección en las principales bahías por organismo y entidad
- Total de gastos de inversión por NAE
- Total de gastos de inversión por NAE y entidad
- Total de gastos de inversión por provincia

### **L.T Control del plan de la ejecución física de las inversiones**

- Control de la ejecución física de las inversiones por DPA
- Control de la ejecución física de las inversiones por NAE
- Control de la ejecución física de las inversiones por entidad
- Control de la ejecución física de las inversiones por organismo

### **3.4 Conclusiones**

Este capítulo abarca la implementación de los subsistemas de integración y visualización de datos. Para realizar la implementación del subsistema de integración se desarrolló el modelo de datos obteniendo 11 tablas distribuidas en dos esquemas. En el desarrollo de los procesos de ETL se realizaron cuatro transformaciones de ellas dos destinadas a la carga de dimensiones y las restantes a la carga de los hechos; fue implementado el trabajo necesario para la ejecución de las transformaciones. Para la implementación del subsistema de visualización fueron diseñados los cubos OLAP, donde se obtuvieron dos cubos multidimensionales; implementados 12 reportes candidatos y establecida la estructura de navegación.

### **CAPÍTULO IV: VALIDACIÓN DEL MERCADO DE DATOS INVERSIONES**

#### **Introducción**

En este capítulo se realizan las pruebas y validaciones al MD Inversiones para garantizar la calidad, eficiencia y obtención de los resultados esperados. Se describen los distintos tipos de pruebas que fueron aplicadas y sus resultados, así como la validación de la solución propuesta.

#### **4.1 Calidad de software**

Un producto de software pierde sentido si no cumple las expectativas del cliente, quien desea que el mismo sea económico, fiable y funcione eficientemente. Por tanto los desarrolladores de software necesitan que los procesos de desarrollo y los productos terminados sean evaluados para tener una garantía de que el software tiene alta calidad. Por ese motivo surgen modelos, técnicas, patrones y normas de calidad que establecen cómo hacer un software con calidad desde los procesos que lo inician hasta el producto terminado. (24)

La calidad de software se pudiera definir como la capacidad que tiene un software para adherirse a un conjunto de requisitos, especificaciones, estándares, normas o procedimientos que deben cumplirse durante el proceso de desarrollo, y que permitan verificar que el producto cumple con los requisitos explícitos establecidos y las expectativas del cliente. (24)

Para que un producto tenga calidad debe cumplir principalmente con los requisitos propuestos por el cliente y con algunas características como:

- **Mantenibilidad:** capacidad del producto de software de ser modificado. Las modificaciones pueden incluir las correcciones, mejoras o adaptaciones del software a cambios en el ambiente, así como en los requisitos y las especificaciones funcionales.
- **Funcionalidad:** capacidad del software para proporcionar funciones que satisfacen las necesidades declaradas e implícitas cuando el software se usa bajo las condiciones especificadas.
- **Portabilidad:** capacidad de producto de software de ser transferido de un ambiente a otro.
- **Confiabilidad:** capacidad del producto de software para mantener un nivel de ejecución especificado cuando se usa bajo las condiciones especificadas.
- **Eficiencia:** capacidad del producto de software para proporcionar una ejecución o desempeño apropiado, en relación con la cantidad de recursos utilizados usados, bajo condiciones establecidas.



- **Usabilidad:** capacidad del producto de software de ser comprendido, aprendido, utilizado y de ser atractivo para el usuario, cuando se utilice bajo las condiciones especificadas. (25)

Estas características se miden mediante la aplicación de diversas pruebas dirigidas a cada una de estas características. La etapa de pruebas es una de las más importante donde se debe aplicar la calidad de software aunque esta es de gran importancia durante todo el ciclo de desarrollo de un producto de software.

### **4.2 Pruebas de calidad de software**

Para determinar la calidad del producto se realizan distintas pruebas de software. Estas pueden aplicarse de forma dirigida a componentes de software o al sistema de software en conjunto, con la finalidad de medir el grado en que el producto cumple con los requerimientos definidos. El proceso de pruebas comienza con la planificación de las mismas, posteriormente la ejecución y el control y como paso final la evaluación de las pruebas.

Existen distintos tipo de pruebas como por ejemplo pruebas de volumen, integración, carga, caja negra, stress, entre otras. Las cuales se clasifican de acuerdo a las características de calidad anteriormente mencionadas. También existen modelos que ayudan y rigen el proceso de pruebas. Durante el desarrollo de la solución propuesta en la presente investigación se utilizó el Modelo V, el cual utiliza el centro DATEC con el fin de que el producto posea calidad y ayude durante el proceso de pruebas.

