

**COMO DISMINUIR LOS RIESGOS DE LOS PROCESOS DE ETL EN EL  
PROYECTO DE INTELIGENCIA DE NEGOCIOS EN UNA EMPRESA DE  
TRANSPORTE**

**LYDA DIANA HENAO DORADO  
200110075010**

**Trabajo de Grado**

**Asesor  
LUIS FELIPE ROSSO RICAURTE  
Director Tecnología Informática  
Empresa Logística Transporte**

**FACULTAD DE INGENIERÍA DE SISTEMAS  
UNIVERSIDAD EAFIT  
MEDELLÍN  
2008**

## TABLA DE CONTENIDO

1.	DISEÑO DE LA INVESTIGACIÓN	5
	TÍTULO	5
1.1.	1.1. Planteamiento del problema	5
1.2.	1.2. Justificación	7
1.3.	1.3. Delimitación de la investigación	8
1.3.1.	1.3.1. En tiempo y espacio	9
1.3.2.	1.3.2. <i>En</i> contenido	9
1.4.	1.4. Objetivos de la investigación	9
1.4.1.	1.4.1. Objetivo general	9
1.4.2.	1.4.2. Objetivos específicos	9
1.5.	1.5. Productos esperados	10
2.	CAPITULO 2. ESQUEMA DE FUNDAMENTOS	11
2.1.	2.1. Marco conceptual	11
2.1.1.	2.1.1. Abreviaturas	12
2.1.2.	2.1.2. Términos técnicos	12
2.1.3.	2.1.3. Definiciones	15
2.1.3.1.	2.1.3.1. OLAP (On – Line Analytical Processing)	15
2.1.3.2.	2.1.3.2. Reporting & Query	16
2.1.3.3.	2.1.3.3. Data Mining	16
2.2.	2.2. Marco contextual	17
2.2.1.	2.2.1. Descripción de la empresa	17
2.2.2.	2.2.2. Misión	18
2.2.3.	2.2.3. Datos generales de la empresa	18
2.2.4.	2.2.4. Reseña histórica	18
2.2.5.	2.2.5. Unidades de negocio de la empresa	19
2.2.6.	2.2.6. DataMarts	20
2.2.6.1.	2.2.6.1. Análisis y diseño lógico – DataMart de Operaciones de una empresa de transporte.	20

2.2.7.	2.2.7. Fuentes de datos de la empresa transportadora.	30
2.3.	2.3. Marco histórico	31
2.3.1.	2.3.1. <i>Referencia histórica de la Inteligencia de Negocios</i>	31
2.3.2.	2.3.2. Historia del Data Warehouse	32
2.4.	2.4. Marco teórico o de referencia	33
2.4.1.		33
	2.4.1. <i>Inteligencia de Negocios</i>	
2.4.1.1.	2.4.1.1. <i>¿Qué es inteligencia?</i>	33
2.4.1.2.	2.4.1.2. <i>¿Qué es negocio?</i>	33
2.4.1.3.	2.4.1.3. <i>Definición y características de la Inteligencia de Negocios</i>	34
2.4.1.4.	2.4.1.4. <i>Datos, información y conocimiento</i>	36
2.4.1.5.	2.4.1.5. <i>Objeto de estudio de la Inteligencia de Negocios</i>	37
2.4.1.6.	2.4.1.6. <i>Arquitectura de un proyecto en Inteligencia de Negocios</i>	37
2.4.1.7.	2.4.1.7. <i>Inteligencia de Negocios en los departamentos de la empresa de transporte.</i>	38
2.4.2.	2.4.2. <i>El proceso ETL: definición y etapas</i>	39
2.4.2.1.	2.4.2.1. <i>Limpieza de datos</i>	40
2.4.2.2.	2.4.2.2. <i>La ciencia de limpiar los datos</i>	41
2.4.2.3.	2.4.2.3. <i>¿Por qué se tienen que limpiar los datos?</i>	42
2.4.3.	2.4.3. <i>Data Warehouse</i>	42
2.4.3.1.	2.4.3.1. <i>Características del Data Warehouse</i>	43
2.4.3.2.	2.4.3.2. <i>Aportes del Data Warehouse a las empresas</i>	43
2.5.	2.5. Marco legal	44
3.	DESCRIPCIÓN TÉCNICA DE LA INVESTIGACIÓN	45
3.1.	3.1. Diseño metodológico	45
3.1.1.	3.1.1. <i>Tipo de investigación</i>	45
3.1.2.	3.1.2. <i>Nivel de investigación</i>	46
3.1.2.1.	3.1.2.1. <i>Pregunta de investigación</i>	47
3.1.3.	3.1.3. <i>Instrumentos de recolección de información</i>	47

3.1.4.	3.1.4. Metodología	48
4.	ANÁLISIS Y RESULTADOS DE LA INVESTIGACIÓN	50
4.1.	4.1. Aplicación de procesos de ETL en Inteligencia de Negocios en empresas de la <i>ciudad</i>	51
4.2.	4.2. Cinco principales etapas para el desarrollo exitoso de los procesos de ETL en el proyecto de Inteligencia de <i>Negocios</i>	58
4.3.	4.3. Riesgos más comunes en el diseño y desarrollo de los procesos de ETL para el proyecto de Inteligencia de Negocios de una empresa de transporte donde se proponen recomendaciones que los <i>disminuyan</i>	63
4.4.	4.4. Proceso de mapeo inicial de los datos y fuentes de destino en los procesos de <i>ETL</i>	69
4.5.	4.5. Estrategias para manejar los problemas de calidad de datos en el proceso <i>ETL</i>	69
5.	CONCLUSIONES DE LA INVESTIGACIÓN Y RECOMENDACIONES	74
6.	ASPECTOS ADMINISTRATIVOS	77
6.1.	6.1. <i>Recursos</i>	77
6.2.	6.2. <i>Presupuesto</i>	77
7.	BIBLIOGRAFÍA	78
8.	ANEXOS	79
8.1.	8.1. Cronograma de <i>actividades</i>	79

# CAPITULO 1

## DISEÑO DE LA INVESTIGACIÓN

### **Título**

Como Disminuir los Riesgos de los Procesos de ETL en el Proyecto de Inteligencia de Negocios en una empresa de transporte.

### **1.1. Planteamiento del problema**

*“Para que una empresa sea competitiva las personas que toman las decisiones necesitan acceder rápida y fácilmente a la información de la empresa y esto se realiza por medio de la Inteligencia de Negocios”<sup>1</sup>.*

La era de la tecnología y de la información vista como un cambio positivo en los paradigmas de las personas, los grupos y las sociedades, ha traído consigo una serie de retos que han impuesto a las organizaciones y a sus gerentes nuevas formas de pensar la empresa, los negocios y la toma de decisiones, es decir, han hecho necesario que se piense una forma “inteligente” de organizar la información y tener los datos a la mano. Esta “inteligencia” tiene como arma secreta el uso efectivo de la información, en especial información altamente integrada que permite a las organizaciones visualizar en detalle cómo se mueve su negocio y así optimizar sus procesos para ser más eficientes.

Es desde este enfoque donde a partir de los años 90 surge entre los

---

<sup>1</sup> [www.aerofitnet.com](http://www.aerofitnet.com). Visitado el 03 de Abril de 2008.

empresarios e ingenieros Estadounidenses el término *Business Intelligence* (Inteligencia de Negocios en español) con el cual se enfatizaba la necesidad de contar en las empresas con datos e información integral que facilitará la toma de decisiones y la posibilidad de ampliar el negocio; además se empieza a centrar el interés de la tecnología informática en los proyectos de Inteligencia de Negocios y sus inmensas posibilidades en el campo de los sistemas.

Los proyectos de Inteligencia de Negocios denominados como una arquitectura y una colección de aplicaciones y bases de datos integradas y orientadas al soporte de las decisiones dentro de las empresas se componen de una serie de procesos logísticos y técnicos para el cumplimiento de su objetivo, entre ellos se destacan la creación de bases de datos que recopilan la información histórica de las empresas y el más importante en los últimos tiempos – y centro de este planteamiento – : el proceso ETL (siglas de Extracción, Transformación y Carga de datos); éste proceso tiene como fin organizar el flujo de los datos en las organizaciones con la ayuda de herramientas como SQL Server 2005 de Microsoft que se emplea en la empresa de transporte para mover los datos desde múltiples fuentes a un almacén de datos, reformatearlos, limpiarlos y cargarlos en otra base de datos.

La idea en sí de un proceso ETL es que este lea los datos primarios de unas fuentes de datos o sistemas transaccionales (en el caso de la empresa de transporte se manejan los sistemas Infocar, Sipres y Excel), los extraiga y los limpie, realizando posteriormente la transformación de los mismos para que al final cargue todos los datos organizados en un almacén o DataMart, es en este momento cuando surge el factor de inteligencia y los datos se hacen disponibles para su utilización por parte de los usuarios, por un lado, y por otro, por parte de los gerentes para la toma de las decisiones al interior de la organización.

En este sentido, la empresa de transporte posee en la actualidad una serie de

procesos operativos que soportan el día a día del negocio, sin embargo, la ausencia de una solución efectiva que minimice los riesgos en cuanto al manejo de datos del proceso ETL (Extracción, Transformación y Carga), y la cual facilite una toma de decisiones efectivas a nivel organizacional, es una dificultad que presenta el área de BI (Business Intelligence) de la empresa, ya que el análisis de la información generada por la operación de datos es alto y complejo; éste proceso cuando es realizado de forma manual y no automatizada genera en algunas ocasiones, el surgimiento de riesgos que entorpecen la carga de datos, además que en la actualidad es un proceso demasiado lento y tedioso, especialmente cuando se deben analizar grandes cantidades de información.

Es por todo esto, que se hace necesario investigar sobre el tema del proceso ETL y sus riesgos en la implementación, para que de este modo se cree una propuesta que los disminuya, consolidando así el acceso a la información de la empresa de manera rápida y con datos más coherentes, e igualmente un modo más seguro para la toma de decisiones inteligentes y efectivas sin la necesidad de tanta carga operativa.

## **1.2. Justificación**

La tesis se llevo a cabo en una empresa que se ha establecido en el sector de logística en transporte urbano, y gracias a su desempeño y buen servicio ha estado en constante crecimiento, dando como resultado el manejo masivo de datos del negocio y creando la necesidad de la utilización de mejores metodologías y herramientas para el análisis del mismo; es a partir de aquí que se crea el área de Inteligencia de Negocios, la cual tiene como misión al interior de la empresa habilitar la ejecución estratégica del negocio mediante el desarrollo y mantenimiento de soluciones que den respuesta ágil, adecuada y efectiva a las necesidades actuales y futuras de la empresa. Desde esta perspectiva, el área de

Inteligencia de Negocios tiene como función sistematizar los procesos de manipulación de datos y reemplazar los antiguos métodos manuales, y cuyo éxito se desprende de realizar los procesos ETL utilizando mejores prácticas de ejecución, llevando un seguimiento para disminuir los riesgos de dichos procesos, asegurando la integridad de los datos y generando como beneficios la optimización del sistema, mejora en las consultas, en el servicio y por último la fijación de una plataforma robusta para soportar el constante crecimiento de la empresa.

Por medio de esta investigación se obtendrán logros importantes a nivel de conocimiento y estrategias que optimizaran los procedimientos y prácticas empresariales dentro del área de Inteligencia de Negocios, donde la correcta planeación, ejecución y seguimiento de los procesos de ETL generara día a día mejores soluciones y desarrollos de nuevas ideas.

El cómo disminuir riesgos de un proceso de ETL en un proyecto de inteligencia de negocios en una empresa de transporte permitiría:

- Identificar proactivamente los riesgos de los procesos de ETL de tal manera que no se impacte el cronograma del proyecto.
- Reducir los costos de consultoría, con el análisis previo de las fuentes de datos, la calidad de los datos y el mapeo de la fuente y el destino.
- Construir una base de conocimiento para el manejo de riesgos en los procesos de ETL y las estrategias de cómo mitigarlos.

Para concluir es importante anotar que en esta investigación quedaran plasmados las bases y conceptos necesarios para enmarcar la importancia que debe darse a los procesos de ETL dentro de Inteligencia de Negocios, ya que con la información suministrada por la presente investigación se verán beneficiados estudiantes, técnicos e ingenieros, profesores e investigadores, ya que podrán acceder al material de consulta e investigación actualizada, veraz y confiable sobre procesos ETL dentro del área de Inteligencia de Negocios de la empresa.

### **1.3. Delimitación de la investigación**

Debido a la gran cantidad de información y existencia de procesos al interior de cualquier empresa, en este caso en la empresa de transporte, la investigación será realizada teniendo en cuenta una delimitación, en tiempo, espacio y contenido.

#### **1.3.1. En tiempo y espacio**

La investigación está delimitada para realizarse en el transcurso del año 2008 – 2009 en la empresa de transporte.

#### **1.3.2. En contenido**

La investigación debido a sus objetivos y metodología se centrara en el proceso ETL y sus riesgos, y desde acá se plantearán las estrategias para la disminución de los mismos.

### **1.4. Objetivos de la investigación**

Los objetivos que se pretenden lograr con la investigación se presentan a continuación.

#### **1.4.1. Objetivo general**

Identificar los principales riesgos que existen al interior de los procesos ETL (Extracción, Transformación y Carga) de las fuentes de datos Sipres, Infocar y Excel en el proyecto de Inteligencia de Negocios en la empresa de transporte, con el fin de crear estrategias que los disminuyan.

### **1.4.2. Objetivos específicos**

- Examinar las cinco principales etapas para el desarrollo exitoso de los procesos de ETL en el proyecto de Inteligencia de Negocios.
- Enumerar los riesgos más comunes en el diseño y desarrollo de los procesos de ETL en el proyecto de Inteligencia de Negocios.
- Proponer cinco posibles recomendaciones que disminuyan los riesgos identificados en el proceso ETL (Sipres, Infocar y Excel).
- Describir el proceso de mapeo inicial de los datos y fuentes de destino en los procesos de ETL.
- Establecer estrategias necesarias para manejar los problemas de calidad de datos en el proceso ETL.

### **1.5. Productos esperados**

Del desarrollo de la investigación en la empresa de transporte se espera obtener:

- Contar con un documento completo y profesional donde se enumeran los riesgos más comunes en el diseño y desarrollo de los procesos de ETL para el proyecto de Inteligencia de Negocios y se proponen estrategias para su disminución.
- Definir de manera sencilla y práctica para su correcta implementación las etapas para el desarrollo exitoso de los procesos de ETL.
- Describir las consideraciones y estrategias para manejar los problemas de calidad de datos.

## **CAPITULO 2**

### **ESQUEMA DE FUNDAMENTOS**

En este capítulo se le brinda al lector una base clara desde la cual podrá comprender la investigación por medio de fundamentos históricos, teóricos y conceptuales, para así contar con una descripción más detallada del problema de investigación y la teoría que da significado a la misma.

En este sentido, el presente esquema se dividirá en los siguientes apartados:

- Marco Conceptual: Expone y define los conceptos claves para la comprensión del texto de la investigación.
- Marco Contextual: Expone el contexto en el cual se implementó la investigación, para el caso la empresa de transporte como espacio de investigación y como espacio donde se implementa el proceso de ETL y la Inteligencia de Negocios.
- Marco Histórico: Expone los antecedentes del problema y del tema investigado.
- Marco Teórico o de Referencia: Expone la teoría del tema principal de la investigación (ETL) sus características y principales operaciones; además los temas secundarios relacionados.
- Marco Legal: Expone las principales normas en las cuales esta enmarcada la investigación.

#### **2.1. Marco conceptual**

A continuación se expone el marco que nombra y define los principales conceptos abordados dentro de la investigación, esto con el fin de hacerlos más claros y entendibles. Dentro de éste podrán visualizarse, tanto abreviaturas utilizadas, como términos técnicos o de difícil comprensión.

