

Universidad Central “Marta Abreu” de las Villas
Facultad de Matemática Física y Computación
Departamento de Ciencia de la Computación



*Título: Mercado de datos para casos delictivos con
manejo de actualizaciones en las tablas de hechos*

*Tesis Presentada en Opción al Título Académico de
Máster en Ciencia de la Computación*

Autor:

Ing. Enrique Alberto Pérez Anchia

Tutores:

Dra. Luisa Manuela González González

MSc. Iván Maykel Cárdenas Tandrón

Santa Clara 2012.

Frase...

Los datos, organizados y empleados debidamente, pueden convertirse en información. La información, absorbida, comprendida y aplicada por las personas, puede convertirse en conocimientos. Los conocimientos aplicados frecuentemente en un campo pueden convertirse en sabiduría, y la sabiduría es la base de la acción positiva.

Mike Cooley

RESUMEN

En el ejercicio de implementar las soluciones de inteligencia de negocio y almacenamiento de datos se presentan situaciones en las que se debe lidiar con la posibilidad de que los hechos cambien con frecuencia.

La presente investigación aborda el diseño y la implementación de una solución para el soporte a la toma de decisiones donde los hechos presentan cambios debido al desconocimiento, imprecisión y vaguedad con que se registran en un sistema operacional, lo cual se traduce en que los hechos ocurridos están sujetos a sufrir actualizaciones o presentar nuevas versiones de la verdad.

La solución fue implantada en el Centro de Tratamiento y Análisis de Información de Seguridad Ciudadana (CTAISC) de la República Bolivariana de Venezuela y cumple con los objetivos deseados por este tipo de centros, donde los análisis de la información en línea demandan la comparación, la asociación, el estudio de tendencias y la localización territorial de los hechos ocurridos y sus actualizaciones.

Las buenas prácticas recomendadas en este trabajo establecen un marco de referencia práctico para los diseñadores de bases de datos multidimensionales que se vean envueltos en la implementación de soluciones con similares requisitos.

ABSTRACT

In the exercise of implementing business intelligence and datawarehouse solutions are situations in which the specialist must deal with the possibility that the facts change frequently.

This research is about the design and implementation of a solution to support decision making where the facts have changed due to ignorance, imprecision and vagueness that occur in an operational system, which means that the events are subject to suffer submit updates or new versions of the truth.

The solution was implemented at the Center of Treatment and Analysis of Public Safety Information (CTAISC) of the Bolivarian Republic of Venezuela and meets the desired objectives of these centers, where online information analysis demand comparison, association, the study of trends and the geographical location of the events and their submit updates.

The best practices establish in this paper give a practical framework for designers of multidimensional databases that would be involved in implementing solutions with similar requirements.

DEDICATORIA

A la memoria de mi abuelo José Luis. Que en paz descanse.

AGRADECIMIENTOS

A mi familia y compañeros de trabajo.

TABLA DE CONTENIDOS

INTRODUCCIÓN	9
CAPÍTULO I “Caso de estudio: Seguimiento de hechos delictivos”	13
1.1 Contexto del caso de estudio	13
1.2 Captación de los hechos delictivos.....	14
1.3 Modelación de los hechos delictivos.....	15
1.4 Hechos delictivos	16
1.5 Datos cualitativos	16
1.5.1 Delito.....	16
1.5.2 Fecha	17
1.5.3 Geografía.....	17
1.5.4 Lugar	18
1.5.5 Perfil de persona.....	18
1.5.6 Órgano de Seguridad.....	19
1.5.7 Expediente.....	19
1.6 Datos cuantitativos	20
1.7 Requisitos de negocio.....	20
1.8 Conclusiones parciales	23
CAPÍTULO II “Conceptos, herramientas y buenas prácticas para el almacenamiento del cambio lento de los hechos”	24
2.1 Toma de decisiones	24
2.2 OLAP	26
2.3 Sistemas de apoyo a la toma de decisiones	27
2.4 Almacenes de datos.....	28
2.5 Integración de datos	34
2.6 Análisis multidimensional de datos.....	35
2.7 Buenas prácticas para el mantenimiento de actualizaciones de los hechos.....	36
2.7.1 Cambio lento tipo 2 en la tabla de hechos.....	36
2.7.2 Tabla auditoria para el control de cambios	37
2.7.3 Tablas snapshots.....	39
2.8 Herramientas de desarrollo.....	41
2.8.1 Pentaho Data Integration.....	41
2.8.2 Oracle Database 10g R2 Enterprise Edition.....	41
2.8.3 Oracle Warehouse Builder 10g	42
2.8.4 PL / SQL Developer.....	42

2.8.5	Oracle Discoverer.....	42
2.9	Conclusiones Parciales.....	43
<i>CAPÍTULO III “Mercado de datos para el seguimiento de hechos delictivos”</i>		44
3.1	Arquitectura del Sistema de Información Nacional de Seguridad Ciudadana	44
3.2	Selección del proceso de negocio.....	45
3.3	Granularidad.....	46
3.4	Dimensiones.....	47
3.5	Tabla de hechos.....	50
3.5.1	Propuesta de solución para manejo de actualizaciones en la tabla de hechos	52
3.5.2	Técnica para integrar los datos.....	59
3.5.3	Pauta para el análisis sobre los datos.....	61
3.6	Conclusiones Parciales.....	62
<i>CONCLUSIONES</i>		64
<i>RECOMENDACIONES</i>		65
<i>REFERENCIAS BIBLIOGRÁFICAS</i>		66