#### **4.2.1 Modelo V**

El Modelo V es una variación del modelo en cascada que demuestra cómo se relacionan las actividades de prueba con las de análisis y diseño. En la siguiente figura se puede ver que la punta de la V es la codificación, con el análisis y el diseño a la izquierda y la prueba y el mantenimiento a la derecha. El Modelo V sugiere que la prueba unitaria y de integración sea hecha también para verificar el diseño, verificando que todos los aspectos del diseño del programa se han implementado correctamente en el código, la prueba del sistema debe verificar el diseño del sistema, asegurando que todos los aspectos del diseño están correctamente implementados. La prueba de aceptación que es dirigida por el cliente, valida los requerimientos. La vinculación entre la parte izquierda y derecha del Modelo V implica que si se encuentran problemas durante la verificación y la validación entonces el lado izquierdo de la V puede ser ejecutado nuevamente para solucionar el problema y mejorar los requerimientos, el diseño y el código antes de retomar las pruebas del lado derecho. (26)

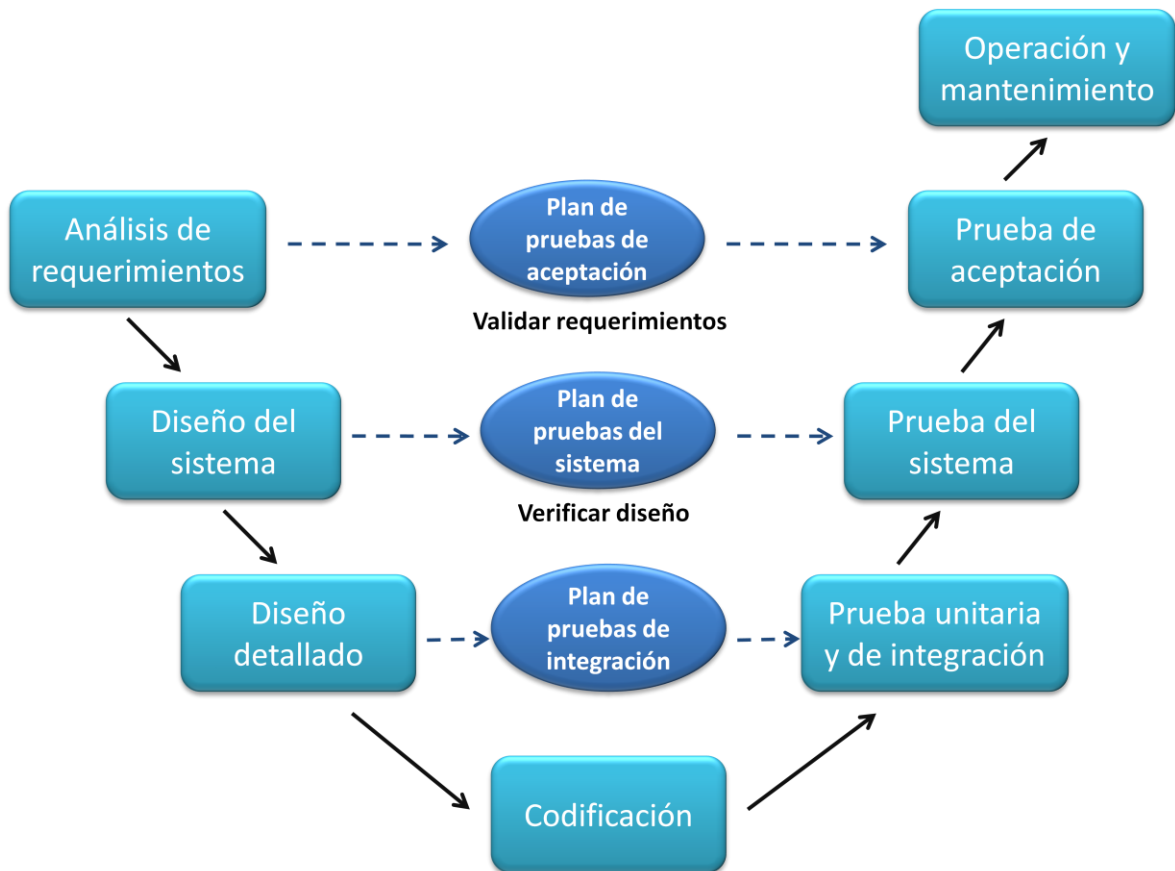


Figura 30: Modelo V

### 4.2.2 Tipos de pruebas

Utilizando el Modelo V fueron aplicadas las siguientes pruebas, por los especialistas del centro:

- **Prueba unitaria:** se prueban los componentes individuales (subprogramas o procedimientos) de un programa. El propósito es descubrir discrepancias entre la especificación de la interface de los módulos y su comportamiento real. Son diseñadas y ejecutadas por el desarrollador una vez terminado el desarrollo de cada componente. (26)
- **Prueba de integración:** se realizan para determinar la integración de los componentes dentro de un sistema y evaluar su correcta interface, funcionalidad y desempeño. Son diseñadas y ejecutadas por el desarrollador cuando la solución está completa junto a los especialistas del centro. (25)
- **Prueba de sistema:** se realizan para determinar el correcto funcionamiento de un sistema y su cumplimiento contra las especificaciones del producto. son ejecutadas por el desarrollador o equipo de pruebas en un ambiente controlado. (25)

Luego de ser aplicadas las pruebas por el centro se realiza la entrega de la solución propuesta a calidad UCI, donde se le realizan una serie de pruebas como son de aceptación, integración, volumen y finalmente de liberación.

Para realizar dichas pruebas en el MD Inversiones fueron utilizados diversos métodos y herramientas como son los diseños de casos de pruebas basados en casos de uso de información y las listas de chequeo para validar los documentos entregables.

### **4.2.3 Casos de prueba**

Los casos de prueba consisten en permitir observar posibles síntomas de defectos. Se puede definir un caso de prueba como “el conjunto de entradas, condiciones de ejecución y resultados esperados desarrollados para un objetivo particular como, por ejemplo, ejercitar un camino concreto de un programa o verificar el cumplimiento de un determinado requisito”. (27)

Toda la capacidad de detección de defectos de las pruebas se basa en la consideración de cualquier tipo de discrepancia entre la salida obtenida y la salida esperada; identificadas con el propósito de hacer una evaluación de aspectos particulares de un elemento objeto de prueba. (28) Por tales motivos resulta imprescindible que el diseño de casos de prueba se apoye en la selección de algunas entradas o situaciones que ejercitar dentro de todo el universo posible.

### **Cómo se ejecutan en un proyecto de software y en qué etapa.**

Una vez que han aprobado todos los CU correspondientes a la iteración se debe establecer cuál es la estrategia de pruebas a seguir. Esta disciplina amerita un mayor esfuerzo en la fase de construcción ya que es el momento en el que se valida el producto de software. Antes, durante y después de esta fase se realizan actividades de verificación para saber si se está construyendo el producto correctamente.

El objetivo de las pruebas es evaluar la calidad del producto apoyándose en un conjunto de buenas prácticas entre las que se destacan:

- Validar que los requerimientos son implementados debidamente.
- Comprobar que el producto de software trabaja según el diseño.

Los casos de prueba son esenciales para todas las actividades de pruebas porque son la base para diseñar y ejecutar los procedimientos de pruebas. Reflejan trazabilidad con los CU, ya que estos muestran una secuencia ordenada de eventos, al describir flujos básicos, flujos alternos, precondiciones y post-condiciones. Si los casos de prueba no son correctos, la calidad del sistema se pone en duda y las

pruebas dejan de ser confiables. El método de diseño de casos de pruebas se realiza en la práctica generando al menos un diseño de caso de prueba por cada caso de uso de información, a partir de la descripción de su interacción.

En el caso de la solución propuesta se han diseñado dos casos de prueba, uno por cada CUI, los mismos se destacan a continuación:

- Caso de prueba Consultar gastos de inversión en ámbito ambiental
- Caso de prueba Consultar plan de ejecución física de inversiones

En el caso de uso Consultar gastos de inversión en ámbito ambiental se analiza la información referente a los gastos de inversión en actividades de protección del aire y el clima, reducción del ruido y las vibraciones, gestión de las aguas, residuos, protección y rehabilitación de los suelos, protección de la biodiversidad y los paisajes, protección contra las radiaciones, investigación y desarrollo, otras actividades de protección del medio ambiente, cuencas hidrográficas de interés nacional y principales bahías. En el cual se tienen en cuenta las secciones (perspectivas de análisis y reportes candidatos). De dichas secesiones inicialmente se tienen en cuenta los distintos escenarios, la descripción de la funcionalidad y el flujo central. Posteriormente cada escenario es analizado desde variables que aplican (perfiles de análisis e indicadores a medir), respuestas del sistema y resultado de la prueba donde se escribe el resultado que se obtiene al realizar la prueba. En el Anexo 1 se muestra como ejemplo el Caso de prueba Consultar gastos de inversión en ámbito ambiental y para mayor profundización de los mismos se pueden consultar los artefactos “Caso de prueba Consultar gastos de inversión en ámbito ambiental” y “Caso de prueba Consultar plan de ejecución física de inversiones” que se encuentran en el expediente de proyecto adjunto a esta investigación.