### 2.1.1. Abreviaturas

- **ETL:** Se refiere a las siglas en inglés de Extract, Transformation y Load, que en español se conocen como Extracción, Transformación y Carga de datos.
- **BI:** Se refiere a las siglas en inglés de Business Intelligence, que en español se conocen como Inteligencia de Negocios.
- **BSC:** Se refiere a las siglas en inglés de Balanced Score Card, que en español se conocen como Balanceo y Estrategia.
- **DSS:** Se refiere a las siglas en inglés de Decisión Support System, que en español se conocen como Sistema al Soporte de la Decisión.
- **OLAP:** Se refiere a las siglas en inglés de On – Line Analytical Processing, que en español se conocen como Procesador Analítico en Línea.
- **SQL Server:** Es una plataforma global de bases de datos que ofrece administración de datos empresariales, con herramientas integradas de inteligencia de negocios. Éste ofrece almacenamiento de datos más seguro y confiable tanto para datos relacionales como estructurados, lo que le permite crear y administrar aplicaciones de datos altamente disponibles y con mayor rendimiento para utilizar en las empresas.

### 2.1.2. Términos técnicos

- **Base de Datos:** Son un conjunto de tablas que permiten almacenar datos de diferentes tipos.

- **Base de Datos Multidimensional:** Conjunto de tablas donde cada una puede asimilarse a un hipercubo – más concretamente si de herramientas OLAP se trata – a un cubo OLAP, donde las dimensiones del mismo corresponden a los campos de dimensiones y el valor almacenado en cada celda del cubo equivale a la métrica o métricas almacenadas en la tabla.
- **Base de Datos Relacional:** Bases de datos construida en un sistema de gestión de base de dato bajo un conjunto de relaciones.
- **Campos:** En las bases de datos es la mínima unidad de información a la cual se puede acceder.
- **Cubo:** Es una forma de estructurar y almacenar la información donde se tienen precalculados indicadores cuyo valor se obtiene del cruce de los hechos con las dimensiones.
- **DataMart:** Nombre técnico que se da a un almacenamiento de datos que contiene información relacionada con un tema específico.
- **Data Warehouse:** Sistema de información que permite el almacenamiento de en un único entorno de la información histórica e integrada proveniente de los distintos sistemas de la empresa y que refleja los indicadores clave asociados a los negocios de la misma.
- **Dimensiones:** Las variables de análisis o dimensiones son las entidades por las cuales se analiza la información. El término entidades hace referencia a los elementos que definen el negocio.
- **Hechos:** Los hechos o movimientos son el registro de los eventos que afectan o definen el valor de los indicadores. De estas se crean las Tablas de Hechos.
- **Herramienta:** Es un objeto elaborado a fin de facilitar la realización de una tarea.
- **Join:** La sentencia Join en SQL permite combinar registros de dos o más tablas en una base de datos relacional.
- **Indicador o Medida Calculada:** Son los valores que se desea medir o calcular. Los indicadores son instrumentos que permiten medir la gestión y

resultados del negocio, por medio de la relación de variables cuantitativas y cualitativas que son comparadas entre una situación presente y un patrón de referencia requerido para medir la productividad y competitividad organizacional.

- **Integration Services:** Microsoft Integration Services es una plataforma para la creación de soluciones de integración de datos y de transformaciones de datos de alto rendimiento. Éste sirve para resolver complejos problemas empresariales mediante la copia o descarga de archivos, el envío de mensajes de correo electrónico como respuesta a eventos, la actualización de almacenes de datos, la limpieza y minería de datos, y la administración de objetos y datos de SQL Server.
- **Inteligencia de Negocios:** La Inteligencia de Negocios es el conjunto de productos y servicios que permiten a los usuarios finales de un sistema acceder y analizar de manera rápida y sencilla la información para la toma de decisiones de negocio a nivel operativo, táctico y estratégico.
- **Metadato:** Son datos que describen otros datos.
- **Miembro:** El término miembro hace referencia a cada uno de los valores que hacen parte o que conforman una variable de análisis.
- **Modelo de Estrella y Modelo Copo de Nieve:** Nombres utilizados para referirse a las dos formas que existen para diseñar la arquitectura de una bodega de datos.
- **Proceso:** Es un conjunto de actividades o eventos que se realizan o suceden (alternativa o simultáneamente) con un determinado fin.
- **Registro:** Un registro es un conjunto de campos que contienen los datos que pertenecen a una misma repetición de entidad. Se le asigna automáticamente un número consecutivo (número de registro) que en ocasiones es usado como índice, aunque lo normal y práctico es asignarle a cada registro un campo clave para su búsqueda.
- **Reporting Services:** Microsoft Reporting Services es una plataforma de reportes basada en servidores, la misma que puede ser empleada para crear y administrar reportes tabulares, de matrices, gráficos y de libre

formato, la información de estos reportes pueden provenir de diferentes orígenes de datos. Los reportes que se definen pueden ser administrados a través de una conexión basada en Web.

- **Sistema:** Es un conjunto ordenado de elementos cuyas propiedades se interrelacionan e interactúan de forma armónica entre sí.
- **Staging:** Es una dimensión de auditoria en la etapa de carga de ETL la cual contiene los datos para ser cargados al DataMart de Operaciones, además permite realizar la integración de los datos y su transformación. Esta dimensión se utiliza para hacer las auditorias de los datos.
- **Tabla:** Tabla en las bases de datos se refiere al tipo de modelamiento de datos, donde se guardan los datos recolectados por un programa. Su estructura general se asemeja a la vista general de un programa de hoja de cálculo.

### 2.1.3. Definiciones

#### 2.1.3.1. OLAP (On – Line Analytical Processing)

Son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos, entre otros (en el caso de la empresa de transporte es el comportamiento de las Operaciones del negocio). Este sistema es típico de los DataMarts.

Dentro de los sistemas OLAP se pueden apreciar tres categorías:

- **ROLAP:** Implementación OLAP que almacena los datos en un motor relacional. Típicamente, los datos son detallados, evitando las agregaciones y las tablas se encuentran normalizadas. Los esquemas más

comunes sobre los que se trabaja son estrella o copo de nieve, aunque es posible trabajar sobre cualquier base de datos relacional.

- **MOLAP:** Implementación OLAP la cual almacena los datos en una base de datos multidimensional. Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado. Estos valores pre calculados son la base de las ganancias de desempeño de este sistema.
- **HOLAP:** Almacena algunos datos en un motor relacional y otros en una base de datos multidimensional.

#### **2.1.3.2. Reporting & Query**

Los sistemas de Reporting & Query, no son almacenes de datos, y poseen como característica principal la complejidad de las consultas, los altos tiempos de respuesta y la interferencia con otros procesos informáticos que compartan su entorno.

La explotación del Data Warehouse mediante Reporting & Query permite flexibilidad de acceso, proporcional a la experiencia y formación del usuario. Es recomendable el mantenimiento de acuerdo a tres niveles de experiencia de los usuarios, así:

- Usuarios poco expertos: Solicitud de informes o consultas según unos parámetros predeterminados.
- Usuarios con cierta experiencia: Solicitud de informes y consultas flexibles mediante una interfaz gráfica de ayuda.
- Usuarios con alta experiencia: Solicitud de informes y consultas, además de escribir la consulta en un lenguaje de interrogación de datos.

#### **2.1.3.3. Data Mining**

El Data Mining es un proceso que, a través del descubrimiento y cuantificación de relaciones predictivas en los datos, permite transformar la información disponible en conocimiento útil de negocio. Esto es debido a que no es suficiente "navegar" por los datos para resolver los problemas de negocio, sino que se hace necesario seguir una metodología ordenada que permita obtener rendimientos tangibles de este conjunto de herramientas y técnicas de las que dispone el usuario. Constituye por tanto una de las vías clave de explotación del Data Warehouse, dado que es este su entorno natural de trabajo.

Se trata de un concepto de explotación de naturaleza radicalmente distinta a la de los sistemas de información de gestión, dado que no se basa en coeficientes de gestión o en información altamente agregada, sino en la información de detalle contenida en el almacén, adicionalmente el usuario no se conforma con la mera visualización de datos, sino que trata de obtener una relación entre los mismos que tenga repercusiones en su negocio.

## **2.2. Marco contextual**

A continuación se expone el marco que define el contexto donde fue desarrollada la investigación: en la empresa de transporte.

### **2.2.1. Descripción de la empresa**

Es una empresa líder en la solución de la logística de transporte urbano de mercancías soportada en vehículos adecuados, tecnología de vanguardia, desarrollos y procesos logísticos, cumplimiento en la promesa de servicio y una excelente selección y capacitación del talento humano.

### **2.2.2. Misión**

Ofrecer soluciones efectivas y únicas fundamentadas en talento humano, sistemas de información y prácticas socialmente responsables, que impacten el valor de su cadena de distribución.

### **2.2.3. Datos generales de la empresa**

### **2.2.4. Reseña histórica**

La empresa se constituyó como compañía transportadora en el año de 1994 en la ciudad de Medellín, obteniendo su licencia de funcionamiento ante el Ministerio de Transporte en 1996.

En 1994, el transporte urbano de carga era prestado en gran proporción por vehículos particulares o públicos contratados directamente por los generadores y remitentes de carga. Dada esta perspectiva, el sistema de trabajo generaba problemas en relación a la escasa oferta de vehículos, la informalidad en la capacidad de carga y la dificultad para garantizar perfiles definidos en el recurso humano implicado en la operación.

Entendiendo esta problemática, se gesta el proyecto de la empresa transportadora, en el concepto de profesionalización del servicio de transporte de carga que se deba movilizar en áreas urbanas, distritales, metropolitanas y sus zonas de influencia; caracterizándose por ofrecerle al mercado un servicio trazable que genere indicadores de gestión, suministre un adecuado y moderno parque automotor operado por recurso humano idóneo y capacitado para operaciones concretas. Además la empresa transportadora esta soportada en desarrollos logísticos para cada operación, interpretando las necesidades de los remitentes de carga, contando con sistemas de seguridad, radiocomunicaciones, monitoreo satelital y sistema de aseguramiento de las mercancías con amplias coberturas.

### 2.2.5. Unidades de negocio de la empresa

La empresa transportadora tiene varias unidades de negocio para ofrecer un servicio más especializado, entre ellas están:

- **Unidad de negocio alimentos:**

Ofrece servicio de transporte para empresas que producen y procesan productos de consumo humano.

- **Unidad de negocio básica y química:**

*Básica:* Ofrece servicio de transporte para empresas fabricantes de insumos básicos para otras industrias de sectores de gran volumen y bajo valor intrínseco de la mercancía.

*Química:* Ofrece servicio de transporte para empresas pertenecientes a sectores altamente regulados que producen bienes derivados de procesos químicos.

- **Unidad de negocio hogar y comercio:**

*Hogar:* Ofrece servicio de Transporte para empresas dedicadas a la manufactura de productos para el hogar.

*Comercio:* Ofrece servicio de transporte para empresas dedicadas a la intermediación comercial en productos nacionales o importados con alto valor y nivel de contacto con el cliente final.

### 2.2.6. DataMarts

Un DataMart es una versión especial del Data Warehouse. Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser, agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

### **2.2.6.1. Análisis y diseño lógico – DataMart de Operaciones de una empresa de transporte.**

En la presente investigación es importante mencionar el tema de análisis y diseño lógico que se maneja en el área de Inteligencia de Negocios de la empresa transportadora, ya que el lector podrá tener una aproximación a cerca de las operaciones que se llevan a cabo con el DataMart de Operaciones y cómo es el proceso en el manejo de la información e investigación que se trabajó.

Para que se pueda obtener un análisis de la situación real de la empresa, se emplean variables del negocio, fuentes de consulta, sistema de cargas, de procesamiento y de reportes que agilicen los procesos para disponer de la información centralizada y mejora a la atención de usuarios dándoles mayor autonomía. En el análisis y diseño lógico, se manejan una serie de preguntas del negocio la cual son los requerimientos o cuestionamientos que el negocio requiere resolver con el objetivo de medir el desempeño del mismo, las preguntas son:

- ¿Cuántos kilogramos se movilizan por mes, por cliente?
- ¿Cuántas unidades de transporte se movilizan por mes, por cliente?
- ¿Cuántos vehículos están asignados a la operación por cliente por mes?
- ¿Cuántos Despachos se realizan por mes?
- ¿Cuáles son los tiempos muertos y efectivos por ruta y por mes?
- Número de Despachos por Tipo de Vehículo
- % Utilización vehicular por vehículo
- Número total de entregas por mes/semana/día
- Número total de recogidas por mes/semana/día
- Número total de Novedades por mes/semana/día

Por otro lado, en cuanto al análisis multidimensional es la funcionalidad de las soluciones de la Inteligencia de Negocios que permite realizar el análisis de información a través de múltiples variables y agrupando grandes volúmenes de información.

El análisis multidimensional se soporta mediante el uso de una tecnología denominada OLAP, creada exclusivamente para soportar de manera eficiente este tipo de análisis.

OLAP (On – Line Analytical Processing) significa procesamiento analítico en línea, es decir, el proceso encargado de interrogar los datos del negocio utilizando una herramienta interactiva que permite analizar grandes volúmenes de información por múltiples variables utilizando estructuras especializadas para esta función. El análisis multidimensional responde preguntas de negocio como ¿Qué pasó? o ¿Por qué pasó?, Por ejemplo, ¿Qué pasó con el número de vehículos movilizadas? ¿Cuáles fueron la cantidad de mercancías movilizadas por operación en Medellín?, ¿Por qué se disminuyeron el número de viajes por producto?

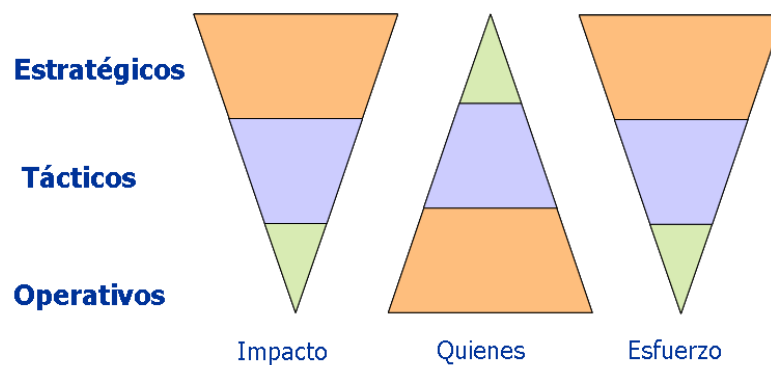
Sin embargo, es importante anotar que el análisis OLAP depende completamente de la persona quien realiza los análisis y quien es responsable de encontrar valor en los datos a través de la combinación del uso de la información y la experiencia y conocimiento del negocio.

Desde esta perspectiva, para llevar a cabo un análisis de información y soporte al proceso de decisiones existen tipos de reportes los cuales se pueden clasificar en tres diferentes grupos: Estratégicos, Tácticos y Operativos, dependiendo del objetivo, las personas o niveles de la organización a quienes esté dirigido y el tipo de decisiones que soportan. Por ejemplo, los reportes estratégicos soportan las decisiones estratégicos y están dirigidos a los niveles ejecutivos de la organización; mientras que los reportes tácticos, soportan las decisiones tácticas y están dirigidos a los niveles gerenciales y directivos y

finalmente los reportes operativos son creados para soportar las necesidades de información de los niveles operativos de la organización.

Basados en el tipo de decisiones que soportan los reportes, el siguiente gráfico representa para cada tipo de reporte, cómo es el impacto dentro de la organización, el número de personas a quienes va dirigido y el nivel de esfuerzo para construirlos.