ÍNDICE DE FIGURAS

FIGURA 1. 1 ESTRUCTURA DE RECOPIACIÓN PARA HECHOS DELICTIVOS.....	16
FIGURA 1. 2 HECHOS DELICTIVOS.....	16
FIGURA 2. 1 PROCESO EMPÍRICO PARA TOMAR DECISIONES.....	24
FIGURA 2. 2 ORGANIZACIÓN JERÁRQUICA Y DEPARTAMENTAL DE UNA EMPRESA.....	26
FIGURA 2. 3 VISTA GENERAL DE UNA TABLA DE HECHOS CON DIMENSIONES Y MEDIDAS.....	33
FIGURA 2. 4 PRINCIPALES PASOS EN EL PROCESO DE ETL.....	34
FIGURA 2. 5 EJEMPLO DE RALPH KIMBALL.....	37
FIGURA 2. 6 TABLA DE HECHOS CON TABLA AUDITORIA PARA CONTROL DE CAMBIO.....	38
FIGURA 2. 7 TABLA SNAPSHOT PARA EL ALMACENAMIENTO DE ACTUALIZACIONES DE LOS HECHOS.....	40
FIGURA 3. 1 ARQUITECTURA DE LA SOLUCIÓN DE SOFTWARE.....	45
FIGURA 3. 2 TABLA DE HECHOS DELITOS DE VIOLACIÓN A LA LEY.....	46
FIGURA 3. 3 DIMENSIÓN LEY.....	47
FIGURA 3. 4 DIMENSIÓN FECHA.....	47
FIGURA 3. 5 DIMENSIÓN GEOGRAFÍA.....	48
FIGURA 3. 6 DIMENSIÓN LUGAR.....	48
FIGURA 3. 7 DIMENSIÓN PERFIL DE PERSONA.....	49
FIGURA 3. 8 DIMENSIÓN ÓRGANO DE SEGURIDAD.....	49
FIGURA 3. 9 DIMENSIÓN EXPEDIENTE.....	50
FIGURA 3. 10 DIMENSIÓN AUDITORÍA.....	50
FIGURA 3. 11 TABLA DE HECHOS PARA ALMACENAR VIOLACIONES A LA LEY (DELITOS).....	51
FIGURA 3. 12 TABLA DE HECHOS CON IDENTIFICADOR EN DIMENSIÓN DEGENERADA.....	53
FIGURA 3. 13 TABLA DE HECHOS CON UN REGISTRO.....	53
FIGURA 3. 14 TABLA DE HECHOS CON UN REGISTRO, MÁS UNA ACTUALIZACIÓN DEL MISMO.....	54
FIGURA 3. 15 TABLA DE HECHOS CON IDENTIFICADOR EN DIMENSIÓN CONFORMADA.....	55
FIGURA 3. 16 TABLA DE HECHOS CON UN REGISTRO.....	55
FIGURA 3. 17 TABLA DE HECHOS CON UN REGISTRO, MÁS UNA ACTUALIZACIÓN DEL MISMO.....	56
FIGURA 3. 18 TABLA DE HECHOS CON IDENTIFICADOR CONFORMADO POR LA COMBINACIÓN DE LAS LLAVES FORÁNEAS DE LAS TABLAS DE DIMENSIÓN ASOCIADAS A LA TABLA DE HECHOS.....	57
FIGURA 3. 19 TABLA DE HECHOS CON UN REGISTRO.....	58
FIGURA 3. 20 TABLA DE HECHOS CON UN REGISTRO, MÁS UNA ACTUALIZACIÓN DEL MISMO.....	59
FIGURA 3. 21 FLUJO DE ETL PARA EL ALMACENAMIENTO DE LAS ACTUALIZACIONES DE LOS HECHOS.....	60

INTRODUCCIÓN

Tomar decisiones a nivel gubernamental está lejos de ser una tarea trivial, el camino para alcanzar los objetivos esperados está lleno de supuestos atajos que pueden llevar al fracaso, una mala decisión puede implicar resultados irreversibles.

Algunos procesos son tan simples y cotidianos que se solucionan muy rápidamente, pero existen otros casos en los cuales las consecuencias de una mala o buena elección pueden tener repercusiones importantes; en particular en un contexto de seguridad ciudadana puede incluso influir directamente en el desarrollo socio económico de un país.

Los sistemas para el soporte a la toma de decisiones juegan un papel determinante en el día de hoy y más aun cuando se trata de tomar decisiones que garanticen la seguridad ciudadana de un país.

El **Centro de Tratamiento y Análisis de Seguridad Ciudadana (CTAISC)** es el ente rector por excelencia, de la República Bolivariana de Venezuela, encargado del análisis situacional cuyo objetivo principal es brindar información oportuna, confiable y primaria al despacho del Vice Ministerio de Seguridad Ciudadana del gobierno venezolano.

En la actualidad el CTAISC cuenta con el **Sistema de Información Nacional de Seguridad Ciudadana (SINSEC)** el cual centraliza la información entregada por el **Cuerpo de Investigaciones Científicas Penales y Criminalísticas (CICPC)**, entre otras.

Recopilar, almacenar y presentar los datos en forma de reportes estadísticos a los niveles superiores requiere de sumo cuidado en la gestión de los datos. Con frecuencia los datos que se reciben desde CICPC no tienen la precisión que se espera producto de la rapidez con que se registran y se envían al CTAISC. En este ámbito es muy común encontrarse con el problema de que los hechos delictivos, por naturaleza y dinámica de trabajo de los entes de seguridad de gobierno, pasan por un proceso de investigación que arroja un resultado no siempre conclusivo, por ejemplo: ***caso cerrado bajo el delito de Homicidio***. Dicho resultado puede verse afectado por la aparición de nuevos

elementos post-investigación que conlleven a una nueva conclusión, por ejemplo: *Suicidio*. En este momento el hecho registrado como *Homicidio* deja de ser válido.

Los cambios se pueden manifestar igualmente en los demás datos del caso como pueden ser la fecha y la geografía. Incluso la cantidad de víctimas o victimarios involucrados en el delito.

Cuando se habla de un hecho aislado no tiene gran impacto, pero en el momento en que el flujo de datos gira en torno a los cientos de registros diarios ya el tema se vuelve delicado. Si se reportan 100 homicidios todos los días durante una semana y al concluir la semana se envía un acumulado semanal, este debería ser de 700 homicidios. Sin embargo, si durante el transcurso de la semana varios casos cambian, los totales no coincidirían con la suma de los diarios.

En términos de sistemas bases de datos operacionales, para gestionar los hechos delictivos, basta con una simple actualización del caso, pero en los sistemas para el soporte a la toma de decisiones las modificaciones deben hacerse con cuidado para no perder la historia.

Hoy el CTAISC está limitado en su capacidad de darle seguimiento a los cambios que puede sufrir un determinado hecho delictivo, lo cual crea descontento en el despacho del vice ministerio y desconfianza en los reporte.

El Centro de Tratamiento y Análisis de Información de Seguridad Ciudadana recibe todos los días varias actualizaciones de los hechos que ya tiene almacenados, realmente el reporte de “ayer” entregado al despacho del vice ministerio fue correcto, pero solo hasta ayer.

Si el CTAISC no tiene en cuenta estas actualizaciones, los números totales al final de la semana o del mes no estarían acorde con lo real. Estaría elaborando reportes con un considerable margen de error, ya que reportaría la suma de las cantidades que se recibieron diariamente cuando en realidad puede ser diferente dadas las actualizaciones.

La información que maneja CTAISC es de interés tanto para el gobierno Venezolano como para los Órganos de Seguridad del Estado. Esta información es clave para orientar la operatividad de los recursos policiales en aras de atacar directamente los hechos delictivos.

En el CTAISC se recibe la información diaria de los hechos delictivos ocurridos en el día anterior, junto con esto llegan nuevas evidencias sobre hechos previamente registrados. Dado que la misión es reportar al gobierno las estadísticas de los hechos delictivos y realizar análisis de tendencia está obligado a mantener un historial de actualizaciones de los hechos que experimenten cambios.

De la problemática anterior se deduce el siguiente **problema científico**: ¿Cómo concebir una solución que brinde soporte a la toma de decisiones de seguridad ciudadana, en un entorno donde los hechos delictivos están sujetos a sufrir actualizaciones?

El **objeto de estudio** del problema se enmarca en las mejores prácticas y soluciones para el diseño de una solución de *BI/DW*¹ y el **campo de acción** en las mejores prácticas para implementar una solución de *BI/DW* que permita el almacenamiento del historial de actualizaciones de los hechos.

Para dar solución al problema científico se plantea como **objetivo general**: diseñar e implementar un mercado de datos para la gestión de hechos delictivos.

Como **objetivos específicos** se plantea:

- ✓ Describir el caso de estudio en el cual se enmarca el problema.
- ✓ Argumentar buenas prácticas para manejar las actualizaciones en la tabla de hechos.
- ✓ Diseñar la arquitectura de la solución al mercado de datos.