#### **4.2.4 Listas de chequeo**

Las listas de chequeo están basadas en la identificación de las técnicas de prueba para evaluar cada sub-característica de las características de calidad. Una lista de chequeo es un formulario de preguntas, las cuales dependen del objetivo para el cual son usadas. (28)

La lista de chequeo que se propone tiene el objetivo de evaluar la calidad de los artefactos que se generan en la etapa de análisis de la solución y la aplicación. Esta tiene gran importancia ya que valida que los requisitos previamente definidos especifican realmente que el sistema cumple con lo que el cliente desea y necesita; además de cubrir otros aspectos aplicables en los temas de análisis.

## Capítulo IV: Validaciones del Mercado de Datos Inversiones

---

Esta lista de chequeo posee varios indicadores a evaluar, los cuales se encuentran distribuidos en tres secciones fundamentales:

- **Estructura del documento:** abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- **Indicadores definidos por la etapa:** abarca todos los indicadores a evaluar de la aplicación.
- **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía, redacción y demás.

Para su mejor comprensión a continuación se explican algunos elementos importantes que forman parte de la estructura de la lista de chequeo.

- **Peso:** define si el indicador a evaluar es crítico o no.
- **Indicadores a evaluar:** son los indicadores que se evalúan al artefacto.
- **Evaluación:** es la forma de evaluar el indicador en cuestión. El mismo se evalúa de 1 en caso de que exista alguna dificultad sobre el indicador y 0 en caso de que el indicador revisado no presente problemas.
- **No procede:** se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- **Cantidad de elementos afectados:** especifica la cantidad de errores encontrados sobre el mismo indicador.
- **Comentario:** especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo.

Para la evaluación de la solución y los artefactos del MD Inversiones se aplicó una lista de chequeo por parte de los especialistas del centro. La misma posee 14 indicadores, de los cuales ocho poseen un peso crítico. Esta lista de chequeo se muestra en el Anexo 2 y para mayor profundización se puede consultar el artefacto “Lista de chequeo MD Inversiones” que se encuentra en el expediente de proyecto adjunto a esta investigación.

### 4.3 Resultado de las pruebas realizadas

Luego de aplicadas las pruebas mediante la lista de chequeo y los casos de prueba siguiendo el Modelo V, se obtuvieron los siguientes resultados:

- **Pruebas unitaria y de integración:** Los especialistas del centro le realizaron pruebas unitarias y posteriormente de integración a la solución propuesta. Las mismas arrojaron como resultado una serie de problemas no críticos que fueron resueltos inmediatamente.
- **Lista de chequeo:** Fue aplicada la lista de chequeo para evaluar la calidad de los artefactos y de la aplicación por los especialistas del centro, obteniéndose 19 NC que fueron resultas de forma inmediata.
- **Pruebas de sistema:** Se le aplicaron pruebas al sistema, mediante los casos de prueba, por el centro de calidad UCI (CALISOFT) arrojando 2 NC en una primera iteración, 4 NC en una segunda iteración y el producto fue liberado luego de una tercera iteración.
- **Pruebas de aceptación:** La aplicación informática fue aceptada por el cliente, evidenciándose que la solución propuesta cumple con todos los aspectos deseados por el mismo.

### **4.4 Conclusiones**

En el capítulo se abarca todo lo referente a los temas de calidad de software y los tipos de pruebas que se aplican para garantizar la misma. Se concretó que el modelo a seguir para garantizar la calidad de la solución a la presente investigación es el Modelo V por el cual se rige el centro DATEC. Siguiendo este modelo fueron utilizados como métodos para realizar las pruebas de software: la lista de chequeo y los casos de prueba orientada a casos de uso de información. Quedó validado el MD Inversiones.