**Figura 4. Tipos de reportes de la empresa.**



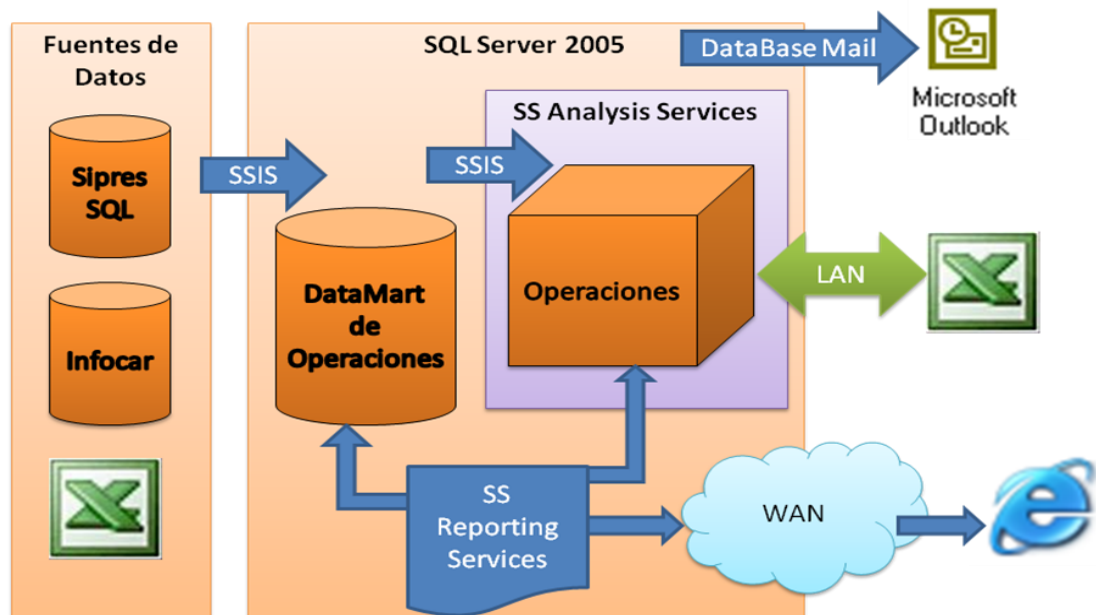
Los reportes Estratégicos por ejemplo, generan un mayor impacto dentro de la organización, por el tipo de información que suministran y las decisiones que soportan, están dirigidos a un número menor de personas y requieren un mayor el esfuerzo para su construcción. Así, mismo, en el otro extremo tenemos los reportes operativos que están dirigidos a un mayor número de personas, no requieren un gran esfuerzo de construcción, y el impacto en la organización es bajo por la información que suministran y porque soportan procesos y decisiones operativas del día a día.

- **Reportes Estratégicos:** Son los reportes que están enfocados a medir los principales indicadores del negocio, área o proceso y que normalmente se presentan de manera integrada con otros indicadores y permiten conocer tendencias y comportamientos.

- **Reportes Tácticos:** Son los reportes enfocados a conocer diferentes medidas e indicadores relacionados con los eventos y transacciones del negocio y que permiten medir el desempeño de procesos y operaciones del negocio.
- Los reportes tácticos también están orientados a permitir el análisis de la información por múltiples variables de manera interactiva.
- **Reportes Operativos:** Son los reportes que suministran información detallada de los eventos y transacciones del negocio o proceso, con el objetivo de identificar todos los elementos, variables e información relacionada con la pregunta operativa a resolver.

Por otro lado el diseño y Arquitectura del DataMart de Operaciones define la distribución física y lógica de los elementos que conforman la solución. Esta arquitectura se basa en tecnología Microsoft y los componentes incluidos son: SQL Server 2005, Analysis Services (SS AS Server), Reporting Services (SS RS), Integration Services (SS IS), SQL Server DataBase Mail, Internet Information Services y Windows 2003 Services en el servidor. Analysis Services requiere que exista al menos un dominio de Windows y en los clientes se asume Windows XP Professional o Windows Vista, Internet Explorer y Excel 2003 ó 2007.

**Figura 5. Arquitectura de ETL en la empresa.**



Es importante nombrar dentro de la arquitectura de ETL el modelo de procesos que define el tipo de carga de datos, el cual permite tomar la información de las operaciones con el objetivo de realizar la carga a la base de datos del sistema. Al sistema se le podrá definir el tipo de carga de datos los cuales son:

- **Carga de datos diferencial:** Es cuando periódicamente se cargan solamente los registros con fecha mayor a la última fecha que se tiene de la última carga.
- **Carga de datos Completa:** Es cuando se carga toda la información del DataMart, borrando toda la información previa que existía en el mismo y volviéndola a cargar, de acuerdo a los datos existentes en ese momento en las fuentes de datos.
- **Carga de datos de un periodo determinado:** Es cuando se establece una fecha inicio y una fecha fin de carga de datos, este tipo de carga de datos es útil principalmente para el reproceso de datos de determinadas fechas. Para este proceso primero se borran los datos que existen para ese rango de fechas y luego se cargan los datos para ese mismo rango desde las fuentes de datos.

- **Carga de datos de los últimos periodos:** Es cuando se carga el DataMart teniendo en cuenta la cantidad de días desde hoy hacia atrás de datos que se desea cargar o actualizar, para este proceso primero se borran los datos que existen en el periodo de tiempo que se definió y luego se procede a realizar la carga de datos de ese mismo periodo de tiempo.

De acuerdo a una tabla de parámetros que se guardará dentro de la base de datos del DataMart contendrá los siguientes campos:

- **idParametro:** Es un valor auto numérico único en la tabla.
- **strCodigoParametro:** Es el nombre del parámetro y es un campo único en la tabla.
- **strValorParametro:** Aquí se almacena el valor del parámetro cuando el dato no es numérico.
- **dblValor:** Valor que toma el parámetro cuando es numérico.
- **strDescripcion:** Almacena una explicación del objetivo del parámetro

La lista de parámetros para la programación del tipo de carga de datos del DataMart serán los siguientes:

- **CargaDatosOperaciones:** Valor numérico para determinar el rango de fechas a cargar en el DataMart de Operaciones. Si el valor es cero (0) se hace una carga completa de toda la historia. Si el valor es mayor a cero se realiza una carga por días, entre la fecha actual y el número de días que este almacenado en el parámetro hacia atrás. Ejemplo, si el valor del parámetro es dos (2) y la fecha actual es marzo 10 de 2008 entonces la carga se realizará entre el 8 de marzo de 2008 y el 10 de marzo de 2008. Por último, si el valor es negativo (<0) indica que la carga será un reproceso y se cargará el periodo de tiempo especificado por los parámetros *FechaInicialCarga* y *FechaFinalCarga*. Cuando el valor es mayor o igual a cero (0), para la carga del DataMart de Operaciones no se tienen en cuenta los valores almacenados en los parámetros *FechaInicialCarga* y *FechaFinalCarga*.

- **FechaInicialCarga:** Fecha a partir de la cual se va a realizar la carga de Operaciones. Solamente se tiene en cuenta si el valor del parámetro *CargaDatosOperaciones* es menor a cero.
- **FechaFinalCarga:** Fecha final de carga para el tema de Operaciones. Solamente se tiene en cuenta si el valor del parámetro *CargaDatosOperaciones* es menor a cero.

La información extraída de las fuentes se estructura de forma que sea rápida para la lectura en los modelos “Estrellas”. Se conocen como “Estrellas”, porque usualmente tienen un centro llamado “Tabla de Hechos” que corresponde al movimiento o los datos que se quieren analizar, y además tienen unas “puntas” llamadas “Tablas Dimensionales” que usualmente corresponden a los maestros y son utilizadas para analizar la información.

Desde aquí se hace necesario realizar los siguientes procesos para la validación de datos:

- Verificación de campos de la clave.
- Verificación de campos requeridos.
- Se verifica integridad referencia en tablas relacionadas.
- Se evalúan las claves duplicadas.
- Se realizan procesos de conversión de tipos de datos (si es necesario).
- En algunos casos se requieren transformaciones de equivalencias como cambiar un campo por datos más dicentes, Ej.: M y F por „Masculino” y „Femenino”.

Por otro lado, el resultado de cada proceso de carga generará un registro en la tabla *tblLogProcesos*. Esta tabla se creará para tal efecto y se podrá consultar mediante un reporte. Cuando se almacene un registro se guardará la siguiente información en la tabla de procesos:

- Tema: el tema que se carga, ejemplo: clientes
- Proceso: el nombre del proceso, ejemplo: carga dimensión de clientes
- Fecha: fecha de ejecución del proceso

- Resultado: si es exitoso o no
- Registros leídos
- Registros ingresados
- Registros actualizados
- Registros rechazados
- Mensaje

### **Carga para tablas de dimensiones**

En el sistema se tienen los siguientes procesos de carga de tablas dimensionales:

**Tabla 3. Tablas dimensionales**

<b>Dimensionales</b>	<b>Descripción</b>
Origen	Se actualiza dimOrigenDestino
Último Origen	Se actualiza dimOrigenDestino
Destino	Se actualiza dimOrigenDestino
Vehículo	Lentamente cambiante por el campo (campos Disponibilidad y Estado son históricos, los demás se actualizan)
Franja Horaria	Se actualiza dimFranjaHoraria.
Moneda	Se actualiza dimMoneda
Conductor	Se actualiza dimConductor.
Geografía	Se actualiza dimGeografía

<b>Dimensionales</b>	<b>Descripción</b>
UEN	Se actualiza dimUEN
Medida	Se actualiza dimMedida
Servicio	Se actualiza dimServicio
Unidad Transporte	Se actualiza dimUnidadTransporte
Tiempo	Se genera automáticamente dimTiempo, de acuerdo a la primera y última fecha de datos que se encuentran en las fuentes de datos
Documento	Se actualiza dimDocumento
Novedad	Se actualiza dimNovedad
Cliente	Se actualiza dimCliente

### **Carga para tablas de hechos**

En el sistema se tienen los siguientes procesos de carga de tablas de Hechos:

**Tabla 4. Tablas de hechos**

<b>Hechos o Movimientos</b>	<b>Descripción</b>
Operaciones	Contiene el movimiento de las operaciones realizadas para cada cliente, teniendo en cuenta que el detalle son los datos que se presentan del tramo entre el último origen y el siguiente destino, durante el Despacho de cada vehículo. Contiene como medidas regulares el número de Despachos, unidades transportadas, número vehículos, número

	recogidas, número entregas, kilómetros recorridos, unidades solicitadas, unidades cargadas, peso transportado, peso solicitado, tiempo cargue, tiempo tránsito, tiempo descargue, tiempo espera cargue, tiempo espera salida cargue, tiempo espera descargue, tiempo espera salida descargue, tiempo entrega, tiempo promesa entrega, tiempo ciclo.
Novedades	Contiene el movimiento con información concerniente a la cantidad de novedades que se han generado durante el transporte de mercancía.

Finalizada esta etapa de carga de tablas de dimensiones y tablas de hechos, se configura el Modelo estrella de operaciones, para este modelo la tabla de hechos o movimientos que es el centro de la estrella es la factOperaciones y el resto de tablas representan las variables de análisis. Ver Anexo 1.

### 2.2.7. Fuentes de datos de la empresa transportadora.

Las fuentes de información para el Data Mart y que se va a tener acceso desde el servidor son: Sipres, Infocar y Excel.

**INFOCAR:** Contiene información principalmente histórica de las operaciones que se han realizado en Imbocar.

**SIPRES:** Contiene la información que se extrae diariamente de las operaciones de la compañía.

**EXCEL:** Contiene la información Histórica y alguna información actual de las operaciones de algunos clientes.

Las diferencias entre las herramientas Infocar y Sipres son:

INFOCAR	SIPRES
Desarrollo de escritorio	Desarrollo WEB
La conexión de un cliente remoto se debe realizar por medio de terminal Server y vpn, si el cliente de imbocar no permite acceso vpn, imbocar debe adquirir una conexión de banda ancha para asegurar la conexión vpn	Por ser Web, con una salida a Internet normal se puede acceder o desde un café Internet
La creación de un campo se realiza modificando el código	Se le puede asignar un campo de captura especial a un cliente determinado por una opción en el menú que se llama atributos variables (sin necesidad de modificar el código)
La importación de la información se realiza por medio de archivos planos	La importación de la información se realiza por medio de archivos planos, Web Services, archivos XML
No interactúa con dispositivos móviles	Interactúa con dispositivos móviles
Tiene clientes desconectados, es decir si en un sitio remoto no se puede colocar Internet, se puede instalar el Infocar de forma local y realizar procesos de replicación de la información para ser actualizada en el servidor	No permite clientes desconectados, deben ser 100% en línea
Por trabajar algunos clientes desconectados la información no puede estar en tiempo real	La información estará en tiempo real.
El ingreso de información es realizado por el coordinador	Por trabajar con dispositivos móviles el ingreso de información estará en un alto porcentaje en manos del conductor, el coordinador solo programara los viajes.
Por no estar 100% no puede tener manejo de disponibilidad de vehículos	Por estar en línea 100% podrá tener manejo de disponibilidad de vehículos entre las operaciones y también podrá verse por colores el estado de cada uno de ellos

## 2.3. Marco histórico

A continuación se expone el marco donde se abordan los antecedentes históricos de los principales temas de la investigación, esto con el fin de posibilitar la descripción de sus orígenes e identificar de una forma clara los cambios que han tenido a través del tiempo.

### 2.3.1. Referencia histórica de la Inteligencia de Negocios

Con base en aportes del libro “El Arte de la Guerra”<sup>2</sup>, se refleja la importancia de compilar y analizar la información para generar conocimiento, no solo de las debilidades y fortalezas propias, sino también las del oponente en la guerra.

A finales del siglo XX las empresas comienzan a ver la necesidad de recopilar y analizar la información para la toma de decisiones, pero esto se ve afectado

<sup>2</sup> Libro sobre tácticas y estrategias militares, inspirado por Sun Tzu, un famoso autor militar. Se presume que probablemente fue escrito hacia el año 500 antes de nuestra era.

por la poca automatización y la escasez de recursos computacionales, lo que llevaba a que los gerentes y demás ejecutivos tomaran decisiones de forma intuitiva.

Con el crecimiento de las empresas y el surgimiento de los sistemas de información, se fueron adoptando maneras más fáciles y estructuradas para el almacenamiento de los datos, aunque se carecía de una buena infraestructura para almacenarlos y procesarlos, generando así incompatibilidades entre los sistemas lo cual afectaba la toma de decisiones estratégica.

Actualmente las organizaciones han ido adoptando la tecnología como medio para administrar grandes volúmenes de datos, generando el nacimiento y consolidación de nuevas tecnologías de almacenamiento y de bodegas de datos, entre ellas los procesos ETL (Extracción, Transformación y carga de datos), los cuales aumentaron la efectividad y velocidad en la recolección y organización de la información, y en conjunto con tecnologías OLAP para la generación de reportes de manera más eficiente, simplificando así el análisis de la información.

### **2.3.2. Historia del Data Warehouse**

Con el inicio y posterior auge de los sistemas y las computadoras, las organizaciones han usado los datos para atender sus necesidades de información. Algunas han creado accesos directo a la información contenida en las aplicaciones, otras han extraído los datos desde sus bases de datos para combinarlos de variadas formas en un intento por atender a las necesidades de información de los clientes. Estos métodos han evolucionado de acuerdo al incremento general de sus clientes y han generado el manejo de nuevas herramientas de almacenamiento, donde la limpieza y calidad de datos posibilita que los ejecutivos de la empresa tomen decisiones estratégicas.

La gestión administrativa reconoce que una manera de elevar su eficiencia está en hacer el mejor uso de los recursos de información que ya existen dentro de la organización. Sin embargo, a pesar de que esto se viene intentando desde hace muchos años, no se tiene todavía un uso efectivo de los mismos debido a la falta de conocimiento sobre las nuevas tecnologías y sistemas de información. Es por esto que la mayoría de las organizaciones hacen lo posible por conseguir buena información, pero el logro de ese objetivo depende fundamentalmente de su arquitectura actual, tanto de hardware como de software.<sup>3</sup>

Actualmente el Data Warehouse provee un ambiente para que las organizaciones hagan un uso eficiente de la información interna que está siendo administrada por diversas aplicaciones operacionales.

## **2.4. Marco teórico o de referencia**

A continuación se expone el marco en el cual se nombran, definen y explican los aspectos teóricos de la investigación, tanto las herramientas como los aspectos que intervienen en el proceso ETL y el proyecto de Inteligencia de Negocios en la empresa transportadora.

### **2.4.1. Inteligencia de Negocios**

Para que una empresa sea competitiva, las personas que toman las decisiones necesitan acceder rápida y fácilmente a la información de la empresa y esto se realiza por medio de la Inteligencia de Negocios.

---

<sup>3</sup> es.geocities.com/info\_datawarehousing/datawarehousing.html. Visitado el 06 de Abril de 2008.

#### **2.4.1.1. ¿Qué es inteligencia?**

Es la capacidad de aprender, entender o afrontar nuevas situaciones, de aplicar conocimiento para cambiar el hábitat en el que nos encontramos.

#### **2.4.1.2. ¿Qué es negocio?**

El negocio es un ámbito en el que actúa una empresa. Para esto, la organización debe tener claro a qué se dedica y qué aspectos quedan excluidos de su actividad empresarial. Es por esto que cualquier organización debe tener identificados sus procesos de negocio, también llamados procesos estratégicos, así como sus procesos de apoyo.