Se define como **hipótesis de investigación**: si se concibe una solución que maneje efectivamente las actualizaciones de los hechos delictivos, entonces se podrá presentar reportes confiables.

El valor práctico consiste en el diseño y la implementación de una solución de inteligencia de negocio orientada al análisis de los índices de seguridad ciudadana que gestione efectivamente las actualizaciones de los hechos que se registran.

El valor teórico consiste en la presentación de buenas prácticas relacionadas con actualizaciones en la tabla de hechos y la adecuación de ellas como una práctica a

¹ BI/DW: Por sus siglas en inglés Business Intelligence and Datawarehouse.

implementar para el seguimiento de hechos delictivos, situación donde los hechos presenten actualizaciones y que por requerimiento(s) del negocio se necesite almacenar esta historia.

El documento está estructurado en tres capítulos que se describen a continuación:

En el capítulo I. **“Caso de estudio: Hechos delictivos”** se describen los antecedentes que dieron lugar a la necesidad de implementar una solución para la toma de decisiones donde los hechos pudiesen ser efectivamente actualizados. Se describe el caso de estudio en el cual se enmarca la problemática existente y se dan las razones por las cuales se hace indispensable implementar una solución de inteligencia de negocios con estas características.

En el capítulo II. **“Conceptos, herramientas y buenas prácticas para el almacenamiento de hechos con cambio lento”** se formalizan los principales conceptos de la investigación y se describen y analizan aspectos de marcada importancia en el diseño de este tipo de soluciones. Se describen las herramientas de desarrollo seleccionadas para darle solución al problema en cuestión. Se exponen lineamientos en forma de prácticas de diseño en situaciones donde los hechos presentan diferentes versiones de la verdad.

En el capítulo III. **“Diseño para el soporte al cambio lento de los hechos delictivos”** se expone la arquitectura de la solución y se describe el diseño de la misma. Consecuentemente con el diseño del almacén de datos se proponen los mecanismos para integrar y analizar los hechos que están propensos a sufrir actualizaciones.

CAPÍTULO I “Caso de estudio: Seguimiento de hechos delictivos”.

En el presente capítulo se hace una descripción de los requerimientos de negocio para el diseño de la solución. Se explica el caso de estudio que da lugar a la necesidad de implementar una solución de inteligencia de negocio que de seguimiento de manera eficiente al historial de las evidencias relacionadas con hechos delictivos. Se describe brevemente el contexto en que se enmarca el caso de estudio.

1.1 Contexto del caso de estudio

Como parte del Convenio de Colaboración correspondiente a la VII Convención Mixta Cuba-Venezuela, se firmó el Contrato para la Creación del Centro de Tratamiento y Análisis de Información de Seguridad Ciudadana (CTAISC) y la metodología para la implantación de centros de emergencias 171 en la República Bolivariana de Venezuela.

El CTAISC se construyó en el **Ministerio del Poder Popular para las Relaciones Interiores y Justicia** (MPPRIJ) y se dotó de tecnología con el fin de resolver los siguientes problemas:

- ✓ Necesidad de integrar la información relacionada con la seguridad ciudadana a nivel nacional.
- ✓ Limitaciones para formular políticas y estrategias nacionales con un carácter proactivo, para coadyuvar a mejorar los niveles de seguridad ciudadana.
- ✓ Dificultades para lograr el seguimiento en tiempo real de situaciones de emergencia relacionadas con la seguridad ciudadana.
- ✓ Limitaciones para realizar el seguimiento y control de las políticas y decisiones estratégicas en relación con hechos y situaciones extraordinarias o relevantes.

En tal sentido el MPPRIJ, atendiendo a su misión institucional de garantizar la seguridad ciudadana mediante la formulación de políticas dirigidas al resguardo de la paz pública, el desarrollo territorial equilibrado y la estabilidad de la nación, asigna al Centro de Tratamiento y Análisis de Información de Seguridad Ciudadana la siguiente misión: “Garantizar la integración y el análisis multidisciplinario de la información generada por las diferentes bases de datos de los Órganos de Seguridad Ciudadana del

país, coadyuvar a la concepción de políticas y estrategias a fin de mejorar los niveles de seguridad ciudadana y brindar seguimiento a hechos y situaciones extraordinarias. Para ello cuenta con tecnología de punta y un talento humano caracterizado por su honestidad, profesionalidad, responsabilidad y sentido de cooperación.”

1.2 Captación de los hechos delictivos

En el caso particular del Cuerpo de Investigaciones Científicas Penales y Criminalísticas (CICPC), Órgano perteneciente al MPPRIJ, gestiona por cada caso de delito un expediente identificador del mismo. Este expediente contiene los datos relativos al hecho al que pertenece, dígame la fecha, la localización geográfica, la ley que se viola, así como la delegación policial que atiende el caso.

El CICPC ejecuta las investigaciones que en materia de criminalidad se realizan en el país, y no solo registra diariamente información sobre los delitos contra la vida e integridad personal, contra la propiedad y contra el estado, sino que sus operaciones dan una idea clara del actuar policial en este sentido.

La integración del CICPC \ SIIPOL (Sistema de Investigación e Información Policial) al CTAISC \ SINSEC-DI2 (Sistema de Información Nacional de Seguridad Ciudadana, Segundo Dominio Informativo) es elemental para la formulación de estrategias de cobertura nacional y local, a partir del conocimiento real de los sucesos delictivos y del actuar policial.

Por la dinámica con la que opera el CICPC y la naturaleza con la que ocurren los hechos, sucede que en ocasiones los datos que describen un hecho experimentan cambios. Entiéndase que la imprecisión, incompletitud, la vaguedad y el desconocimiento, son factores que están siempre presentes en los datos de un hecho delictivo.

El CICPC envía diariamente los detalles de los hechos delictivos ocurridos a nivel nacional. Los casos policiales son investigados constantemente, lo que conduce a que casos que en un momento dado se consideran cerrados, puedan de nuevo reabrirse por la aparición de nuevas evidencias o elementos que hacen que este tipo de procesos manejen hechos muy susceptibles a las actualizaciones.

Ante la declaración de un hecho determinado, supóngase un robo, las autoridades que recogen los detalles del hecho tienen la obligación de registrarlo en su sistema operacional tal y como lo describe el denunciante, no obstante el CICPC siempre abre un expediente del caso y lo investiga a profundidad. Esto se traduce en que comúnmente, aunque no siempre, los casos se pueden ver afectados por la aparición de nuevos elementos.

Los hechos que se registran en un día pueden variar al otro o al siguiente, incluso meses o años después. Varios pueden ser los motivos que provoquen esta situación, entre ellos:

- ✓ El denunciante no estaba claro del lugar donde se produjo el robo.
- ✓ El denunciante acudió a las autoridades unos días después del robo y en ese momento no recuerda con exactitud la fecha del mismo.
- ✓ El denunciante no conoce con exactitud la ubicación geográfica en la que se produjo el hecho.
- ✓ Se denuncia un robo cuando en realidad fue un auto robo.

Ante esta situación los especialistas de CICPC, con ayuda del sistema operacional SIIPOL, actualizan el caso con la posible pérdida de información que a los efectos estratégicos es deseable preservar.