### **CONCLUSIONES GENERALES**

Al concluir el desarrollo del MD Inversiones se puede plantear que fueron cumplidos todos los objetivos establecidos, por lo que se arribó a las siguientes conclusiones:

- Se refinó el análisis y diseño del MD Inversiones cumpliendo con todo lo propuesto en el negocio, posibilitando la posterior implementación.
- Se implementaron los subsistemas de integración y visualización de datos referentes al área de inversiones, propiciando la consistencia de los datos y la disponibilidad de la información para su consulta por parte de los usuarios, ayudando así en el proceso de toma de decisiones.
- Se validó la calidad del MD Inversiones mediante la aplicación de pruebas, logrando la aceptación del cliente.

### **RECOMENDACIONES**

- Que se tome como referencia el contenido de esta investigación para próximos sistemas de mercados y AD que se desarrollen.
- Integrar el MD Inversiones al Sistema de Información de Gestión Estadística (SIGE) para lograr la seguridad de la fuente de datos así como la integridad.



### REFERENCIAS BIBLIOGRÁFICAS

1. Orellanos, Norah Velazco. Estadísticas de Inversión Extranjera Directa en los Países de la Comunidad Andina.
2. Secretaria de Economía . [En línea]  
<http://www.economia.gob.mx/swb/work/models/economia/Resource/516/1/images/EstadInverMexicoUE.pdf>.
3. **B., Eduardo Abedrapo.** *Principales aspectos del Sistema Nacional de Inversiones de Chile.*
4. ONE. [En línea] <http://www.one.cu>.
5. **Galicia, Business Intelligence.** SINNEXUS. [En línea] 14 de enero de 2010.  
[http://www.sinnexus.com/business\\_intelligence/datawarehouse.aspx](http://www.sinnexus.com/business_intelligence/datawarehouse.aspx).
6. **Kimball, R.** *The Data Warehouse ETL Toolkit. Practical Techniques for extracting, cleaning, conforming, and delivering data.* s.l. : WILEY PUBLISHING, INC., 2002.
7. **Ruiz, Tomás Martínez, Navarro, Almudena Quevedo y Visuete, Michael.** *UNIVERSIDAD DE CASTILLA-LA MANCHA ESCUELA SUPERIOR DE INFORMÁTICA, modelos avanzados de bases de datos.* Quito : s.n., Marzo, 2006.
8. **Vázquez, L. C.** Tecnología e Internet. [En línea] 27 de octubre de 2009. <http://gestiopolis.com>.
9. **Herrera, C.** Adictos al trabajo. [En línea] 30 de octubre de 2007.  
<http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse4>.
10. **Kimball, Ralph.** *El Juego de Herramientas del Almacén de Datos.* 1996.
11. **Nader, Ing. Javier.** *Sistema de apoyo gerencial universitario. Tesis de Magister en Ingeniería de Software.* 2003.
12. **Beltrán, Carlos Patricio López.** *Análisis, Diseño e implementación de un Data Mart para la Dirección Financiera y Recursos Humanos de la Escuela Politécnica del Ejército para una toma de decisión efectiva.* SANGOLQUI : s.n., 10 de abril de 2007.
13. Sinnexus. [En línea] [http://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx).
14. **Ing. Alberto Límia Navarro<sup>1</sup>, Ing. Mabel Medina Rodriguez<sup>2</sup>, Ing. Daulemys Rigo Portillo<sup>3</sup>, Ing. Madelys Cuesta Villa<sup>4</sup>, Ing. Yamila Mateu Romero<sup>5</sup>, Ing. Julio Ernesto Ortiz<sup>6</sup>, Ing. Osniel Hernandez Calvo<sup>7</sup>, Ing. Marisleydis Socas Alvez<sup>8</sup> , Anisley Delfino Rodríg.** *METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD.*
15. **Technologies, B. P.** Visual Paradigm. [En línea] 2 de noviembre de 2010. <http://www.visual-paradigm.com>.
16. Mailxmail. [En línea] 7 de febrero de 2010. <http://www.mailxmail.com/curso-sistemas-bases-datos/sgbd-sistemas-gestores-bases-datos-introduccion..>