#### **2.4.1.3. Definición y características de la Inteligencia de Negocios**

Inteligencia de Negocios puede ser tomado de muchas maneras, pero todas tienen algo común: analizar datos para tomar decisiones. Ya que es un término un poco técnico, dentro de esta investigación se tomará la siguiente definición:

*Inteligencia de Negocios es el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (Reporting, análisis OLTP / OLAP, alertas) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.<sup>4</sup>*

---

<sup>4</sup> www.sinnexus.com. Visitado el 03 de Abril de 2008.

En otras palabras, es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.

Este conjunto de metodologías, aplicaciones y tecnologías tienen en común las siguientes características:

- **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.

De acuerdo a su nivel de complejidad se pueden clasificar las soluciones de Inteligencia de Negocios en:

- **Cubos OLAP (On Line Analytic Processing).**
- **Consultas e informes simples (Reporting & Query).**
- **Data Mining o minería de datos:** Las empresas suelen recabar información sobre producción, mercados y clientes, pero en realidad el éxito del negocio depende de la visión para intuir cambios o nuevas tendencias. Las aplicaciones de data mining identifican tendencias y comportamientos para extraer información y descubrir las relaciones en bases de datos que revelen comportamientos poco evidentes.

Dentro de los proyectos de Inteligencia de Negocios unos de los componentes fundamentales y quizás el más crítico es el proceso de integración de datos en un repositorio que permita almacenar la información ya consolidada para ser

explotada por herramientas de análisis. El proceso que soporta esta característica es ETL (Extract, Transform and Load).

**Figura 1. Inteligencia de Negocios<sup>5</sup>**



#### **2.4.1.4. Datos, información y conocimiento**

Para lograr que un proyecto de Inteligencia de Negocios sea exitoso se debe tener en cuenta que lo más importante es el conocimiento generado para la toma de decisiones, sin embargo este término es confundido muchas veces con dato o información, y por esto es importante diferenciarlos.

- **Dato:** Los datos son la mínima unidad semántica, y se corresponden con elementos primarios de información que por sí solos son irrelevantes como apoyo a la toma de decisiones. Los datos pueden ser una colección de hechos almacenados en algún lugar físico como un papel, un dispositivo electrónico o la mente de una persona.
- **Información:** La información se puede definir como un conjunto de datos procesados que tienen relevancia, propósito y contexto, y que por lo tanto son de utilidad para quién debe tomar decisiones, al disminuir su

<sup>5</sup> [www.sinnexus.com](http://www.sinnexus.com). Visitado el 03 de Abril de 2008.

incertidumbre. La información es la comunicación de conocimientos o inteligencia, y es capaz de cambiar la forma en que el receptor percibe algo, impactando sobre sus juicios de valor y sus comportamientos.

*Información = Datos + Contexto (añadir valor) + Utilidad (disminuir la incertidumbre)*

- **Conocimiento:** El conocimiento se deriva de la información, así como la información se deriva de los datos. Para que la información se convierta en conocimiento es necesario realizar acciones como:
  - Comparación con otros elementos.
  - Predicción de consecuencias.
  - Búsqueda de conexiones.
  - Conversación con otros portadores de conocimiento.

#### ***2.4.1.5. Objeto de estudio de la Inteligencia de Negocios***

El principal objetivo de la teoría de inteligencia de negocios es realizar acciones y procesos para tomar decisiones de alta calidad. Estas acciones y procesos incluyen los clientes, competidores, socios de negocios, entorno económico y operacional.

#### ***2.4.1.6. Arquitectura de un proyecto en Inteligencia de Negocios***

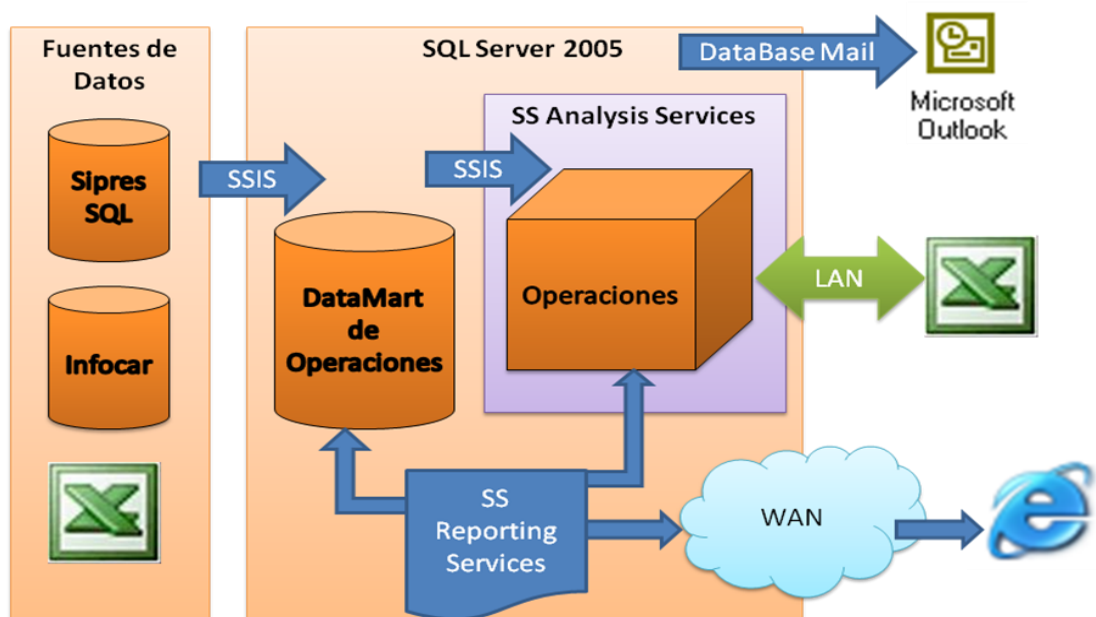
Un proyecto de Inteligencia de Negocios parte de las fuentes de datos de una organización sobre los cuales suele ser necesario aplicar una transformación estructural para optimizar su proceso analítico (en el caso de la empresa de transporte las fuentes de datos son Sipres, Infocar y Excel). Para tal fin se realiza el proceso de ETL. Este proceso actúa como comunicador entre las fuentes de datos y los Data Warehouse o DataMart, y cuyo principal objetivo

consiste en extraer y cargar datos para evitar la saturación de los servidores de la organización.

La información resultante ya depurada y consolidada se almacena en un Data Warehouse, el cual puede servir como base para la construcción de distintos DataMarts los cuales se caracterizan por poseer la estructura óptima para el análisis de los datos de un área respectiva de la empresa (en el caso de la empresa de transporte el área de Operaciones), mediante bases de datos analíticas (OLAP).

Para finalizar se obtiene el producto solución de la Inteligencia de Negocios, datos, información y conocimiento para la toma de decisiones. Ver figura 2, La arquitectura de inteligencia de Negocios de la empresa de transporte.

**Figura 2. Arquitectura de Inteligencia de Negocios de la empresa de transporte.**



#### **2.4.1.7. Inteligencia de Negocios en los departamentos de la empresa de transporte.**

Inteligencia de Negocios como un sistema de solución empresarial da beneficios y actúa sobre los diferentes departamentos de la empresa ya que todos manejan algo en común: datos, información y lo más importante, conocimiento. Es así como Inteligencia de Negocios actúa sobre los departamentos de la empresa:

**Tabla 1. Inteligencia de Negocios en los departamentos de la empresa**

<b>INTELIGENCIA DE NEGOCIOS</b>	
<b>DEPARTAMENTO</b>	<b>FUNCIÓN</b>
Mercadeo y Ventas	Estudia e identifica los comportamientos y necesidades del cliente, así como el análisis de la competencia y productos sustitutos.
Gestión Humana	Gestiona el recurso humano de la empresa, manejando a la vez la relación necesidades del empleado con el cumplimiento de sus labores.
Finanzas y Contabilidad	Administra la actividad microeconómica de la empresa para el desarrollo de procesos financieros en general.
Operación y Logística	Coordina las operaciones de transporte desde el origen de la carga hasta su destino.
Gestión Vehicular	Administra el parque automotor y habilita los vehículos para la carga y transporte de la mercancía.
Seguridad	Maneja los sistemas de seguridad, radiocomunicaciones, monitoreo satelital y el sistema de aseguramiento de las mercancías.

#### **2.4.2. El proceso ETL: definición y etapas**

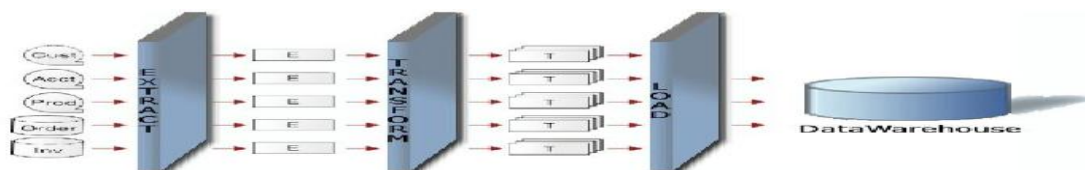
ETL son las siglas en ingles de Extraer, Transformar y Cargar (Extract, Transform and Load). ETL se define como el proceso de extracción, transformación y carga de los datos, que es parte del ciclo de vida de una implementación de BI. ETL permite a las organizaciones mover o extraer datos desde múltiples fuentes, transformarlos, limpiarlos, convertirlos y cargarlos a

una fuente destino, como puede ser un DataMart o un Data Warehouse, todo esto para apoyar un proceso de negocio.

ETL parte de 3 etapas bases las cuales apoyan la implementación de BI:

- **E: Extraer:** Etapa donde se extraen y se obtienen los datos de las distintas fuentes de datos tanto internas como externas de la empresa. La extracción selecciona los datos en un formato listo para iniciar el proceso de transformación.
- **T: Transformar:** Etapa donde se filtra, limpia, depura, homologan y se agrupan los datos para su posterior organización.
- **L: Carga:** Etapa donde se organizan, actualizan y cargan los datos en la base de datos destino.

**Figura 3. Proceso ETL<sup>6</sup>**



El objetivo de ETL es extraer, transformar y cargar los datos directamente a una base de datos (en el caso de la empresa de transporte la base de datos para la carga será el DataMart de Operaciones) la cual contiene tablas especialmente creadas para los datos de paso o temporales (proceso conocido como *staging*), estas tablas se conocen como Tablas Dimensionales y Tablas de Hecho, las cuales pueden ser actualizadas en cada proceso de carga.

Es así como teniendo la información en staging se elabora el proceso de transformación de los datos que posteriormente pasaran al DataMart de

---

<sup>6</sup> [www.wikipedia.org](http://www.wikipedia.org). Visitado el 2 de septiembre de 2008.

Operaciones. La transformación se hará con el lenguaje propio de la base de datos, en el caso del DataMart de Operaciones es SQL 2005. Ya con los datos transformados en la base de datos, se pasa al proceso de carga al DataMart de Operaciones y finalmente se validan los datos de paso si es conveniente.

#### **2.4.2.1. Limpieza de datos**

Dentro del proceso de ETL es importante ejecutar una etapa la cual se encuentra entre la extracción y transformación de datos, a esta etapa se le conoce con el nombre de Limpieza de Datos.

La limpieza detecta y elimina errores, rellena atributos vacíos, y resuelve inconsistencias. En este sentido, la limpieza de datos es necesaria porque si se cargan datos duplicados o valores no asignados no se obtienen datos con calidad y los resultados de análisis van a ser inexactos y poco fiables.

#### **2.4.2.2. La ciencia de limpiar los datos**

Para limpiar los datos se pasan por 6 fases: descomponer, estandarizar, verificar, emparejar, buscar información relacionada y documentar.

Ejemplo: Ricardo Restrepo y Diana de Restrepo referidos por Vélez Freddy CII 16 #1315 ap 109 aa 1234 Cartagena Atlántico.

Estos campos se han ingresado en cinco columnas en la base de datos, llamadas direc1 hasta direc5.

##### **1. Descomponer:**

Destinatario Nombre (1): Ricardo

Destinatario Apellido (1): Restrepo

Destinatario Nombre (2): Diana

Destinatario Apellido (2): Restrepo

Relación con el destinatario: Referidos por

Apellido persona relacionada: Vélez

Nombre persona relacionada: Freddy

Parte inicial de la dirección: Cll 16

Número dirección: #1315

Complementos dirección: ap 109

Apartado aéreo: aa 1234

Ciudad: Cartagena

Departamento: Atlántico

## **2. Estandarizar:**

Cll \_ Calle

#1315 \_ No. 13 15

ap 109 \_ apartamento 109

aa 1234 \_ A.A. 1234

## **3. Verificar:**

Atlántico \_ Bolívar

**4. Emparejar:** Buscar si Ricardo Restrepo o Diana Restrepo aparecen con otros registros y asegurarse de que todos los registros tengan la misma dirección.

**5. Buscar información relacionada:** Buscar en otros sistemas si Ricardo y Diana son casados.

**6. Documentar:** Todos los resultados en los metadatos para la próxima carga.

### ***2.4.2.3. ¿Por qué se tienen que limpiar los datos?***

Los datos que vienen de las fuentes de datos son:

- **Incompletos:** Atributos sin valor, falta de atributos necesarios para el contexto o el valor del atributo se tiene agregado.
- **Con ruido:** Contienen errores o caracteres incorrectos. Ej.: salario = -10
- **Inconsistentes:** Contienen discrepancias. Ej.: edad = 42, fecha de nacimiento: 20/12/1982.

- En unas celdas aparecen primero A, B, C como valores validos y luego aparecen 1, 2, 3.
- Registros duplicados con parte de información diferente.

Todo lo dicho hasta acá sobre el proceso de ETL y sus etapas parte de una metodología global la cual facilita la toma de decisiones en las empresas y provee que los datos se organicen de manera lógica, a esto que nos referimos es a la Inteligencia de Negocios que a continuación se presenta.

### **2.4.3. Data Warehouse**

Data Warehouse es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. La creación de un Data Warehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar un proyecto completo y fiable de Inteligencia de Negocios.

#### **2.4.3.1. Características del Data Warehouse**

En la siguiente tabla se describen las principales características del Data Warehouse:

**Tabla 2. Características del Data Warehouse**

<b>CARACTERISTICAS DEL DATA WAREHOUSE</b>	
<b>INTEGRADO</b>	Los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas.
<b>TEMÁTICO</b>	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.
<b>HISTÓRICO</b>	El tiempo es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente.
<b>NO VOLÁTIL</b>	El almacén de información de un Data Warehouse existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

#### **2.4.3.2. Aportes del Data Warehouse a las empresas**

- El Data Warehouse proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos; obteniendo un valor añadido para el negocio de dicha información.
- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.

En conclusión según lo visto del proceso ETL y el Data Warehouse se puede entender que:

*“Hay una extensa gama de herramientas en el mercado para cumplir esta funcionalidad sobre entornos de tipo Data Warehouse, por lo que se puede elegir el software más adecuado para cada problemática empresarial concreta.”<sup>7</sup>*

<sup>7</sup> www.csi.map.es. Visitado el 08 de Junio de 2008.

## **2.5. Marco legal**

Esta investigación está enmarcada bajo las leyes Colombianas de propiedad intelectual, las cuales se pueden observar continuación:

- Ley No. 23 del 28 de Enero de 1982 sobre derechos de autor.
- Decisión 351 del 17 de Diciembre de 1993 de derechos de autor y derechos conexos.
- Decreto No. 117 del 14 de Enero de 1994, por el cual se reglamenta la Decisión 344 de la Comisión del Acuerdo de Cartagena.

Además de ser realizado bajo las normas establecidas en el Instituto Colombiano de Normas Técnicas ICONTEC.

## **CAPITULO 3**

### **DESCRIPCIÓN TÉCNICA DE LA INVESTIGACIÓN**

#### **3.1. Diseño metodológico**

En el presente capítulo se abordara el diseño metodológico, entendiendo éste como la relación clara y concisa de cada una de las etapas de la investigación, en el cual se remite al lector al método por el que se investigo – tipo y nivel de investigación – y los instrumentos que se utilizaron para la recolección y análisis de la información, los cuales – guiados por los objetivos de la investigación – permitieron identificar los principales riesgos existentes en los procesos de ETL (Extracción, Transformación y Carga) de las fuentes de datos

---

Sipres, Infocar y Excel en el proyecto de Inteligencia de Negocios en la empresa de transporte., con el fin de crear estrategias que los disminuyeran.