1.3 Modelación de los hechos delictivos.

Los datos con los que debe trabajar el CTAISC están divididos en dos grupos, los que aportan un valor cuantitativo y los que proveen información cualitativa. En cualquier caso, los datos no siempre tienen un cien por ciento de fidelidad, esto se traduce en que un determinado hecho puede verse afectado por la aparición de nuevos elementos que provoquen la actualización del mismo, una nueva verdad del hecho puede hacerse presente. En este punto se está lidiando con la posibilidad de que un hecho presente una nueva versión de la verdad en momentos posteriores.

Los datos que se desean recopilar sobre los *hechos delictivos* son los referentes a: *delito, fecha, geografía, lugar, perfiles de personas, órganos de seguridad y expedientes de casos delictivos*. La **figura 1.1** muestra la estructura de almacenamiento para el caso de estudio.



Figura 1. 1 Estructura de recopilación para hechos delictivos.

1.4 Hechos delictivos

Los hechos delictivos son la diana central en el proceso de consulta y análisis de información y elaboración de reportes. La figura 1.2 muestra el punto central de análisis del modelo dimensional que almacena los hechos delictivos.



Figura 1. 2 Hechos Delictivos.

1.5 Datos cualitativos

1.5.1 Delito

Un delito consiste en una violación a alguno de los aspectos recogidos en el código penal de Venezuela. El CTAISC tiene identificado nueve delitos dentro del código penal que están contemplados en un plan de seguridad dado que son los de mayor impacto negativo en el bienestar de la población y el desarrollo social del país. Esta estructura permite organizar los reportes por delitos de interés en cuanto a:

- ✓ Mayor impacto (*violación, homicidios, secuestro*).

- ✓ Más comunes (*robo de vehículos, robo de motos*).
- ✓ Más violentos (*homicidio, hurto, robo a mano armada*).

La organización del tesoro delictivo está dada por una clasificación jerárquica en cuanto a *clase, tipo y delito*. El plan de seguridad anteriormente mencionado consiste en un atributo de la estructura *Delito* que marca aquellas violaciones a la ley que el gobierno venezolano considera atender con mayor prioridad y en base a los cuales se solicitan la mayoría de los análisis.

1.5.2 Fecha

Marca de tiempo de cuando ocurre un hecho delictivo, en cuanto a año, mes, día. Este aspecto que permite la comparación de dos o más periodos de tiempo. Los datos históricos que se recopilan son desde 1990 hasta la actualidad con el objetivo de comparar los años previos al gobierno bolivariano y los años posteriores a este en cuanto a los índices de seguridad ciudadana y percepción del riesgo en la población.

Las fecha de ocurrencia de los hechos es uno de los aspectos básicos en los reportes, siempre se desea saber en qué momento ocurren los hechos delictivos. Esta estructura permite enmarcar el espacio de búsqueda de las consultas en base a la época escogida para analizar, incluso se pueden hacer comparaciones de dos o más épocas similares de años diferentes.

1.5.3 Geografía

Este atributo permite ubicar en el espacio las aéreas o sectores de la población con mayor o menor violencia. Esto es un punto importante a la hora de distribuir los recursos operativos para combatir los delitos. La distribución geográfica de Venezuela se organiza en estados que tienen municipios y estos a su vez tienen parroquias.

Esta estructura aporta conjuntamente con la *Fecha* la base del análisis sobre todos los informes que se generan. La organización jerarquía está dada por:

- ✓ Estados (*Distrito Capital, Miranda, Zulia*).
- ✓ Municipios (*Libertador, Baruta, Maracaibo*).
- ✓ Parroquias (*Antímano, El Cafetal, Bolívar*).

Adicionalmente está diseñada para almacenar los cambios que puedan ocurrir dado una nueva distribución político-geográfica, manteniendo un historial de las antiguas distribuciones.

1.5.4 Lugar

A pesar de poder ubicar geográficamente las zonas delictivas, con esta estructura se puede identificar cuáles son las zonas y los lugares en donde ocurren con mayor o menor incidencia los hechos delictivos. Contiene básicamente dos propiedades las cuales son:

- ✓ Zona (*urbana, rural*).
- ✓ Lugar (*hospital, centro comercial, parque*).

Estas propiedades no presentan relación, desde un punto de vista jerárquico, sin embargo son dos variables que por decisiones de diseño dadas las características del negocio están estrechamente relacionadas. En estas situaciones es conveniente que convivan bajo la misma estructura.

1.5.5 Perfil de persona

El CTAISC estudia el comportamiento de los elementos de investigación como es el caso de las personas. Para esto recopila datos referentes al perfil físico social de las personas entre los cuales están la nacionalidad, el sexo, la edad, el nivel educativo, la ocupación laboral y el estado civil. Esto permite hacer estudios sobre qué tipo de personas intervienen con mayor o menor frecuencia en delitos, ya sea como víctimas o victimarios.

Las personas son el principal elemento de investigación en cuanto a seguridad ciudadana. En esta estructura se agrupan varias propiedades que aportan información acerca del perfil físico y social de los individuos involucrados en hechos delictivos:

- ✓ Nacionalidad (*venezolano, cubano, boliviano, colombiano*).
- ✓ Sexo (*masculino, femenino*).
- ✓ Edad y rangos de edad (*45-50, 60-65*).
- ✓ Nivel escolar (*primaria, secundaria, técnico, universitario*).

- ✓ Ocupación laboral (*medico, informático, chofer, retirado*).
- ✓ Estado civil (*soltero, casado, divorciado, viudo*).

En este caso, como en la estructura *Lugar* anteriormente presentada, las propiedades no tienen relación directa. Solo se agrupan por un concepto que generaliza el sentido conceptual de las propiedades que agrupa.

1.5.6 Órgano de Seguridad

Los hechos delictivos son atendidos por las entidades policiales. Estos se encargan de elaborar los primeros reportes, incluso de actualizar los casos en la medida que pueda ir apareciendo nueva información sobre los mismos. Esta estructura contiene las siguientes propiedades:

- ✓ Tipo (*CICPC*).
- ✓ Entidad superior (*Delegación a la cual se subordina*).
- ✓ Nombre (*nombre de la sub-delegación*).
- ✓ Siglas (*nombre en forma de siglas*).

Los Órganos de Seguridad están desagregados en todo el país por Delegaciones, Sub-delegaciones y unidades policiales de acuerdo con la distribución geográfica de Venezuela.

1.5.7 Expediente

En esta estructura se recogen los datos del expediente que identifica cada caso. Los hechos delictivos son asociados a un identificador del caso en cuestión. Los expedientes experimentan cambios de estado (*cerrado, abierto, reabierto*) en dependencia del avance que se tenga en la investigación del caso.

- ✓ Tipo (*policial, forense, accidentalidad*).
- ✓ Numero (*ABCDE-12345*).
- ✓ Estado (*abierto, cerrado, reabierto*).
- ✓ Relevancia (*alta, media, baja*).

1.6 Datos cuantitativos

Los datos cuantitativos son constituyen el componente aritmético en la elaboración de reportes. Aportan las cantidades de cada aspecto que se desea analizar o medir. En base a estos números es que se puede elaborar los reportes estratégicos.

En cuanto a los hechos delictivos los aspectos medibles son:

- ✓ Cantidad de fallecidos.
- ✓ Cantidad de lesionados.
- ✓ Cantidad de detenidos.