17. **Perissé, M. C.** *Una Metodología Simplificada*. Argentina : s.n., 2001.
18. **Díaz, Néstor A.** Tienda Linux. [En línea] 22 de marzo de 2003.  
[http://soporte.tiendalinux.com/portal/Portfolio/postgresql\\_ventajas\\_html](http://soporte.tiendalinux.com/portal/Portfolio/postgresql_ventajas_html).
19. Datacleaner. [En línea] <http://datacleaner.eobjects.org/>.
20. Pentaho. [En línea] [http://www.pentaho.com/products/data\\_integration/](http://www.pentaho.com/products/data_integration/).
21. Pentaho Corporation. [En línea] 2005. <http://pentaho.com>.
22. Apache Tomcat. [En línea] <http://www.tomcat.apache.org>.
23. [En línea] [http://www.db-system.com/portal/page?\\_pageid=33,7665&\\_dad=portal&\\_schema=PORTAL](http://www.db-system.com/portal/page?_pageid=33,7665&_dad=portal&_schema=PORTAL).
24. **Tecky Wong Leyva, Maibel Ferrat Cardo.** *Procedimiento para el análisis de los resultados de las evaluaciones de software en la UCI*. La Habana, Cuba : s.n., Junio, 2011.
25. Calisoft. [En línea] <http://calisoft.uci.cu/tmp/documentos/normas/iso/NC-ISO-IEC%209126-1.pdf>, .
26. **Pérez, Beatriz.** *Tesis de maestría "Proceso de Testing Funcional Independiente"*. Montevideo, Uruguay : s.n., 2006.
27. **IEEE.** *IEEE Std. 610 Computer dictionary, IEEE*. 1990.
28. **Edumilis Méndez, María Pérez, Luis E. Mendoza.** *Departamento de Procesos y Sistemas, Edificio de Matemáticas y Sistemas, Laboratorio de Investigación en Sistemas de Información (LISI)*. Venezuela : s.n.

## BIBLIOGRAFÍA

- Apache Tomcat. [En línea] <http://www.tomcat.apache.org>.
- B., Eduardo Abedrapo. *Principales aspectos del Sistema Nacional de Inversiones de Chile*.
- Beltrán, Carlos Patricio López. *Análisis, Diseño e implementación de un Data Mart para la Dirección Financiera y Recursos Humanos de la Escuela Politécnica del Ejército para una toma de decisión efectiva*. SANGOLQUI : s.n., 10 de abril de 2007.
- Calisoft. [En línea] <http://calisoft.uci.cu/tmp/documentos/normas/iso/NC-ISO-IEC%209126-1.pdf>, .
- Dario Bernabeu, Ricardo. *Metodología propia para la construcción de un Data Warehouse*. 2007.
- Datacleaner. [En línea] <http://datacleaner.eobjects.org/>.
- DB system [En línea] [http://www.db-system.com/portal/page?\\_pageid=33,7665&\\_dad=portal&\\_schema=PORTAL](http://www.db-system.com/portal/page?_pageid=33,7665&_dad=portal&_schema=PORTAL).
- Díaz, Néstor A. Tienda Linux. [En línea] 22 de marzo de 2003. [http://soporte.tiendalinux.com/portal/Portfolio/postgresql\\_ventajas.html](http://soporte.tiendalinux.com/portal/Portfolio/postgresql_ventajas.html).
- Edumilis Méndez, María Pérez, Luis E. Mendoza. *Departamento de Procesos y Sistemas, Edificio de Matemáticas y Sistemas, Laboratorio de Investigación en Sistemas de Información (LISI)*. Venezuela : s.n.
- Galicia, Business Intelligence. SINNEXUS. [En línea] 14 de enero de 2010. [http://www.sinnexus.com/business\\_intelligence/datawarehouse.aspx](http://www.sinnexus.com/business_intelligence/datawarehouse.aspx).
- García Gerardo, Clemente. *Tesis doctoral: Un Sistema para el Mantenimiento de Almacenes de Datos*. 2008.
- Herrera, C. Adictos al trabajo. [En línea] 30 de octubre de 2007. <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse4>.
- IEEE. *IEEE Std. 610 Computer dictionary,IEEE*. 1990.
- Ing. Alberto Límia Navarro<sup>1</sup>, Ing. Mabel Medina Rodríguez<sup>2</sup>, Ing. Daulemys Rigo Portillo<sup>3</sup>, Ing. Madelys Cuesta Villa<sup>4</sup>, Ing. Yamila Mateu Romero<sup>5</sup>, Ing. Julio Ernesto Ortiz<sup>6</sup>, Ing. Osniel Hernandez Calvo<sup>7</sup>, Ing. Marisleydis Socas Alvez<sup>8</sup> , Anisley Delfino Rodríguez. *METODOLOGÍA PARA EL DESARROLLO DE SOLUCIONES DE ALMACENES DE DATOS E INTELIGENCIA DE NEGOCIO EN CENTALAD*.
- Inmon, Bill. *Building the Data Warehouse*. 2005. 4ta edición.
- Iznaga, Yonelbys. *Tesis: Sistema Data Warehouse*. 2008.
- Kimball, R. *The Data Warehouse ETL Toolkit. Practical Techniques for extracting, cleaning, conforming, and delivering data*. s.l. : WILEY PUBLISHING, INC., 2002.
- Kimball, Ralph. *The Data Warehouse ETL*. 2004. 1ra edición.
- Kimball, Ralph. *El Juego de Herramientas del Almacén de Datos*. 1996.