### **3.1.1. Tipo de investigación**

La presente investigación es de *Tipo Cualitativo*, debido a que su principal interés es el de describir y examinar cualidades del proceso ETL a la vez que se identifican riesgos en su proceso y se proponen recomendaciones que los disminuyan.

El objetivo de las investigaciones de tipo cualitativo es la construcción de conocimiento en torno a un tema específico, en este caso ETL, en donde se buscan identificar relaciones causales, comprender procesos y generar estrategias de solución, características de la presente investigación. En este sentido Bryman plantea:

“La investigación cualitativa epistemológicamente se preocupa por la construcción de conocimiento... En efecto, problemas como examinar la dinámica y la lógica de los procesos, se convierte en una constante de las diversas perspectivas cualitativas”.<sup>8</sup>

Dicho esto, es importante plantear que las investigaciones de tipo cualitativo utilizan como una de sus fuentes principales de recolección de información instrumentos como la encuesta y la entrevista, esto con el fin de conocer la situación actual del objeto de estudio y crear conocimiento del mismo. Por consiguiente, la primera actividad de estrategia de análisis que se desarrollará dentro de la investigación es dar a conocer mediante una entrevista estructurada cómo se han venido aplicando los procesos ETL en Inteligencia de Negocios en algunas empresas de la ciudad de Medellín, esto con el fin de poseer información significativa que aporte a la investigación y contribuya a los

---

<sup>8</sup> BRYMAN citado por RODRÍGUEZ G, Gregorio et al en: Metodología De La Investigación Cualitativa. Aljibe. Málaga 1996. p.37.

procesos de ETL de la empresa de transporte, además los resultados que arrojen las entrevistas sumadas al conocimiento que se tenga del negocio permitirá disminuir las probabilidades de riesgo que ETL puedan tener en el proyecto de Inteligencia de Negocios de la empresa de transporte.

### **3.1.2. Nivel de investigación**

La presente investigación, la cual tiene como objetivo identificar los riesgos de los procesos ETL en el proyecto de Inteligencia de Negocios en una empresa de transporte, se desarrollara mediante un *Nivel Descriptivo*, este tipo de investigaciones procura brindar una comprensión del funcionamiento de un proceso y de las maneras en que se comportan los factores y elementos que lo componen y desde aquí aplicar el análisis de información describiendo todas su dimensiones, en este caso se describe el objeto a estudiar que es el proceso de ETL.

Los estudios de este nivel, son los que abordan situaciones que habitualmente ocurren en condiciones naturales más que los que se basan en situaciones experimentales, y generalmente no parten de hipótesis explícitas, pero sí de preguntas de investigación que actúan como herramientas orientadoras de todo el proceso investigativo. En este sentido, la presente investigación es de nivel descriptivo, ya que si bien existen teorías que describen los riesgos que existen en los procesos de ETL, y que sirven como punto de referencia, esta investigación se plantea como objetivo identificar el proceso, riesgos y estrategias para disminuirlos al interior del manejo de datos de una empresa de transporte.

#### **3.1.2.1. Pregunta de investigación**

¿Cuáles son los principales riesgos que existen en los procesos de ETL (Extracción, Transformación y Carga) de las fuentes de datos Sipres, Infocar y

Excel en el proyecto de Inteligencia de Negocios en la empresa de transporte y cómo se pueden disminuir para que no afecten las operaciones del negocio?

### **3.1.3. Instrumentos de recolección de información**

Los instrumentos de recolección de información utilizados en la presente investigación son:

- Formato de encuesta: Consiste en la construcción planeada de un formato de preguntas a responder; este instrumento no pretende ser un sistema de información como tal sino un punto de recolección de información para una posterior entrevista y análisis de datos.
- Entrevista estructurada: Encuentro cara a cara entre el investigador y el sujeto entrevistado que tiene como objetivo recolectar información de un tema específico, en este caso los procesos de ETL en la Inteligencia de Negocios. La entrevista se basa en un marco de preguntas predeterminadas o formato de encuesta la cual guía todo el proceso de entrevista, es decir, las preguntas se construyen antes de que inicie la entrevista.

Es importante destacar que se utilizaran las siguientes fuentes de información, que aunque no sean instrumentos propiamente dichos para la investigación de tipo cualitativo, sirven para la recolección de información válida y sustancial:

- Revisión de otras investigaciones donde se aborda el proceso ETL y el Mapeo de Datos.
- Reuniones con los asesores de la empresa de transporte y con los empleados de la empresa proveedora de servicios, los cuales en el presente desarrollan un proyecto de Inteligencia de Negocios.

Cabe destacar que los instrumentos de recolección de información son el medio o medios por los cuales el investigador toma o recopila una información

para luego procesarla y obtener resultados que permitan lograr el objetivo de la investigación.

#### **3.1.4. Metodología**

La metodología de la presente investigación se basó en un estudio cualitativo y descriptivo del proceso ETL, sus características, etapas, riesgos y estrategias de disminución en el proyecto de Inteligencia de Negocios en una empresa de transporte. Para este fin se llevaron a cabo 4 actividades que orientaron la recolección, análisis y síntesis de información, lo cual permitió cumplir los objetivos propuestos de la investigación.

##### **ACTIVIDAD 1:**

Se realizó una exploración teórica de los conceptos de ETL, esto ayudó a identificar sus características y etapas e ir describiendo cuáles podrían ser los riesgos más comunes en su implementación.

##### **ACTIVIDAD 2:**

Se visitaron empresas que tenían implementadas soluciones de Inteligencia de Negocios en sus procesos de ETL y con base a esta experiencia se llevó a cabo un análisis de información, lo que permitió aplicar lo aprendido e identificado al interior de la empresa transportadora.

##### **ACTIVIDAD 3:**

Fui acompañada y asesorada por el personal de la empresa transportadora para el análisis de las fuentes de información empleadas dentro de la empresa, (Sipres, Infocar y Excel), esto con el fin de conocer su estructura y cómo son los modelos de base de datos en el proceso ETL.

##### **ACTIVIDAD 4:**

En esta actividad se llevó a cabo:

- Análisis de resultados del estudio de las fuentes y la calidad de los datos; a partir de aquí se describieron posibles riesgos de los procesos de ETL.
- Identificación de riesgos; de aquí se planteó estrategias para disminuir los mismos.
- Generación de resultados de la investigación de acuerdo con los objetivos para la entrega de productos.

## **CAPITULO 4**

### **ANÁLISIS Y RESULTADOS DE LA INVESTIGACIÓN**

Para que las organizaciones puedan aspirar a tener una alta productividad y un desempeño eficiente que las introduzca en la excelencia deben atender las necesidades de desarrollo de sus estructuras de procesamiento de datos e información, esto con el fin de elevar la eficiencia y eficacia en el sentido organizacional, además de posibilitar que se organice la información de manera más inteligente, estratégica y segura donde la toma de decisiones es imprescindible para el mejoramiento. Pero toda toma de decisiones implica aceptar un riesgo, lo que es indudable es que el objetivo es minimizar ese riesgo, aquí es donde entran en juego el proceso de ETL, donde la disminución

de los riesgos proporcionara información de relevancia logística para la empresa transportadora, objetivo de la presente investigación.

En este sentido, se ha desarrollado la presente investigación, la cual describe los procesos de ETL que ejecutan las fuentes de datos en una empresa de transporte, lo cual permitió conocer su estructura y funcionamiento. A la vez se identificaron los errores en su implementación y propusieron y establecieron estrategias para el manejo y disminución de los mismos.

Desde esta perspectiva, se presenta a continuación el análisis y resultados de la investigación llevada a cabo al interior de la empresa transportadora teniendo como aporte los objetivos específicos de la investigación.

Es importante aclarar que para el inicio del presente análisis se dará a conocer cómo se han venido aplicando los procesos ETL en Inteligencia de Negocios en algunas empresas de la ciudad de Medellín, esto con el fin de poseer información significativa que aporte a la investigación y contribuya a los procesos de ETL de la empresa transportadora, además los resultados que arrojen las entrevistas aunadas al conocimiento que se tenga del negocio permitirá disminuir las probabilidades de riesgo que ETL puedan tener en el proyecto de Inteligencia de Negocios de una empresa de transporte.

#### ***4.1. Aplicación de procesos de ETL en Inteligencia de Negocios en empresas de la ciudad***

Se hace necesario para los objetivos de la presente investigación comprender los riesgos que pueden presentar el proceso de ETL en la toma de decisiones estratégicas al interior de algunas organizaciones del medio. La actividad en este sentido que se hizo dentro de la investigación fue la visita a empresas que han implementado proyecto de Inteligencia de Negocios y que han manejado el proceso ETL, las cuales les han permitido tener una visión global de la

información de forma compartida y distribuida para todos los departamentos de la organización.

En este sentido, en el presente punto del análisis y resultados de la investigación se nombraran algunas experiencias que las empresas visitadas tuvieron durante y después de implementar el proceso de ETL dentro del área de Inteligencia de Negocios desde 11 enfoques de preguntas:

**1. Soluciones de Inteligencia de Negocios implementadas al interior de las organizaciones:** Dentro de las soluciones informáticas implementadas por las organizaciones se encuentra el manejo de herramientas como:

- IBM (Base de datos – Software de ETL).
- Hyperion (Cubos y Visualización).
- Plataforma MS SQL SERVER 2005.
- Análisis OLAP.
- Reporteo: Informes que tienen una muy buena presentación y lee los datos desde el cubo o fuentes de datos como Data Mart y fuentes de datos relacionales.
- Balance Score Scard.

Todas estas herramientas sirven al interior de las organizaciones para la clasificación y administración estratégica de la información procedente de las bases de datos.

**2. Metodologías y tecnologías empleadas en el área de Inteligencia de Negocios:** Dentro de las metodologías y tecnologías informáticas orientadas a mejorar la gestión inteligente de la organización se encuentran la implementación de las siguientes:

- Metodología de desarrollo de DataMarts propia de la organización.
- Base de datos relacional – DB2 8.2 para Windows.
- ETL – Data Warehouse Center.

- Hyperion Essbase – Creación de cubos (modelos multidimensionales de análisis).
- Hyperion Analyzer – Visualización y análisis de información contenida en cubos.
- Metodología OLAP.
- Reporteo.
- Balance Score Scard.
- Microsoft SQL Server.
- La metodología que se aplica es de Ralph Kimball.

Todas estas metodologías y tecnologías informáticas permiten a los equipos directivos de las organizaciones controlar los negocios desde la administración de las bases de datos en cada área y así tomar decisiones estratégicas.

**3. Herramientas empleadas para el proceso ETL:** Al interior de las organizaciones visitadas se emplean los siguientes procesos de ETL para la integración de los datos:

- ETL – Data Warehouse Center de IBM para Windows.
- Visual Fox Pro para preparar información en formato adecuado para algunos modelos de DataMart.

Estos procesos de ETL permiten a las bases de datos de las organizaciones extraer datos del entorno origen, transformarlos según las necesidades estratégicas del negocio y cargar estos datos en los entornos destino.

**4. Fuentes de integración en ETL:** Las fuentes utilizadas para la integración de datos en los procesos de ETL al interior de las organizaciones visitadas son las siguientes:

- Información del ERP (Enterprise Resource Planning).
- Información de Excel.

- Información de Sistemas Satélites o complementarios del ERP.
- Extracciones de base de datos Informix, SQL.
- Archivos planos.
- Archivos Excel.

Estas fuentes sirven para que la información proporcionada por las bases de datos se integre en un mismo destino y de una forma estratégica.

**5. Administración de procesos ETL:** Según la opinión generalizada dada por los ingenieros perteneciente al área de Inteligencia de Negocios de las organizaciones visitadas, refiriéndose a la administración de los procesos de ETL, se puede determinar que: *“Los procesos son adecuados al nivel de detalle de información que requiere el DataMart y se ejecutan con la periodicidad y validez requerida de acuerdo al modelo, además son contruidos en la herramienta de Data Warehouse de IBM”*. Así mismo, la administración y ejecución de los procesos de ETL al interior de las organizaciones pasa por los siguientes momentos:

- Se definen y catalogan las fuentes de datos.
- Se definen los destinos de información.
- Se define el proceso de transformación de la información y tablas temporales que contendrán los datos en proceso de transformación y limpieza de acuerdo a las reglas acordadas con el usuario de la información.
- Se define la periodicidad del proceso y su automatización.
- Se ejercen los controles para su ejecución, mantenimiento, adaptación y desempeño.
- Se evalúan las inconsistencias.

Desde este punto de vista, cada proceso es valorado y administrado de acuerdo a las necesidades de la organización y a las decisiones que se toman.

**6. Riesgos en la implementación de ETL:** Dentro de los riesgos más sobresalientes al momento de implementar procesos de ETL al interior de las áreas de Inteligencia de Negocios de las organizaciones visitadas se encuentran los siguientes:

- La no implementación de controles adecuados de calidad que garanticen que la información que se extrae de las fuentes corresponda a la almacenada en el DataMart.
- Que la calidad de datos no sea óptima.
- Realizar cambios innecesarios debido al crecimiento o cambios que surjan en el negocio.
- Realizar una mala elección en las fuentes de información con respeto a su calidad y correcta actualización.

La evaluación minuciosa de los riesgos presentes en el proceso de ETL permitirá en un plazo de tiempo razonable que el mismo se mejore continuamente y se planteen así estrategias de solución adecuadas acordes.

**7. Inconvenientes en la implementación antes y después de ETL y etapas de solución:** Dentro de los inconvenientes y estrategias de solución más relevantes al momento de implementar procesos de ETL al interior de las áreas de Inteligencia de Negocios de las organizaciones visitadas se encuentran los siguientes:

- Falla en la de calidad de datos en las fuentes de información. La estrategia de solución para éste inconveniente fue involucrar al usuario clave del área para que asimilara el proceso de construcción y calidad del DataMart y lo irradiara en su área, ejerciendo de esta forma un proceso de calidad desde la fuente (proactivo y no reactivo).

- La homologación de conceptos no era clara. Desde acá se creó una metodología propia, en donde se da una conceptualización a usuarios mostrando que se necesita y que se puede hacer.
- Entrevistas con usuarios líderes y gerentes.

La evaluación de los inconvenientes aunada con estrategias de solución efectiva y eficaz permitirá en un futuro que los procesos de ETL se mejoren continuamente y se planteen estratégicamente así soluciones adecuadas.

**8. Problemas y mitigación en los procesos de ETL:** Dentro de los problemas de ejecución y estrategias de mitigación más relevantes al momento de implementar procesos de ETL al interior de las áreas de Inteligencia de Negocios de las organizaciones visitadas se encuentran los siguientes:

- Fallas en la ejecución de los ETL en procesos programados. La estrategia de mitigación para éste problema fue programar alertas vía e-mail de problemas en el proceso e incorporar monitoreo diario de los mismos.
- Problemas de conectividad con la fuente de datos. Desde acá se crearon controles de ejecución que impidieran la ejecución parcial del modelo cuando se detectasen problemas de comunicación.
- Problemas en la calidad de datos. Debido a que es uno de los problemas más comunes en el proceso de ETL, se implementaron como estrategias de mitigación las siguientes:
  - Que el usuario se sienta dueño de los datos.
  - Mantener actualizado el sistema de información.
  - Tener especial cuidado en matrícula de clientes y productos.
  - Tener varias bases de datos con tablas que contiene la información replicada, la cual se debió realizar evaluación de la información de cada fuente y decidir extraer la información de la base de datos más actualizados y confiables.

La evaluación de los riesgos permitirá en un plazo de tiempo razonable que los procesos se mejoren continuamente y se planteen estratégicamente así soluciones adecuadas.

**9. Aprendizajes (ensayos, errores y aciertos):** Después de haber considerado en los anteriores puntos de pregunta los procesos de ETL y las estrategias implementadas dentro del área de Inteligencia de Negocios en las organizaciones visitadas, se enfatizó en los aprendizajes a los cuales llegaron las mismas en su implementación.

Dentro de lo que no volverían a hacer se encontró:

- No iniciar un proceso de construcción de ETL sin conocer claramente los requerimientos del usuario o sin aclarar el alcance de la información y los datos que requieren.