Tanto los datos cuantitativos como los datos cualitativos, que conforman y caracterizan los hechos delictivos, están sujetos a cambiar debido a la propia naturaleza con la que ocurren y describen estos hechos.

1.7 Requisitos de negocio

El CTAISC necesita tomar las decisiones correctas en el momento apropiado soportado por una solución de inteligencia de negocios mediante la cual se recopile almacenen y analicen los datos relevantes a partir de los datos extraídos de los sistemas operacionales de los órganos de seguridad.

1.7.1 Fidelidad y transparencia en los reportes

Poder saber que el CICPC le envió un caso de homicidio en una fecha determinada y que posteriormente le envió una actualización del mismo caso, ahora como suicidio, igualmente con cambios en la geografía, la fecha y el lugar donde ocurrió el hecho, le da un alto grado de credibilidad ante los máximos dirigentes del país ya que en todo momento pueden justificar los cambios en los indicadores. Los reportes elaborados por CTAISC son fuente de información nacional. Es importante que la población no se sienta engañada cuando se informan los resultados en la combatividad a la violencia. La seguridad ciudadana es inversamente proporcional a la percepción del riesgo.

1.7.2 Análisis de desempeño laboral

Igualmente, el seguimiento del historial de actualizaciones de los hechos da la posibilidad de realizar análisis sobre el desempeño laboral de los Órganos de Seguridad del Estado. Poder medir los índices de desempeño laboral de las autoridades policiales

en las Delegaciones y Sub-delegaciones garantiza establecer mecanismos para evitar la imprecisión o vaguedad a la hora de registrar los hechos delictivos. Los reportes del CTAISC no solo van dirigidos al gobierno, también están preparados para que puedan ser interpretados por los Órganos de Seguridad del Estado a modo de retroalimentación.

1.7.3 Reportes de retroalimentación

Los órganos de seguridad del Estado como CICPC necesitan de información acerca de dónde y cómo operar ante ciertas situaciones, sin esta información los Órganos pueden caer en la ineffectividad al no saber cómo emplear objetivamente los recursos humanos y materiales con que cuentan. Venezuela es un país muy grande, por muchos recursos con que cuentan los Órganos de Seguridad, si no se pueden caracterizar bien las áreas de delictividad, se disminuye considerablemente la capacidad operativa.

Otro aspecto importante para la operatividad es saber con anterioridad el comportamiento que presentan las variables que forman a los índices de seguridad ciudadana

1.7.4 Análisis predictivo

Poder predecir y determinar tendencias son aspectos que influyen directamente en la capacidad operativa de los Órganos. De esta manera se pueden atacar con mayor fuerza las áreas de mayor índice delictivo. CTAISC tiene la responsabilidad de aportar información predictiva tanto a los dirigentes del país como a las autoridades policiales. Se hace necesario anteponerse a los hechos, estar preparados a todos los niveles para disminuir los delitos conociendo lo que sucede en ciertos periodos del año o en ciertos sectores geográficos.

1.7.5 Misión de CTAISC

De manera general y acorde con la Gaceta Oficial de la República Bolivariana de Venezuela, el rol social de CTAISC, **recientemente convertido en Observatorio Venezolano de Seguridad (OVS)**, se basa en:

- ✓ Recopilar, discriminar, procesar, analizar y cotejar todos los datos cualitativos y cuantitativos de los indicadores de Criminalidad y Violencia así como los concernientes a planes, programas, y proyectos en materia de seguridad

ciudadana aplicados a nivel nacional, a fin de garantizar la efectividad y eficacia de los mismos.

- ✓ Realizar estudios relacionados a los fenómenos que afectan la Seguridad Ciudadana, igualmente la realización de encuestas de Victimización.
- ✓ Establecer mecanismos a través de la página Web de Observatorio y otros medios para la publicación de los estudios y datos oficiales de Seguridad Ciudadana.
- ✓ Elaborar mapas situacionales y criminalísticos sobre la incidencia delictiva a nivel nacional.
- ✓ Promover y coordinar el desarrollo de investigaciones en materia de seguridad ciudadana.
- ✓ Elaborar informes sobre la problemática delictiva, violencia, sistema judicial, situación penitenciaria, conflictos (económicos, políticos, sociales y culturales) y momentos coyunturales, que permitan conocer los factores que lo originan, así como posibles alternativas para su solución.
- ✓ Elaborar propuestas de escenarios que se pudieran presentar de acuerdo a las situaciones coyunturales suscitadas en el país en materia de seguridad ciudadana.
- ✓ Analizar los datos estadísticos proporcionados por los entes de control social formal del Estado, en materia de seguridad ciudadana.
- ✓ Recomendar programas para minimizar la incidencia delictiva y conflictos que generan inestabilidad en los niveles de seguridad ciudadana del país en base a los estudios realizados en la materia.
- ✓ Diseñar e implementar la plataforma tecnológica y sistemas integrales de información a nivel nacional, estatal y municipal y promover la creación de unidades estadísticas por cada órgano de seguridad ciudadana.
- ✓ Actualizar de forma periódica la información recopilada y velar por la calidad y confiabilidad de la información producida en su seno.

- ✓ Asistir al viceministro o viceministra con competencia en materia de seguridad ciudadana en la planificación, coordinación y dirección de las actividades del Observatorio.
- ✓ Establecer el criterio técnico, así como las directrices a emplear en el diseño, planificación, estructuración y formulación de programas de prevención inherentes a la Seguridad Ciudadana.

El CTAISC tiene la responsabilidad de guiar los pasos estratégicos del gobierno venezolano, analizar los datos recopilados y organizarlos en forma de reportes que soporten las decisiones a tomar, y al mismo tiempo analizar el desempeño laboral de los órganos de seguridad del estado.

En el CTAISC se recibe la información diaria de los hechos ocurridos, junto con esto llegan los hechos modificados. Dado que la misión es reportar al gobierno las estadísticas de los hechos delictivos y realizar análisis de tendencias está obligado a mantener un historial de versiones de los hechos que experimenten cambios.

1.8 Conclusiones parciales

En este capítulo se han expuesto las características, la misión y los objetivos que persigue el CTAISC. Se han expuesto las particularidades y los requerimientos que sustentan la necesidad de diseñar una solución de inteligencia de negocios con la cual el CTAISC pueda desempeñar efectivamente su rol social.

CAPÍTULO II “Conceptos, herramientas y buenas prácticas para el almacenamiento del cambio lento de los hechos”.

Para una correcta comprensión de los temas que se abordan en el presente documento es necesario dominar los conceptos básicos relacionados con la inteligencia de negocios y sus soluciones, todo esto dentro del marco de la compleja actividad de tomar decisiones. A continuación se exponen un grupo de conceptos y terminologías asociadas al campo de la inteligencia de negocios y sus aplicaciones.

2.1 Toma de decisiones

La toma de decisiones como así se conoce, no es más que el proceso que consiste en concretar la elección entre varias alternativas.

Uno de los aspectos más importantes dentro del sector laboral tanto estatal como de la actividad privada, es la toma de decisiones, que es el proceso durante el cual la persona debe escoger entre dos o más alternativas (PALACIOS, 2009).