- Kimball, Ralph y Ross, Margy. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modelin*. 2002. 2da edición.
- Mailxmail. [En línea] 7 de febrero de 2010. <http://www.mailxmail.com/curso-sistemas-bases-datos/sgbd-sistemas-gestores-bases-datos-introduccion..>
- Medina Mustelier, Doris. *Tesis: Técnicas de Extracción, Transformación y Carga de Datos del Sistema de Información Nacional de Seguridad Ciudadana en la República Bolivariana de Venezuela*. 2009.
  - *Metodología para el desarrollo de Almacenes de Datos y BI*. Ciudad de la Habana, UCI : s.n.
  - Nader, Ing. Javier. *Sistema de apoyo gerencial universitario. Tesis de Magister en Ingeniería de Software*. 2003.
  - ONE. [En línea] <http://www.one.cu>.
  - Orellanos, Norah Velazco. *Estadísticas de Inversión Extranjera Directa en los Países de la Comunidad Andina*.
  - Pentaho. [En línea] [http://www.pentaho.com/products/data\\_integration/](http://www.pentaho.com/products/data_integration/).
  - Pentaho Corporation. [En línea] 2005. <http://pentaho.com>.
  - Pérez, Beatriz. *Tesis de maestría "Proceso de Testing Funcional Independiente"*. Montevideo, Uruguay : s.n., 2006.
  - Perissé, M. C. *Una Metodología Simplificada*. Argentina : s.n., 2001.
  - Ruiz, Tomás Martínez, Navarro, Almudena Quevedo y Visuete, Michael. *UNIVERSIDAD DE CASTILLA-LA MANCHA ESCUELA SUPERIOR DE INFORMÁTICA, modelos avanzados de bases de datos*. Quito : s.n., Marzo, 2006.
  - Secretaria de Economía . [En línea] <http://www.economia.gob.mx/swb/work/models/economia/Resource/516/1/images/EstadInverMexicoUE.pdf>.
  - Sinnexus. [En línea] [http://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx).
  - Technologies, B. P. Visual Paradigm. [En línea] 2 de noviembre de 2010. <http://www.visual-paradigm.com>.
  - Tecky Wong Leyva, Maibel Ferrat Cardo. *Procedimiento para el análisis de los resultados de las evaluaciones de software en la UCI*. La Habana, Cuba : s.n., Junio, 2011.
  - Vázquez, L. C. Tecnología e Internet. [En línea] 27 de octubre de 2009. <http://gestiopolis.com>.
  - Visuete Naranjo, Michael Wladimir. *Tesis: Análisis, diseño e implementación de un data mart para el área de sismología del departamento de geofísica de la escuela politécnica nacional*. 2006. 1ra edición.
  - Zepeda Sánchez, Leopoldo. *Borrador Tesis de Doctorado: Metodología para el Diseño Conceptual de Almacenes de Datos*. 2008.

## ANEXOS

## Anexo 1: Caso de prueba Consultar gastos de inversión en ámbito ambiental

Escenario	Descripción	Variables que aplican		Respuesta del sistema	Flujo central
		Perfiles de análisis	Indicadores a medir		
<b>EC 1.1</b> Gastos de inversión en el medio ambiente por organismo y entidad	Permite visualizar el reporte con las variables presentes en el mismo.	Indicadores Trimestre Entidad Organismo UM	Gastos total	Se muestra la tabla con los valores correspondientes a cada escenario y su gráfico.	Se autentica el usuario y entra al sistema. El actor despliega el A.A.G SIGOB. El actor despliega el A.A Inversiones. El actor selecciona el LT Control del plan de la ejecución física de las inversiones. El actor selecciona el reporte que desea visualizar. Se verifica que estén disponibles en las filas y las columnas todos los perfiles de análisis y los indicadores a medir.
<b>EC 1.2</b> Gastos de inversión en el Medio Ambiente por provincia y entidad		Indicadores DPA Trimestre Entidad UM	Gastos total		
<b>EC 1.3</b> Gastos de inversión en Ámbito ambiental por organismo		Indicadores Trimestre Organismo UM	Gastos total		
<b>EC 1.4</b> Gastos de protección en cuencas hidrográficas por organismo y entidad		Indicadores Trimestre Organismo Entidad UM	Gastos total		
<b>EC 1.5</b> Gastos de protección en las principales bahías por organismo y entidad		Indicadores Trimestre Entidad Organismo UM	Gastos total		
<b>EC 1.6</b> Total de gastos de inversión por NAE		Indicadores NAE Trimestre UM	Gastos total		
<b>EC 1.7</b> Total de gastos de inversión por NAE y entidad		Indicadores NAE Trimestre Entidad UM	Gastos total		