Dentro de lo que volvería a hacer igual se encontró:

- Aplicar controles de verificación de los totales de la información que están en los sistemas fuentes y en el DataMart para garantizar al usuario la integridad de la información.
- Considerar al usuario como ente primordial en el proceso ETL.

El aprendizaje y la asimilación del mismo en el manejo de procesos y herramientas de ETL permiten que las organizaciones a través del tiempo mejoren sus metodologías, tecnologías y prácticas en el manejo inteligente de los datos.

**10. Razones por las que las empresas deben utilizar la Inteligencia de Negocios:** A manera de conclusión se pueden emplear como razones claves para un área de Inteligencia de Negocios las siguientes sugerencias:

- Hacer uso “inteligente” de la información que generan internamente los sistemas transaccionales de la compañía, con el fin de apoyar el proceso de toma de decisiones.
- Homologar la información que servirá para la toma de decisiones, buscando un lenguaje organizacional entendible por todos y no meramente departamental.
- Facilitar el acceso a la información dentro de la organización, democratizando el uso de la misma.
- Servir como control de calidad a la información que se genera en los sistemas fuente de información.
- Contar desde las áreas con información única y entendible en lo relacionado al manejo de la Inteligencia de Negocios.
- Siempre se deben tener muy en cuenta a los usuarios líderes de cada área en la cual se vaya a implementar Inteligencia de Negocios, debido a que son ellos los que nos brindan orientación sobre los procesos y nos ayudaran a realizar la verificación de la información que se esta cargando en la bodega de datos.

**11. Razones para la implementación de ETL:** A modo de retroalimentación y aprendizaje vicario entre las organizaciones visitadas y la investigadora, se plantea que para la implementación de procesos de ETL es útil tener en cuenta las siguientes razones:

- El proceso de ETL también se puede utilizar para los procesos de migración de Datos.
- Además son útiles para procesos repetitivos evitando desarrollar trabajos manuales para transformar información.

***4.2. Cinco principales etapas para el desarrollo exitoso de los procesos de ETL en el proyecto de Inteligencia de Negocios***

A modo de introducción práctico, y retomando elementos del Marco Teórico, se entenderá ETL como el proceso destinado a extraer información desde las fuentes de datos de una empresa, para así limpiarlos, transformarlos a los formatos de utilización y cargarlos en el repositorio final; todo esto con el fin de organizar los datos de la empresa y así optimizar la toma de decisiones. Vemos desde acá que ETL más allá de su definición es un proceso que consta de múltiples etapas en su implementación las cuales son necesarias clarificar, por esto a continuación se nombra cada una y se especifica la forma en la cual se implementa dentro de la empresa de transporte,

ETAPA	DESCRIPCIÓN
<b>EXTRACCIÓN DE DATOS</b>	<p>Etapa en la cual se extraen los datos desde las fuentes de datos: Sipres, Infocar y Excel. Estas fuentes cumplen la función de guardar información básica de la empresa. Los procesos se correrán diariamente por medio de la herramienta Integration Services, se extraerán los datos necesarios desde las fuentes, los validarán y estructurarán en tablas de hechos y tablas dimensionales que se almacenarán en el motor relacional de SQL Server.</p> <p>Extraer significa en términos simples llevar los datos (variables de análisis) de un lugar a otro; el propósito es leer y mover los datos de las fuentes de datos a tablas bases con prefijo SQL (se guardan los datos como vienen, ya en la transformación se cambian y estandarizan), todos estos datos a su vez se guardan</p>

	<p>dentro del DataMart de Operaciones. Para llevar cabo este proceso es importante nombrar que para extraer los datos de las fuentes de datos, se deben definir 2 restricciones: la fecha de la carga y las variables de análisis a requerir, esto se hace con el fin de no cargar información innecesaria.</p> <p>Dentro de la extracción de datos se realizan las siguientes operaciones:</p> <ul style="list-style-type: none"> <li>• Levantamiento de información: Contempla la definición de los datos que requiere el nuevo sistema y la identificación de las fuentes de datos.</li> <li>• Análisis de calidad: Se verifica la calidad de los datos, con el propósito de establecer los tiempos reales para las etapas de, Limpieza y Transformación/Mapeo de Datos.</li> </ul> <p>El proceso a realizar básicamente en la empresa será el de llevar los datos que provienen de las distintas fuentes de datos de la compañía a un conjunto de tablas en SQL Server 2005 mediante procesos de extracción que se diseñarán para este propósito con la herramienta de ETL Integration Services 2005.</p>
ETAPA	DESCRIPCIÓN
LIMPIEZA DE DATOS	<p>La limpieza de datos constituye una de las etapas dentro de la transformación de datos. Esta etapa contempla la depuración de los datos, es decir, entre la extracción (donde se crean las tablas con prefijo SQL) y la transformación (donde se crean las tablas con prefijo TMP: Tablas Temporales) existe un proceso donde los datos son limpiados para que la información que se vaya cargando y transformando sea de mayor calidad.</p> <p>El propósito de la limpieza consiste en detectar y eliminar errores, rellenar atributos vacíos y resolver inconsistencias; además, corregir, estandarizar y completar los datos.</p> <p>En sí, la limpieza de datos dentro del proceso ETL identifica:</p> <ul style="list-style-type: none"> <li>• Datos redundantes.</li> </ul>

	<ul style="list-style-type: none"> <li>• Valores atípicos.</li> <li>• Valores perdidos.</li> </ul> <p>Es necesario anotar que los datos que no pueden ser limpiados ingresan a un campo dentro de la tabla TMP llamado “Tipo de Inconsistencias” es decir, la información acerca de los registros que poseen problemas de calidad de datos se guardará en una tabla de inconsistencias que se encontrará dentro de la base de datos del DataMart de Operaciones, a partir de esta tabla se podrá consultar mediante un reporte que se elaborará en Reporting Services la información de los datos inconsistentes, de manera que el administrador del sistema pueda visualizar que datos no cumplieron con la validación de calidad y pueda repararlos desde las fuentes de datos y luego de reparados vuelvan a ser cargados al DataMart.</p>
--	---

ETAPA	DESCRIPCIÓN
<b>TRANSFORMACIÓN/ MAPEO DE DATOS</b>	<p>Etapa en la cual se transforman los datos (variables de análisis) para homologarlos y estandarizarlos.</p> <p>Esta etapa busca que los datos transformados se adapten al formato preciso o modelo lógico de la fuente destino, la cual en la empresa es el DataMart de Operaciones.</p> <p>Esta etapa también contempla la definición de los diccionarios con las reglas de transformación y el mapeo de campos entre las fuentes de origen y destino donde se realizan las relaciones necesarias entre las tablas bases SQL para la creación de las tablas temporales con</p>

prefijo TMP, que permitan obtener los datos indispensables para cargar la Tabla de Hecho y Tablas de Dimensiones. Es decir, los datos que se encuentran en las tablas SQL se insertan e integran dentro de una tabla TMP mediante la herramienta de ETL Integration Services 2005. Desde este proceso la Tabla de Hecho y las Tablas de Dimensiones se cargan de datos.

Es importante nombrar que también con esta etapa se inicia el proceso de calidad de datos.

ETAPA	DESCRIPCIÓN
CARGA	<p>Etapa donde se cargan los datos resultantes (Tabla de Hecho y Tabla de Dimensiones) en el almacén de datos históricos individuales de la empresa de transporte (DataMart de Operaciones) y en las aplicaciones OLAP o Cubo.</p> <p>El propósito de esta etapa es definir procesos de carga de datos y sincronización. Es importante que la carga de datos se realice de forma adecuada, porque si se cargan datos duplicados o valores no asignados no se obtienen datos con calidad y los resultados de análisis van a ser inexactos y poco fiable.</p> <p>Dentro de la carga de datos se realizan las siguientes operaciones:</p>

- Cargas reales: Se realiza la carga real de los datos desde el antiguo sistema al nuevo sistema.

ETAPA	DESCRIPCIÓN
CONCILIACIÓN/ VALIDACIÓN	<p>En esta etapa se identifican problemas en los datos no detectados en las etapas anteriores y se realiza la revisión final de la información cargada al DataMart de Operaciones.</p> <p>Existe conciliación/validación por:</p> <ul style="list-style-type: none"> <li>Tipo de dato.</li> <li>Datos en blanco.</li> <li>Datos nulos.</li> <li>Cantidad de caracteres.</li> <li>Precisión de longitud de datos.</li> </ul>

#### ***4.3. Riesgos más comunes en el diseño y desarrollo de los procesos de ETL para el proyecto de Inteligencia de Negocios de una empresa de transporte donde se proponen recomendaciones que los disminuyan***

El siguiente punto de análisis y resultado de la investigación plantea en una vía los riesgos más comunes identificados en el proceso de ETL para el proyecto de Inteligencia de Negocios de la empresa y en otra vía las recomendaciones para disminuirlos, esto con el fin de cumplir los objetivos específicos de la investigación y responder la pregunta de investigación del diseño metodológico.

- Se caiga el servicio del motor de bases de datos de SQL Server, de las fuentes de datos.

La estrategia de disminución para este riesgo es contar en la empresa de transporte con un motor SQL Server certificado en calidad por la firma creadora, además que sea legal y permita la actualización, todo esto para que soporte el proceso de ETL y la arquitectura del DataMart de

Operaciones, además que se integre a las operaciones realizadas por las fuentes de datos Sipres, Infocar y Excel.

- No realizar una limpieza de datos la cual traería como consecuencia que la toma de decisiones del negocio se haga de forma incorrecta, a la vez que se contaría en la bases de datos con datos redundantes, valores atípicos y valores perdidos lo cual crearía desconfianza.

La estrategia de disminución para este riesgo es que cada vez que se extraigan, transformen y carguen datos de las fuentes de datos internas y externas de la empresa de transporte se realice el proceso de limpieza total el cual se puede programar en la arquitectura de ETL. En este sentido, la limpieza de datos es necesaria porque si se cargan datos duplicados o valores no asignados no se obtienen datos con calidad y los resultados del proceso van a ser inexactos y poco fiables.

- El proceso y las etapas de ETL no estén bien definidos.

La estrategia de disminución para este riesgo es contar primeramente con personal especializado en el proceso de ETL. Además antes de montar la arquitectura el ETL y el DataMart de Operaciones de la empresa de transporte es necesario contar con los datos necesarios del negocio con los cuales se construirán las variables de análisis, las tablas dimensionales y las tablas de hechos ya que estas son necesarias e imprescindibles dentro del proceso ETL. Es muy importante para disminuir este riesgo definir el qué, cómo y cuándo se hará la extracción, transformación y carga de datos.

- Dar un tratamiento diferente a los datos de las fuentes externas y de las fuentes internas, es decir suponer que los datos de las fuentes internas no están contaminados y no darles el tratamiento adecuado, o suponer que los datos de las fuentes externas vienen ya transformados.

La estrategia de disminución para este riesgo es pasar todos los datos por un proceso de limpieza y homologación, el cual asegure que la carga de

datos se haga con calidad. Es importante recordar que en ETL y en la Inteligencia de Negocios no se puede dar nada por supuesto, todo debe ser planeado, medido y verificado.

- No realizar chequeos constantes en los procesos de conciliación/validación de datos.

La estrategia de disminución para este riesgo es realizar con frecuencia procesos de conciliación/validación que además de purificar los datos antes de ponerlos en el DataMart de Operaciones, también posibilite realizar mensual o trimestralmente auditorias de calidad de datos, es decir, cuanto más frecuentes sean las comprobaciones, mejor será la calidad de los datos.

- No capacitar a los usuarios directos del proceso de ETL en el manejo básico del sistema.

La estrategia de disminución para este riesgo es capacitar a los usuarios que tienen que ver con el proceso ETL de la empresa de transportes desde una forma sencilla pero práctica en el manejo básico del proceso, ya que usuarios capacitados es muy posible que no cometan errores a la hora de introducir datos.

- Fallas en la ejecución de ETL en procesos programados.

La estrategia de disminución para este riesgo es incorporar un sistema de monitoreo diario del proceso y programar alertas al responsable de este subproceso a través de vía e – mail, para que desde acá se detecte el error en la extracción de datos y poder corregirlo óptimamente.

- Problemas de conectividad con la fuente de datos.

La estrategia de disminución para este riesgo es que cuando se detecten problemas de conectividad, automáticamente se impide la ejecución parcial

del proceso, para así captar el error, devolver el proceso de conectividad y mitigarlo, esta acción permite ahorrar tiempo de ejecución.

- Problemas en la calidad de datos: Debido a que es uno de los problemas más comunes en el proceso de ETL.

Las estrategias de disminución para este riesgo son:

- Que el usuario se sienta dueños de los datos.
  - Mantener actualizado el sistema de información mediante validación en la captura y alimentación de datos, además de la implementación de chequeos de exactitud y pertinencia donde se verifique que los datos si pertenezcan al proceso y a la base de datos con la que se este trabajando.
  - Se debe evitar que los archivos de Excel puedan ser modificados o eliminados por personas no adecuadas, para ello se propone que exista un administrador del sistema quien sea responsable de la manipulación de los datos contenidos en estos archivos.
  - Definir políticas de Backups y actualización de bases de datos, realizando evaluación y verificación de la información de cada fuente para garantizar que siempre se tenga una versión única, confiable y valida. en este sentido, en la empresa se hará Backup en la noche para que no afecte procesos diurnos.
- Suponer que la calidad de datos es responsabilidad única del área de Inteligencia de Negocios. En algunas organizaciones existe la tendencia a suponer erróneamente que el área de sistemas es la única responsable de la calidad de los datos, cuando esta tarea es responsabilidad de todas las áreas del negocio; se sabe que el tratamiento de datos es función del área de Inteligencia de Negocios, pero la calidad de datos es responsabilidad de todos.

En este sentido, la estrategia de disminución de este riesgo para este caso es emplear procedimientos de estandarización – como las auditorias – para

todas las áreas de la empresa los cuales definan cómo se van manejar los datos y cómo se va a asegurar la calidad de los mismos, ya que si los datos no son correctos, los resultados no serán precisos.

- Traer datos que contengan fechas más antiguas de las que necesitamos cargar o campos que no se estén requiriendo, todo esto puede demorar la carga de datos.

La estrategia de disminución para este riesgo es emplear los cuatro tipos de carga que se definieron en el Documento de Especificaciones DataMart de Operaciones las cuales son:

- **Carga de datos diferencial:** Periódicamente se cargan solamente los registros con fecha mayor a la última fecha que se tiene de la última carga.
- **Carga de datos completa:** Se carga toda la información del DataMart de Operaciones borrando toda la información previa que existía en el mismo y volviéndola a cargar, de acuerdo a los datos existentes en ese momento en las fuentes de datos.
- **Carga de datos de un periodo determinado:** Se establece una fecha inicio y una fecha fin de carga de datos, este tipo de carga de datos es útil principalmente para el reproceso de datos de determinadas fechas. Para este proceso primero se borran los datos que existen para ese rango de fechas y luego se cargan los datos para ese mismo rango desde las fuentes de datos
- **Carga de datos de los últimos periodos:** Se carga el DataMart de Operaciones teniendo en cuenta la cantidad de días desde hoy hacia atrás de datos que se desea cargar o actualizar, para este proceso primero se borran los datos que existen en el periodo de tiempo que se definió y luego se procede a realizar la carga de datos de ese mismo periodo de tiempo.

- Redundancia e inconsistencia de datos: Debido a que los archivos que mantienen almacenados los datos son creados por diferentes fuentes internas y externas a la empresa, existe la posibilidad que si no se controla detalladamente el almacenamiento, se pueda originar un duplicado de información, lo cual generaría una redundancia de datos y aumentaría el tiempo en el cual se realiza la carga y acceso a los datos. Además puede originarse inconsistencia de datos, es decir diversas copias de un mismo dato no concuerdan entre sí, por ejemplo: que se actualice el teléfono de un transportador de la empresa en una fuente y que en otra fuente permanezca el anterior.

La estrategia de disminución para este riesgo es contar con un control riguroso de sistema, el cual además de sincronizar de manera paralela los datos de las fuentes de datos Sipres, Infocar y Excel para prevenir redundancias, controlará los datos que se extraen, transforman y cargan al DataMart de Operaciones. Cabe anotar también que cuando se tiene un adecuado control y conocimiento de las fuentes externas se puede disminuir la probabilidad que se tengan datos incorrectos.