El proceso para tomar decisiones es empírico, una vez elegida y aplicada una decisión solo resta definir y analizar el nuevo problema resultado de la previa decisión. A esto se enfrentan las organizaciones en su proceso de desarrollo. Una incorrecta elección hace más largo el camino para alcanzar los objetivos, una correcta decisión hace que la empresa de un paso adelante y no tomar decisiones es una mala decisión. En **la figura 2.1**, se muestra el proceso empírico para tomar una decisión.



Figura 2. 1 Proceso empírico para tomar decisiones.

En algunos casos, por ser tan simples y cotidianos, este proceso se realiza de forma implícita y se soluciona muy rápidamente, pero existen otros casos en los cuales las

consecuencias de una mala o buena elección pueden tener repercusiones en la vida y si es en un contexto laboral en el éxito o fracaso de la organización para los cuales es necesario realizar un proceso más estructurado que puede dar más seguridad e información para resolver el problema.

Las decisiones se pueden clasificar teniendo en cuenta diferentes aspectos, como lo es la frecuencia con la que se presentan. Se clasifican en cuanto a las circunstancias que afrontan estas decisiones sea cual sea la situación para decidir y como decidir:

Las **decisiones programadas o estructuradas** son aquellas que se toman frecuentemente, es decir son repetitivas y se convierte en una rutina tomarlas. La persona que toma este tipo de decisión no tiene la necesidad de diseñar una solución, sino que simplemente se rige por la que se ha seguido anteriormente.

Las **decisiones no programadas o no estructuradas** son aquellas que se toman ante problemas o situaciones que se presentan con poca frecuencia, o aquellas que necesitan de un modelo o proceso específico de solución, por ejemplo: ***Lanzamiento de un nuevo producto al mercado***, en este tipo de decisiones es necesario seguir un modelo de toma de decisión para generar una solución específica para este problema en concreto.

En toda empresa o institución existe suele existir una jerarquía que determina el tipo de acciones que se realizan dentro de ella y, en consecuencia, el tipo de decisiones que se deben tomar.

Administrativamente una empresa se puede dividir en tres niveles jerárquicos:

- ✓ Nivel estratégico: alta dirección, planificación global de toda la empresa.
- ✓ Nivel táctico: planificación de los subsistemas empresariales.
- ✓ Nivel operativo: desarrollo de operaciones cotidianas (diarias/ rutinarias).

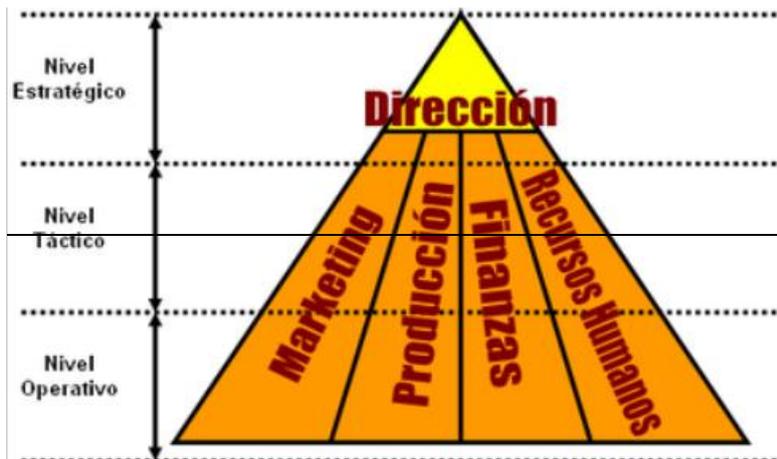


Figura 2. 2 Organización jerárquica y departamental de una empresa.

Conforme se sube en la jerarquía de una organización, la capacidad para tomar decisiones no programadas o no estructuradas adquiere más importancia, ya que son este tipo de decisiones las que atañen a esos niveles. Por tanto, la mayor parte de los programas para el desarrollo de gerentes pretenden mejorar sus habilidades para tomar decisiones no programadas, por regla general enseñándoles a analizar los problemas en forma sistemática y a tomar decisiones lógicas. A medida que se baja en esta jerarquía, las tareas que se desempeñan son cada vez más rutinarias, por lo que las decisiones en estos niveles serán más estructuradas o programadas.

2.2 OLAP

Es el acrónimo en inglés de procesamiento analítico en línea (*On-Line Analytical Processing*). Es una solución utilizada en el campo de la llamada Inteligencia empresarial (o *Business Intelligence*) cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales (o *Cubos OLAP*) que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales (*OLTP*).

Es una categoría de tecnología de software que permite a los analistas, directivos y ejecutivos acceder a los datos de forma rápida, consistente e interactiva a través de una amplia variedad de vistas de la información que han sido obtenidas de datos sin procesar para reflejar la dimensionalidad real de la empresa como la entiende el usuario (Fernández, 2005).

Los datos son clasificados en diferentes dimensiones y pueden ser vistas unas con otras en diferentes combinaciones para obtener diferentes análisis de los datos que contienen.

En estos modelos los datos son vistos como cubos los cuales consisten en categorías descriptivas (*dimensiones*) y valores cuantitativos (*medidas*) (Ibarra, 2005).

Las herramientas OLAP son las aplicaciones que se encargan de formar cubos multidimensionales de este tipo de soporte y de analizarlos con el objetivo de producir y obtener la información más completa para las empresas. Gracias a estas herramientas los usuarios corporativos tienen la oportunidad de sacar el máximo partido a las bases de datos de información.

Esta tecnología es muy utilizada en el área de marketing, ventas, informes, etc., especialmente porque las respuestas a consultas complejas se obtienen muy rápidamente y además porque puede obtener los datos tanto de una fuente externa como de una base interna.

Hay diversos tipos de implementaciones de la tecnología OLAP, las que varían según el tipo de motor en el que se almacenan los datos. De esta manera, las más comunes se clasifican en ROLAP, que almacena los datos en un motor relacional por lo que se puede tener un acceso veloz a ellos y MOLAP, que almacena los datos en una base de datos multidimensional, entre otras.

2.3 Sistemas de apoyo a la toma de decisiones

El concepto de sistema de soporte a las decisiones (*DSS por sus siglas en inglés Decision support system*) es muy amplio, debido a que hay muchos enfoques para la toma de decisiones y debido a la extensa gama de ámbitos en los cuales se toman. Estos sistemas de apoyo son del tipo OLAP o de minería de datos, que proporcionan información y soporte para tomar una decisión.

Los sistemas de apoyo para la toma de decisiones son sistemas que ayudan en el análisis de información de negocios. Su propósito es ayudar a la administración para que **marque tendencias, señale problemas y tome decisiones inteligentes**. La idea básica es recolectar datos operacionales del negocio y reducirlos a una forma que pudiera ser usada para analizar el comportamiento del mismo y modificarlos de una manera inteligente (Pérez, 2006).

Un DSS puede adoptar muchas formas diferentes. En general, se puede decir que un DSS es un sistema informático utilizado para servir de apoyo, más que automatizar, el

proceso de toma de decisiones. La decisión es una elección entre alternativas basadas en estimaciones de los valores de esas alternativas. El apoyo a una decisión significa ayudar a las personas que trabajan solas o en grupo a reunir inteligencia, generar alternativas y tomar decisiones. Apoyar el proceso de toma de decisión implica el apoyo a la estimación, la evaluación y/o la comparación de alternativas. En la práctica, las referencias a DSS suelen ser referencias a aplicaciones informáticas que realizan una función de apoyo (Alter, 1980).