<b>EC 1.8</b> Total de gastos de inversión por provincia		Indicadores DPA Trimestre UM	No aplica		
--	--	------------------------------	-----------	--	--

Anexo 2: Lista de chequeo aplicada al MD Inversiones

<b>Estructura del documento</b>					
<b>Peso</b>	<b>Indicadores a evaluar</b>	<b>Eval</b>	<b>(NP)</b>	<b>Cantidad de elementos afectados</b>	<b>Comentarios</b>
crítico	1. ¿Los entregables contienen las secciones obligatorias de la plantilla estándar definidas para un expediente de proyecto? (portada, control de versiones, reglas de confidencialidad, tabla de contenidos y contenido). Ver expediente de proyecto.				
<b>Indicadores definidos en el desarrollo</b>					
<b>Peso</b>	<b>Indicadores a evaluar</b>	<b>Eval</b>	<b>(NP)</b>	<b>Cantidad de elementos afectados</b>	<b>Comentarios</b>
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?				
crítico	2. ¿Los reportes son configurables a través de la interfaz del sistema?				
	3. ¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?				
crítico	4. ¿No existen restricciones para construir cubos OLAP con dimensiones y niveles de agregación ilimitados?				
crítico	5. ¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades, conformando nuevos reportes?				
	6. ¿El sistema responde de una forma rápida a la información que le sea solicitada?				

	por el usuario?				
	7. ¿El sistema refleja cualquier lógica del negocio para poder responder a preguntas específicas?				
crítico	8. ¿El sistema garantiza la confidencialidad y seguridad de acceso a los datos por rol de los usuarios?				
	9. ¿Los datos e información derivados del proceso de análisis realizado mediante la aplicación, apoyan la toma de decisiones en la Institución?				
crítico	10. ¿Los cambios en los datos se reflejan automáticamente en los reportes de forma instantánea?				
Semántica del documento					
<b>Peso</b>	<b>Indicadores a evaluar</b>	<b>Eval</b>	<b>(NP)</b>	<b>Cantidad de elementos afectados</b>	<b>Comentarios</b>
crítico	1. ¿Se han identificado errores ortográficos en los entregables?				
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?				
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?				

### GLOSARIO DE TÉRMINOS

1. **AD:** Almacén de Datos
2. **BD:** Base de Datos.
3. **Caso de Uso:** Secuencias de acciones que el sistema puede llevar a cabo interactuando con sus actores, incluyendo alternativas dentro de las secuencias.
4. **Centros Informantes:** Los Centros Informantes son las empresas u organismos que suministran información a las oficinas de estadísticas en sus diferentes niveles.
5. **DATEC:** Centro de Tecnologías de Gestión de Datos.
6. **ETL:** Por sus siglas en inglés: Extracting, Transforming, Loading. Procesos de extracción, transformación y carga de los datos a una base de datos.
7. **HOLAP:** Procesamiento Analítico en línea Híbrido (Hybrid Online Analytical Process).
8. **Indicadores:** Se dice de la variable que puede tomar un valor de una determinada unidad de medida y de un determinado tipo de datos (generalmente numérico). Los indicadores de la ONE están bien definidos y tienen un código único que los identifica.
9. **MD:** Mercado de Datos.
10. **Modelo:** Especie de planilla diseñada para almacenar la información estadística.
11. **MOLAP:** Procesamiento Analítico en línea Multidimensional (Multidimensional Online Analytical Processing).
12. **NAE:** Es un nomenclador, para la organización de las actividades económicas de las empresas.
13. **OLTP:** Procesamiento de Transacciones en Línea (On Line Transactional Processing).
14. **OLAP:** Procesamiento de Transacciones en Línea (On Line Analytical Processing).
15. **ONE:** Oficina Nacional de Estadísticas.
16. **ROLAP:** Procesamiento Analítico en Línea Relacional (Relational On Line Analytical Processing).
17. **SGBD** Sistemas de gestión de bases de datos.