- Dificultad para tener acceso a los datos necesarios para responder una pregunta de negocio: El DataMart de Operaciones de la empresa de transporte como un sistema de base de datos debe contemplar un entorno de datos que le facilite la toma de decisiones y el manejo de los datos desde las variables de análisis. Es decir, las variables de análisis que se formulan para la creación de las Tablas de Dimensiones, las cuales facilitan responder las preguntas del negocio y la arquitectura que se monte en el DataMart de Operaciones deben facilitar el acceso a los datos y la toma de decisiones en el negocio. Para un ejemplo supongamos que el gerente necesite averiguar los nombres de todos los clientes a los cuales se les haya transportado mercancía en la ciudad de Medellín en el periodo de enero – febrero de 2007. El usuario final pide al área de Inteligencia de Negocios que genere la lista correspondiente. Puesto que esta situación no

fue prevista en el diseño del sistema, no existe ninguna aplicación de consulta que permita este tipo de solicitud, esto ocasiona una deficiencia del sistema y por ende un riesgo.

Desde esta perspectiva, las estrategias de disminución para este riesgo son:

- Primero: Formular de manera clara y coherente las variables de análisis del negocio, esto se hace sentándose cara a cara con el usuario final.
  - Segundo: Construir la arquitectura que contenga las variables de análisis formuladas en el paso anterior, y que soporte el proceso ETL.
- Riesgo en la actualización de datos concurrentes: De manera inocente suele suceder que algunas veces para mejorar el funcionamiento global del sistema y obtener un tiempo de respuesta más rápido, se permite que múltiples usuarios actualicen los datos simultáneamente. En un entorno así la interacción de actualizaciones concurrentes puede dar por resultado datos inconsistentes.

La estrategia de disminución para este riesgo es contar con la supervisión de un experto el cual supervise, valga la redundancia, el proceso de actualización y carga de datos.

- Problemas de seguridad: La información que se maneja en la empresa es importante, aunque unos datos lo son más que otros, por eso se debe contar con un control de acceso a los mismos. No todos los usuarios pueden visualizar alguna información, por tal motivo para que el DataMart de Operaciones sea confiable debe mantener un grado de seguridad que garantice la protección de los datos.

La estrategia de disminución para este riesgo es contar con un supervisor al cual se le asignara una contraseña que no caducará; como el supervisor va a tener conexión al DataMart de Operaciones solo tendrá permisos de lectura sobre la misma mas no de escritura, por lo tanto no podrá modificar los datos.

#### ***4.4. Proceso de mapeo inicial de los datos y fuentes de destino en los procesos de ETL***

Los campos cuyo valor es generado, obtienen su fuente a partir de cálculos que se crean a través de procedimientos almacenados en el motor de bases de datos de SQL Server 2005 (Database Engine). Ver Anexo 3.

#### ***4.5. Estrategias para manejar los problemas de calidad de datos en el proceso ETL***

Antes de iniciar es importante nombrar cuales son los problemas que podemos encontrar con los datos, los cuales afectarían su calidad:

- Demasiados datos sin filtrar.
- Datos corruptos o con ruido.
- Datos redundantes.
- Datos irrelevantes.
- Excesiva cantidad de datos.
- Atributos perdidos.
- Valores perdidos.
- Poca cantidad de datos.
- Datos fracturados.
- Datos incompatibles o que no concuerdan.
- Múltiples fuentes de datos.
- Inexactitud o información sucia.

Cabe anotar que para prevenir problemas en la calidad de datos en el proceso ETL en la empresa es necesario realizar un sistema de auditoria y control donde se deben tener en cuenta las siguientes preguntas:

- ¿Cuántos datos son necesarios?

- ¿Cuántos registros o filas?
- ¿Cuántos campos o columnas?
- ¿Cuánta historia?
- ¿Cuál fue el último proceso que actualizó los registros de la tabla de hechos?
- ¿Cómo puedo encontrar los registros cargados hoy?
- ¿Cuántas filas fueron cargadas hoy? ¿Cuántas fueron descartadas por errores?
- ¿Qué procesos terminaron exitosamente?
- ¿Cuántos registros tenía la tabla de hechos antes de la carga? ¿Cuántos después?

El sistema de auditoría, control y creación de informes garantiza la coherencia de los datos, realiza el análisis de impacto y supervisa constantemente la calidad de la información.

### ***El impacto de la mala calidad de datos***

- Impide la Inteligencia de Negocios: Se toman las decisiones incorrectas o se toman decisiones tarde.
- No posibilita que se entienda el negocio.
- Crea informes erróneos con defectos en el análisis.
- Incrementa los costos de gestión.
- Crea diferencias entre aplicaciones que puede requerir trabajos de conciliación de registros.
- Daña la relación con el cliente.
- Imposibilita ofrecer un buen servicio y un trato personalizado.
- No pueden identificarse duplicados de datos.

Todo esto se traduce en reducción de ingreso, incremento de gastos, oportunidades de negocio pérdidas, insatisfacción de los clientes y riesgos de incumplimiento de normas.

### ***Beneficios de la calidad de datos***

- Corrige errores de manera proactiva.
- Habilita oportunidades de negocio.
- Mejora la exactitud.
- Mejora la toma de datos.
- Mejora la confianza.
- Enfoca los esfuerzos de calidad.
- Reduce costos.
- Proporciona una visión única de los elementos de las bases de datos y da visibilidad de las posibles agrupaciones que existen.
- Mejora el mantenimiento de clientes y del servicio ofrecido, además da confianza a los usuarios de la información.
- Reduce las ineficiencias operacionales.
- Aumenta el conocimiento de los datos y permite conocer dónde se están produciendo los errores.
- Maximiza el éxito de las iniciativas y proyectos de BI y por consiguiente se mejora la toma de decisiones.

Explicado todo lo anterior se nombraran las consideraciones y/o estrategias para manejar los problemas de calidad de datos:

1. Identificar la información necesaria para el negocio: En las empresas se manejan grandes cantidades de información y datos, por esto es necesario para mejorar la calidad de los mismos identificar cuál es la información que tiene un mayor impacto en las operaciones del negocio.
2. Establecer criterios de calidad de datos: Cuando se tiene la información necesaria del negocio se definen si los datos que se tienen son los precisos para la empresa.

3. Realizar un chequeo inicial para detectar posibles problemas de calidad de datos: En este paso se lleva a cabo un breve diagnóstico, el cual permite medir la calidad actual de la información, para identificar donde se encuentran los mayores problemas en calidad e identificar cuales son las áreas a intervenir.
4. Automatizar indicadores para la calidad de datos: Con esto se pretende evaluar de forma periódica la calidad de datos. Estos indicadores deben ser claros para que los responsables de la calidad de datos puedan monitorearlos y optimizar así la calidad de los datos.
5. Definir la persona responsables de la calidad de datos: Esta persona debe ser apta para el monitoreo de los indicadores de calidad de datos y efectuar planes de acción para mejorar los mismos.
6. Diagnósticos de calidad de datos: En este paso se identifican las causas por las cuales puede existir mala calidad de datos, todo esto con el fin de planificar estrategias de acción para su mejora. Cuando ocurre esto se crean controles preventivos y correctivos para la mejora del indicador con un responsable que ejecutara la estrategia.
7. Monitoreo y revisión de los indicadores de calidad de datos: Los indicadores de calidad de datos deben ser revisados periódicamente para tener una mejora continua.
8. Finalmente, una vez establecidos los indicadores de calidad de datos, se regresa al punto uno para identificar información que tenga impacto en el negocio y que sea necesario su medición y mejora.

## **CAPITULO 5**

### **CONCLUSIONES DE LA INVESTIGACIÓN Y RECOMENDACIONES**

Después de finalizada la investigación y haber analizado el proceso de ETL, en el proyecto de Inteligencia de Negocios dentro de la empresa de transporte se puede llegar a las siguientes conclusiones:

- Para esta empresa de transporte es imprescindible tener organizada en bases de datos su información estratégica, para que de esta forma los directivos puedan optimizar la toma de decisiones. Es decir, dentro de las empresas que desean ser parte de las tecnologías, la creación de un área de Inteligencia de Negocios facilitará y aportará en gran medida al desarrollo de proyectos estratégicos, tácticos y operativos.
- No emplear el proceso de ETL de manera adecuada traería riesgos como la recolección, transformación y carga de información y datos redundantes

y no actualizados (unidad de la información y que me garantice que sea integra, buena valida y consistente). De esta forma, la experiencia que han tenido las organizaciones con respecto a la implementación adecuada del proceso de ETL ha podido mejorar la actualización de la información en las bases de datos, beneficiando así el negocio a nivel de eficiencia, eficacia y efectividad.

- Cuando se ejecute el proceso de carga, se recomienda que se haga en un horario que no entorpezca otros procesos que se estén ejecutando en el mismo horario y la misma maquina.
- Parece ser que la mayoría de problemas, inconvenientes y riesgos en la implementación de ETL se deben a la calidad poco óptima de los datos. Desde aquí la evaluación de los riesgos permitirá en un plazo de tiempo razonable que los procesos se mejoren continuamente y se planteen estratégicamente así soluciones adecuadas.
- Es necesario para el procesado de datos no encriptar los objetos, como los procedimientos quedan almacenados dentro del DataMart de Operaciones, si se llega a encriptar quitan la posibilidad de que otro proveedor no pueda modificarlas o hacer corrección para su mejora.
- Se recomienda que haya una buena documentación técnica de los procesos de ETL para que cuando un proveedor o persona quiera entender el funcionamiento y quiera hacer modificación sobre estos procesos, tenga de donde consultar. Además es importante para disminuir riesgos en la implementación de ETL contar con un manual de administración del sistema en cuanto a datos.
- Es necesario capacitar a los usuarios de los procesos de ETL en la práctica inteligente de los mismos dentro del proyecto de Inteligencia de Negocios.
- El resultado final del proceso de ETL es poder contar con consultas dinámicas para responder las preguntas del negocio.
- El éxito de la investigación radicó en entender que la forma más adecuada de minimizar los riesgos en el proceso ETL de la empresa de transporte es la posibilidad de mejorar las prácticas de ejecución en los procesos.

- El área de Inteligencia de Negocios de la empresa se debe dar a conocer con las otras áreas para sacar adelante el negocio y optimizar el proceso de toma de decisiones.
- Se recomienda contar con un código administrado, el cual es un código con un buen manejo de excepciones y errores de programación que se ve muy completo a nivel de control de errores en el proceso de ETL, y permite guardar información de auditoría para mirar cada proceso de carga que se ha realizado, o si falla permite ver en donde fallo exactamente.

## **CAPITULO 6**

### **ASPECTOS ADMINISTRATIVOS**

#### **6.1. Recursos**

Para la realización de la investigación se contó con la colaboración de:

Asesor metodológico

Luis Felipe Rosso Ricaurte

Investigadora

Lyda Diana Henao Dorado

## 6.2. Presupuesto

Concepto	Cantidad	Unidad	Valor Unitario	Subtotal
Asesor	40	Hora	\$ 35.000	\$ 1.400.000
Investigador	160	Hora	\$ 20.000	\$ 3.200.000
Equipo de Computo			\$ 1.170.000	\$ 1.170.000
Resma de Papel	2	Unidad	\$ 11.500	\$ 23.000
Impresiones	150	Hoja	\$ 200	\$ 30.000
Otros	-	-	\$ 200.000	\$ 200.000
Total General				\$ 6.023.000

## CAPITULO 7

### BIBLIOGRAFÍA

- KIMBALL, Ralph. The Data Warehouse toolkit. United States, John Wiley & Sons, 1996.
- LOSTHIN, David. Business Intelligence: The Savvy Manager's Guide. United States, Addison Wesley, 2003.
- POE, Vidette with contributions by Laura L. Reeves. Building a Data Warehouse for Decision Support. United States, Prentice Hall PTR, 1996.
- <http://etl-tools.info/es/>
- <http://www.datanalytics.com/etl.html>
- <http://www.tdwi.org>
- <http://www.dmreview.com>
- <http://www.scribd.com/doc/2182415/NORMAS-ICONTEC>

- <http://es.wikipedia.org/wiki/ETL>

## CAPITULO 8

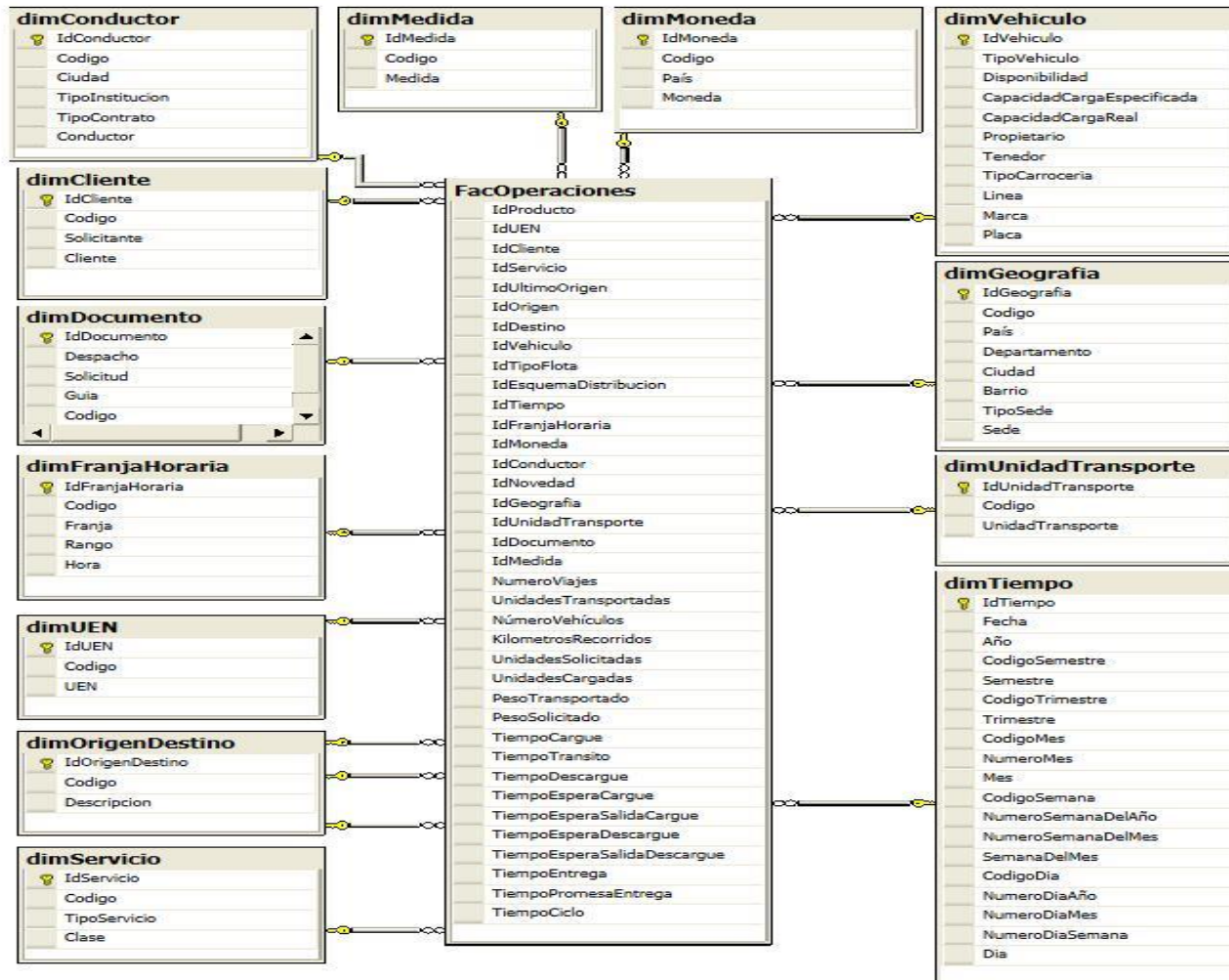
### ANEXOS

#### 8.1. Cronograma de actividades

ACTIVIDAD	PARTICIPANTES	FECHAS DE ACTIVIDADES AÑO 2007/2008
Conocimiento de la Empresa	Lyda Diana Henao Dorado Luis Felipe Rosso	NOV 2007 / SEMANA TRES
Definición del tema del proyecto de grado	Lyda Diana Henao Dorado Jaime A. Díaz	DIC 2007/ SEMANA DOS
Construcción del Anteproyecto	Lyda Diana Henao Dorado Jaime A. Díaz	DIC 2007/SEMANA TRES
Estudio teórico de los conceptos ETL para identificar los principales componentes y conocer así cuales son los riesgos más comunes.	Lyda Diana Henao Dorado	ENE 2008/ SEMANA CUATRO, FEB 2008 / SEMANA UNO, DOS.
Visitar empresas que hayan implementado soluciones BI	Lyda Diana Henao Dorado	FEB 2008 / SEMANA TRES CUATRO.
Examinar las cinco principales etapas para el desarrollo exitoso de los procesos de ETL y la fase de Inteligencia de Negocios.	Lyda Diana Henao Dorado	MAR 2008 / SEMANA UNO
Identificar diez riesgos encontrados dentro de los procesos de ETL examinados en la fase de Inteligencia de Negocios con el fin de proponer cinco posibles recomendaciones que lo mitiguen	Lyda Diana Henao Dorado	MAR 2008 / SEMANA DOS
Definir los pasos necesarios e indispensables para analizar y diseñar los procesos de ETL y las soluciones en Inteligencia de Negocios.	Lyda Diana Henao Dorado	MAR 2008 / SEMANA TRES
Diseñar el mapeo inicial de los datos y fuentes destino en los procesos de ETL.	Lyda Diana Henao Dorado	MAR 2008 / SEMANA CUATRO
Generar y establecer diez estrategias necesarias para manejar los problemas de calidad de datos.	Lyda Diana Henao Dorado	ABR 2008 / SEMANA UNO, DOS
Estudio sobre herramientas SQL Server 2005	Lyda Diana Henao Dorado	FEB 2008 / SEMANA DOS Y TRES

## Anexo 1

### Modelo estrella de Operaciones



## Anexo 2

### FORMATO DE ENTREVISTA

**Objetivo General:** Visitar empresas que hayan implementado soluciones BI y conocer como ha sido el proceso de ETL.

**Metodología:** Se llevara a cabo una serie preguntas para investigar como es el manejo del área BI y todo sobre el proceso de ETL.