2.4 Almacenes de datos

Los almacenes de datos como comúnmente se conocen (*DW por sus siglas en inglés datawarehouse*), surgen en la década de los años 90 cuando los sistemas de bases de datos operacionales se volvieron insuficientes en cuanto a la profundidad, flexibilidad y robustez en el análisis de los datos.

Un DW es una base de datos corporativa de apoyo a la toma de decisiones que se caracteriza por integrar datos crudos de una o más fuentes distintas, depurando y almacenando la información necesaria de forma organizada para luego procesarla, permitiendo su análisis desde múltiples perspectivas y con grandes velocidades de respuesta. Permite a los directivos que lo utilizan, tener una visión más completa e integral de los procesos dado que el resultado de su implementación es conocimiento acerca del funcionamiento de la organización (Chaudhuri, y otros, 2011).

Bill Inmon definió a principio de los años 90 que un almacén se caracteriza por ser (Inmon, 2005):

- ✓ **Orientado a tema:** Los datos almacenados brindan información sobre un tema o asunto en particular en lugar de concentrarse en la dinámica de las transacciones de la organización.
- ✓ **Integrado:** los datos cargados en el DW pueden provenir de diferentes fuentes y son integrados para dar una visión global coherente.
- ✓ **Variables en el tiempo:** El DW se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones. Lo cual implica que todos los datos deben estar asociados a un período de tiempo específico.

- ✓ **No volátil:** Los datos son estables en el DW, se agregan y modifican datos, pero los datos existentes no son removidos.

El diseño de un almacén de datos debe cumplir dos importantes objetivos. La correcta centralización de los datos que provienen de los sistemas fuentes y proveer a los analistas de una consolidada y rica fuente de información que permita generar reportes que guíen eficientemente a una empresa o institución por el camino del éxito.

Según Ralph Kimball, considerado el principal promotor del enfoque dimensional para el diseño de almacenes de datos, define un DW como: Una copia de los datos transaccionales específicamente estructurados para la consulta y el análisis.

El modelo dimensional es esencialmente una técnica de diseño lógico que busca presentar los datos en una plataforma estándar que sea intuitiva y de alto rendimiento en el acceso a los datos. Es inherentemente dimensional y se adhiere a la disciplina que usa el modelo relacional pero con algunas restricciones. A continuación se describen las principales particularidades del modelo dimensional (Kimball, y otros, 1998).

- ✓ Todo modelo dimensional está compuesto de una tabla con clave múltiple, llamada tabla de hechos y un conjunto de tablas más pequeñas llamadas tablas de dimensión.
- ✓ Cada tabla de dimensión tiene una clave única principal que se corresponde con exactamente uno de los componentes de la clave múltiple de la tabla de hecho.

Una tabla de hechos es la tabla primaria en cada modelo dimensional y está destinada a contener las medidas de la empresa. Existen varios tipos de tablas de hechos entre las cuales están (Kimball, y otros, 2002):

- ✓ **Transaction Fact Tables:** Representan los eventos que suceden en un determinado espacio de tiempo. Se caracterizan por permitir analizar los datos con el máximo nivel de detalle.
- ✓ **Periodic Snapshot Fact Tables:** Son tablas de hechos usadas para recoger información de forma periódica a intervalos de tiempo regulares. Dependiendo de la situación medida o de la necesidad de negocio este tipo de tablas de hechos son una agregación de las anteriores o están diseñadas específicamente.

- ✓ ***Accumulating Snapshot Fact Tables***: Representan el ciclo de vida completo de una actividad o proceso que tiene un principio y un final. Se caracterizan por presentar múltiples tablas de dimensión que caracterizan los eventos presentes en un proceso.

Una tabla de dimensión forma parte del conjunto de tablas de dimensión asociadas a la tabla hechos. Cada dimensión es definida por su clave primaria que sirve como base para la integridad referencial con cualquier tabla de hechos. Las tablas de dimensión contienen atributos con naturaleza no aditiva que son la base para limitar y agrupar dentro de las consultas al almacén de datos. Los tipos de tabla de dimensión más comunes son (Kimball, y otros, 2002):

- ✓ ***Causal dimensions***: Aquellas que proporcionan información adicional sobre la causa de un evento (*por ejemplo promoción en el negocio de las ventas*).
- ✓ ***Multiple date or time-stamp dimensions***: Aquellas usadas para hacer marcas de tiempo en las tablas de hechos, comúnmente en las tablas de hechos *Accumulating Snapshot Fact Tables*.
- ✓ ***Degenerate dimensions***: Son dimensiones que no tienen ningún atributo y por tanto no tienen una tabla específica de dimensión. Solo se incluye para ellas un identificador en la tabla de hechos, que identifica completamente a la dimensión (*por ejemplo el identificador alfanumérico de factura de una venta*).
- ✓ ***Role-playing dimensions***: Se dan cuando una misma dimensión interviene en una tabla de hechos varias veces (por ejemplo la fecha en una tabla de hechos donde se registran varias fechas referidas a conceptos diferentes), es necesario reutilizar la misma dimensión pues no tiene sentido crear tantas dimensiones como uso se hagan de ellas.
- ✓ ***Junk dimension***: Es del tipo de dimensión que agrupa atributos de baja cardinalidad como sexo, estado civil, nivel escolar, entre otros. Al mismo tiempo pueden ser atributos que no guardan relación directa entre ellos a pesar de poder ser agrupados bajo una dimensión.
- ✓ ***Audit dimension***: Se crean cuando se requiere mantener una traza de los datos para propósitos de auditoría y calidad.

Vale aclarar que una dimensión puede presentar al mismo tiempo dos o más de los tipos anteriormente expuestos. En el caso de las dimensiones que presentan cambios (*Causal dimension*) existen varias técnicas, en dependencia de los requerimientos del negocio, para gestionar adecuadamente los datos. Este tipo de dimensiones son también conocidas como dimensiones de cambio lento (SCD por sus siglas en inglés *Slowly Change Dimension*). A continuación se describen las principales técnicas (Kimball, y otros, 2004):

- ✓ ***Slowly Change Dimension Type 1:*** Sobre-escribir. Cuando se produce un cambio en los valores de un atributo, se sobre-escribe el valor antiguo con el nuevo sin registrar una historia. Esto significa perder toda la historia del dato, y cuando se haga un análisis se verá la información histórica desde el punto de vista actual.

Ejemplo para una dimensión proveedor:

Llave_proveedor	Código_proveedor	Nombre_proveedor	Provincia_proveedor
123	PRO_1	JUAN	VILLA CLARA

Si el proveedor cambia su localización geográfica, la tabla actualizada debe sobre-escribir este registro.

Llave_proveedor	Código_proveedor	Nombre_proveedor	Provincia_proveedor
123	PRO_1	JUAN	PLACETAS

- ✓ ***Slowly Change Dimension Type 2:*** Añadir fila. Cuando hay un cambio, se crea un nuevo registro en la tabla. El nuevo registro tiene una nueva clave primaria de forma que, por ejemplo un cliente, puede tener varios registros en la tabla de dimensión según se van produciendo los cambios. Los cambios en el caso de un cliente pueden ser de cualquiera de sus datos personales. Adicionalmente se puede incluir un versionado que contemple las fechas de validez de los registros o bien una bandera que marque el registro activo.