¿Cuál es la solución de BI en la empresa?

1. Cuando hablamos de BI nos estamos refiriendo al conjunto de metodologías y tecnologías, que se orientan a mejorar la gestión inteligente de la empresa, que permiten a los equipos directivos controlar los negocios. ¿Qué metodologías y que tecnologías emplean?
2. ¿Qué herramientas emplean para el proceso ETL?
3. ¿Cuáles son las fuentes que ustedes integran?
4. ¿Cómo son los procesos de ETL?
5. ¿Cuáles son los riesgos más altos que ven ETL?
6. Al emplear el proceso de ETL, ¿Que inconvenientes tuvieron durante y después de llevar a cabo este proceso? ¿Y qué etapas emplearon para superarlo?
7. ¿Qué problemas han tenido en los procesos ETL y como lo mitigaron?
8. ¿Que volvería usted hacer igual o no en los procesos de ETL?
9. Existe alguna otra razón fuerte que yo debo tener en cuenta para los procesos de ETL.
10. Nombrar cuatro razones por las que las empresas deben utilizar la inteligencia de Negocios.

## Anexo 3

### Dimensión Conductor

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	SP_DRIVERS	SP_DRIVERS.DRIFIRSTNAME + " " + SP_DRIVERS.DRIFIRSTLASTNAME + " " + SP_DRIVERS.DRISECONDLASTNAME	Nombre Completo del Conductor	dimConductor	Conductor
SQL	Generada	Generada	IdConductor	Clave Primaria autonumérica de la dimensión	dimConductor	IdConductor
SQL	Sipres	SP_DRIVERS	DRIIDENTIFICATION	Código de la dimensión basado en el documento de identidad del Conductor	dimConductor	Código
SQL	Sipres	SP_CITIES	CITNAME	Ciudad de residencia del conductor	dimConductor	Ciudad

## Dimensión Geografía

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	Geografía	País	Nombre del País en donde se ejecuta la operación	dimGeografía	dimGeografía
Excel	Excel	Geografía	Departamento	Nombre del departamento en donde se genera la operación	dimGeografía	dimGeografía
SQL	Sipres	SP_CITIES	CITNAME	Nombre de la ciudad en la cuál se realiza la operación	dimGeografía	dimGeografía
SQL	Sipres	SP_NEIGHBORHOODS	NEINAME	Nombre del barrio	dimGeografía	dimGeografía
Excel	Excel	Tipo Sedes	TipoSede	Tipo de sede a la que se le está transportando la mercancía	dimGeografía	dimGeografía
SQL	Sipres	SP_CITIES	CITAREACODE-CITCODE	Código de la dimensión	dimGeografía	dimGeografía
SQL	Sipres	SP_BRANCHOFFICES.	BRONAME	Nombre de la Sucursal dueña de la operación	dimGeografía	dimGeografía
SQL	Generada	Generada	IdGeografia	Clave Primaria auto numérica de la dimensión	dimGeografía	dimGeografía

### Dimensión Moneda

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
Excel	Excel	Moneda	País	Nombre del País	dimMoneda	País
Excel	Excel	Moneda	Moneda	Moneda del País	dimMoneda	Moneda
SQL	Generada	Generada	IdMoneda	Clave Primaria auto numérica de la dimensión	dimMoneda	IdMoneda

### Dimensión UEN

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
Excel	Excel	UEN	UEN	Nombre de la unidad estratégica del negocio	dimUEN	UEN
SQL	Generada	Generada	IdUEN	Clave Primaria auto numérica de la dimensión	dimUEN	IdUEN

### Dimensión Cliente

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	SP_CLIENTS	CLINIT	Código de la dimensión basado en el Nit del Cliente	dimCliente	Código
SQL	Sipres	SP_CLIENTS	CLINAME	Nombre del Cliente	dimCliente	Cliente
SQL	Sipres	SP_REQUEST S	REQAPPLICANT	Solicitante del servicio o formato	dimCliente	Formato
SQL	Generada	Generada	IdCliente	Clave Primaria auto numérica de la dimensión	dimCliente	IdCliente

### Dimensión Servicio

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	SP_SERVICETYPES	UUID	Código del tipo de servicio	dimServicio	Código
SQL	Sipres	SP_SERVICETYPES	STYNAME	Modalidad del Servicio (Urbano, Exportación,...)	dimServicio	Modalidad
SQL	Sipres	Fórmula	si (Cantidad de guías por despacho)>1, entonces exclusivo, sino combinado	Tipo de Servicio, si es exclusivo o combinado, Exclusivo una sola guía por viaje, combinado contempla varias guías en un solo viaje	dimServicio	TipoServicio
SQL	Generada	Generada	IdServicio	Clave Primaria auto numérica de la dimensión	dimServicio	IdServicio

### Dimensión Vehículo

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	SP_VEHICLETY PES	VETNAME	Tipo de Vehículo	dimVehiculo	TipoVehículo
				Indica si el vehículo es fijo o de apoyo para la operación	dimVehiculo	Disponibilidad
SQL	Sipres	SP_VEHICLES	VEHPLATE	Placa del Vehículo	dimVehiculo	Placa
SQL	Sipres	SP_VEHICLES	VEHSPECCAPACITY	Capacidad de carga según las especificaciones del vehículo	dimVehiculo	CapacidadCarga Especificada
SQL	Sipres	SP_VEHICLES	VEHREALCAPACITY	Capacidad real de carga	dimVehiculo	CapacidadCarga Real
SQL	Sipres	SP_OWNERS	OWNFIRSTNAME+' '+ OWNFIRSTLASTNAME +' '+OWNSECONDLASTN AME	Nombre del propietario del vehículo	dimVehiculo	Propietario
				Nombre del Tenedor del vehículo hace referencia básicamente al Conductor del vehículo	dimVehiculo	Tenedor
SQL	Sipres	SP_VEHICLEBO DYWORKTYPE S	VBWDESCRIPTION	Tipo de Carrocería	dimVehiculo	TipoCarrocería
SQL	Sipres	SP_VEHICLELI NES	VELNAME	Nombre de la clasificación dada a la línea de vehículos	dimVehículo	Linea
SQL	Sipres	SP_VEHICLEBR ANDS	VBRDESCRIPTION	Marca del vehículo	dimVehículo	Marca

SQL	Generada	Generada	IdVehiculo	Clave Primaria auto numérica de la dimensión	dimVehículo	IdVehiculo
SQL	Sipres	SP_VEHICLETY PES	VETNAME	Tipo de Vehículo	dimVehiculo	TipoVehículo

### Dimensión Origen Destino

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	SP_STREETA DDRESSES	STAALIAS	Origen	dimOrigenDestino	Origen
SQL	Sipres	SP_STREETA DDRESSES	STAALIAS	Destino	dimOrigenDestino	Destino
SQL	Generada	Generada	IdOrigenDestino	Clave Primaria auto numérica de la dimensión	dimOrigenDestino	IdOrigenDestino

### Dimensión Esquema Distribución

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
				Identificador del Esquema de distribución	dimEsquemaDistribucion	Código
SQL	Generada	Generada	IdEsquemaOperacion	Clave Primaria auto numérica de la dimensión	dimEsquemaDistribucion	IdEsquemaDistribucion

## Dimensión Tiempo

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Generada	Generada	Año		dimTiempo	Año
SQL	Generada	Generada			dimTiempo	CodigoSemestre
SQL	Generada	Generada	Semestre		dimTiempo	Semestre
SQL	Generada	Generada			dimTiempo	CodigoTrimestre
SQL	Generada	Generada	Trimestre		dimTiempo	Trimestre
SQL	Generada	Generada			dimTiempo	CodigoMes
SQL	Generada	Generada	Mes		dimTiempo	Mes
SQL	Generada	Generada			dimTiempo	CodigoSemana
SQL	Generada	Generada			dimTiempo	NumeroSemanaDelAño
SQL	Generada	Generada			dimTiempo	NumeroSemanaDelMes
SQL	Generada	Generada			dimTiempo	SemanaDelMes
SQL	Generada	Generada			dimTiempo	CodigoDia
SQL	Generada	Generada			dimTiempo	NumeroDiaAño

SQL	Generada	Generada	Dia		dimTiempo	NumeroDiaMes
SQL	Generada	Generada			dimTiempo	NumeroDiaSemana
SQL	Generada	Generada			dimTiempo	Dia
SQL	Generada	Generada	IdVehiculo	Clave Primaria auto numérica de la dimensión	dimVehículo	IdVehiculo
SQL	Sipres	SP_VEHCLETY PES	VETNAME	Tipo de Vehículo	dimVehiculo	TipoVehículo

### Dimensión Franja Horaria

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	SP_TIMESLOT	TISNAME	Nombre de la franja horaria	dimFranjaHoraria	Franja
SQL	Sipres	SP_TIMESLOT	TISTIMEBEGIN - TISTIMEEND	Hora de inicio y fin de la franja horaria	dimFranjaHoraria	Rango
SQL	Sipres	Fórmula con datos provenientes de la tabla SP_TIMESLOT	Horas que se encuentran entre la hora de inicio de la franja (TISTIMEBEGIN) y la hora fin de la franja (TISTIMEEND)	Hora en que se ejecuta la operación	dimFranjaHoraria	Hora
SQL	Sipres	SP_TIMESLOT	UUID	Código único de la franja horaria	dimFranjaHoraria	Código
SQL	Generada	Generada	IdFranjaHoraria	Clave Primaria auto numérica de la dimensión	dimFranjaHoraria	IdFranjaHoraria
SQL	Generada	Generada	Hora	Hora de ejecución de la operación	dimHora	Hora

SQL	Generada	Generada	Minuto	Minuto de ejecución de la operación	dimHora	Minuto
SQL	Generada	Generada	IdHora	Clave Primaria auto numérica de la dimensión	dimHora	IdHora

### Dimensión Unidad Transporte

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres				dimUnidadTransporte	Unidad Transporte
SQL	Sipres				dimUnidadTransporte	Código
SQL	Generada	Generada	IdUnidadTransporte	Clave Primaria auto numérica de la dimensión	dimUnidadTransporte	IdUnidadTransporte

### Dimensión Documento

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	SP_DISPATCHS	DISNUMBER	Número consecutivo que se le asigna al despacho en el momento de su creación.	dimDocumento	Despacho
SQL	Sipres	SP_GUIDES	GUINUMBER	Numero asignado de la guía de Transporte	dimDocumento	Guía
SQL	Sipres	SP_REQUESTS	REQNUMBER	Número de la solicitud	dimDocumento	Solicitud

SQL	Sipres	SP_REQUESTS, SP_GUIDES	REQNUMBER- GUINUMBER	Código de la dimensión	o dimDocument o	Código
SQL	Generada	Generada	IdDocumento	Clave Primaria auto numérica de la dimensión	o dimDocument o	IdDocumento

### Dimensión Medida

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	SP_UNITSOFWEIG HT	UOWNAME	Nombre asignado a la unidad de peso.	dimMedida	Medida
SQL	Sipres	SP_UNITSOFWEIG HT	UUID	Identificador de la medida	dimMedida	Código
SQL	Generada	Generada	IdMedida	Clave Primaria auto numérica de la dimensión	dimMedida	IdMedida

## Dimensión Producto

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	Atributo Variable		Se toma del campo de atributos variables SP_ATTRIBUTEVALUES.ATVVALU E donde el campo SP_ATTRIBUTEDEFINITION.ATDD ESCRPTION sea igual a 'Categoría Producto'	dimProducto	Categoría
SQL	Sipres	Atributo Variable		Se toma del campo de atributos variables SP_ATTRIBUTEVALUES.ATVVALU E donde el campo SP_ATTRIBUTEDEFINITION.ATDD ESCRPTION sea igual a 'SubCategoría Producto'	dimProducto	SubCategoría
SQL	Sipres	Atributo Variable		Se toma del campo de atributos variables SP_ATTRIBUTEVALUES.ATVVALU E donde el campo SP_ATTRIBUTEDEFINITION.ATDD ESCRPTION sea igual a 'Producto'	dimProducto	Producto
SQL	Sipres	Atributo Variable		Se toma del campo de atributos variables SP_ATTRIBUTEVALUES.ATVVALU E donde el campo SP_ATTRIBUTEDEFINITION.ATDD ESCRPTION sea igual a 'Código Producto'	dimProducto	Código
SQL	Generada	Generada	IdProducto	Clave Primaria	dimProducto	IdProducto

## Dimensión Hechos

Tipo de Fuente	Fuente de datos	Tabla u Hoja de Origen	Campo Origen	Descripción	Tabla Destino	Campo Destino
SQL	Sipres	Contador del número de guías	Cuenta el campo SP_GUIDES.GUINUMBER	Medida regular que cuenta la cantidad de viajes realizados	FactOperaciones	Numero Viajes
SQL	Sipres	Sumatoria Unidades Transportadas	suma el valor de SP_LOADED_UNITS.LOUPACKTYPEUID y si el resultado es un valor nulo con respecto a la guía que se esté pivoteando entonces debe de sumarse el valor de SP_LOADEDUNITS.GUITRANSPORTEUNITS	Medida regular que suma las unidades transportadas	FactOperaciones	Unidades Transportadas
SQL	Sipres	Contador de vehículos	Cuenta el campo SP_VEHICLES.VEHPLATE	Medida regular que cuenta la cantidad de vehículos	FactOperaciones	Número Vehículos
SQL	Sipres	Contador de viajes que se realizan desde un destino hasta un origen			FactOperaciones	Número Recogidas
SQL	Sipres	Contador de viajes que se realizan desde un origen a un destino			FactOperaciones	Número Entregas
SQL	Sipres	Cálculo de duración (Hora fin cargue - Hora inicio cargue), estas horas son atributos variables	El valor "Hora fin cargue" se toma del campo de atributos variables SP_ATTRIBUTEVALUES.ATVVALUE donde el campo SP_ATTRIBUTEDEFINITION.ATDDDESCRIPTION sea igual a 'Hora de terminación del cargue'; El valor "Hora inicio cargue" se toma del campo de atributos variables SP_ATTRIBUTEVALUES.ATVVALUE donde el campo SP_ATTRIBUTEDEFINITION.ATDDDESCRIPTION sea igual a 'Hora Inicio Cargue'.	Es tiempo que transcurre entre el inicio de cargue de mercancía a un vehículo y el fin del proceso de cargue	FactOperaciones	Tiempo Cargue