Llave_proveedor	Código_proveedor	Nombre_proveedor	Provincia_proveedor	Fecha_desde	Fecha_hasta
123	PRO_1	JUAN	VILLA CLARA	01/ene/2012	

La actualización de la localización geográfica en este caso sería como sigue.

Llave_proveedor	Código_proveedor	Nombre_proveedor	Provincia_proveedor	Fecha_desde	Fecha_hasta
123	PRO_1	JUAN	VILLA CLARA	01/ene/2012	23/dic/2012
123	PRO_1	JUAN	PLACETAS	23/dic/2012	

- ✓ *Slowly Change Dimension Type 3*: Añadir columna. Cuando se produce un cambio, se guarda el valor anterior en una columna distinta, actualizando el campo con el nuevo valor. De esta forma para cada fila se tiene el valor anterior y el valor actual lo cual indica que se tendrá una historia limitada del dato. El número de columnas depende del nivel de antigüedad que se desee guardar.

Llave_proveedor	Código_proveedor	Nombre_proveedor	Provincia_original	Fecha	Provincia_actual
123	PRO_1	JUAN	VILLA CLARA	01/ene/2012	VILLA CLARA

La actualización del hecho se limita a la versión anterior, como se muestra a continuación.

Llave_proveedor	Código_proveedor	Nombre_proveedor	Provincia_original	Fecha	Provincia_actual
123	PRO_1	JUAN	VILLA CLARA	01/ene/2012	PLACETAS

Una segunda actualización provocaría la pérdida de información. En otra variante pudiera implementarse de manera tal que mantuviera las dos últimas actualizaciones en vez de la última y la original. Adicionándole más columnas se puede definir la cantidad de actualizaciones que se requiere mantener.

Llave_proveedor	Código_proveedor	Nombre_proveedor	Provincia_original	Fecha	Provincia_actual
123	PRO_1	JUAN	VILLA CLARA	01/ene/2012	REMEDIOS

A lo anterior se puede agregar que las tablas de hechos contienen campos llamados medidas, los cuales aportan el valor cuantitativo del hecho en cuestión, así como las dimensiones aportan el valor cualitativo.

De esta manera cada tabla de hecho se relaciona con sus correspondientes dimensiones además de la medida/(s) que contiene y así se conforma lo que en la teoría se conoce como *Cubo Multidimensional de Datos*.

Los cubos multidimensionales están diseñados para soportar el análisis de grandes volúmenes de datos, cuando por el contrario en las bases de datos operacionales se busca la optimización y normalización del diseño, en los almacenes de datos cambian los principios debido a la gran cantidad de datos que deben almacenar. Cuestiones como la eliminación de relaciones foráneas que complejizan la gestión de los datos en las

consultas, son relativamente desechadas a la hora de diseñar bases de datos para el soporte a la toma de decisiones.

A continuación, **la figura 2.3**, muestra una vista de la estructura general de una tabla de hechos con dimensiones y medidas.

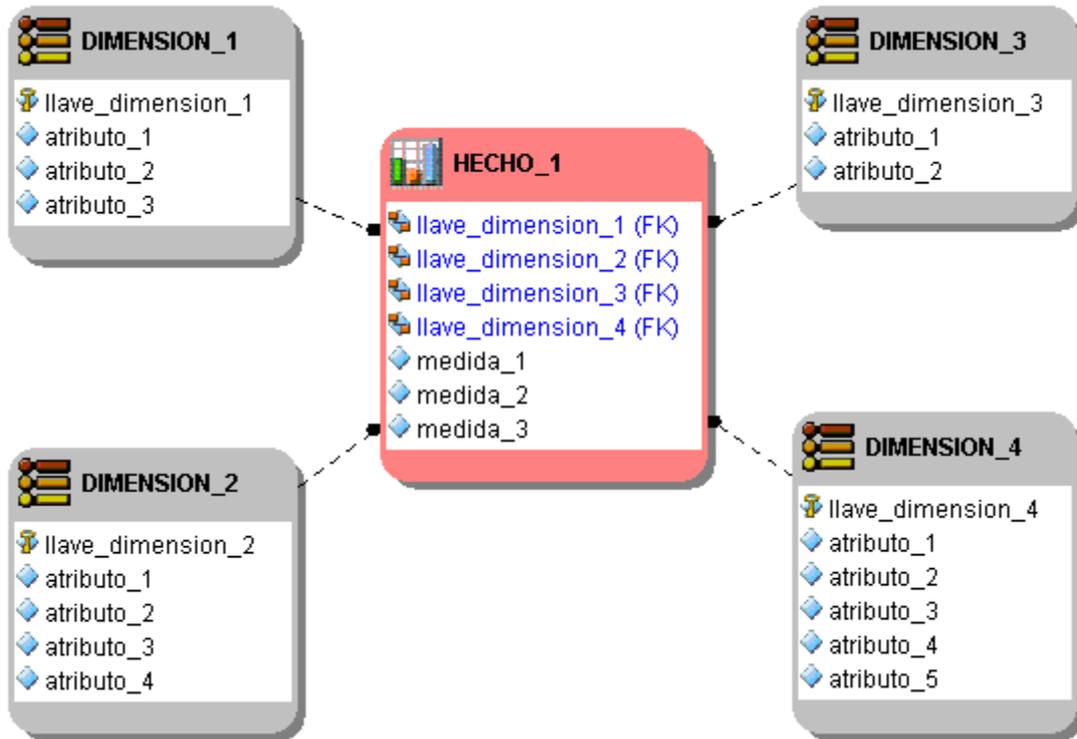


Figura 2.3 Vista general de una tabla de hechos con dimensiones y medidas.

De manera general, las bases de datos multidimensionales se utilizan principalmente para crear aplicaciones OLAP y pueden verse como bases de datos de una sola tabla, su peculiaridad es que por cada dimensión tienen un campo (o columna), y otro campo por cada métrica o hecho, es decir estas tablas almacenan registros cuyos campos son de la forma:

$$d1, d2, d3, \dots, f1, f2, f3, \dots$$

Donde los campos d_i hacen referencia a las dimensiones de la tabla, y los campos f_i a las métricas o hechos que se quiere almacenar, estudiar o analizar.

Cada una de estas tablas puede asimilarse a un hipercubo, o más concretamente, si de herramientas OLAP se trata, a un cubo OLAP, donde las dimensiones del mismo se corresponden con los campos de dimensiones de la tabla (campos $d_i \dots$) y el valor

almacenado en cada celda del cubo equivale a la métrica o métricas (campos *fi...*) almacenadas en la tabla.

2.5 Integración de datos

El principal propósito de esta capa consiste en fusionar los datos de heterogéneas plataformas en un formato estándar para la base de datos destino (T. Moss, y otros, 2003).

La integración de datos es un proceso que consta de tres actividades principales que se realizan sobre los datos, extracción, transformación y carga (*ETL* como se conoce en inglés *extract, transformation and load*). Los pasos a seguir para realizar estas actividades se muestran en **la figura 2.4**.

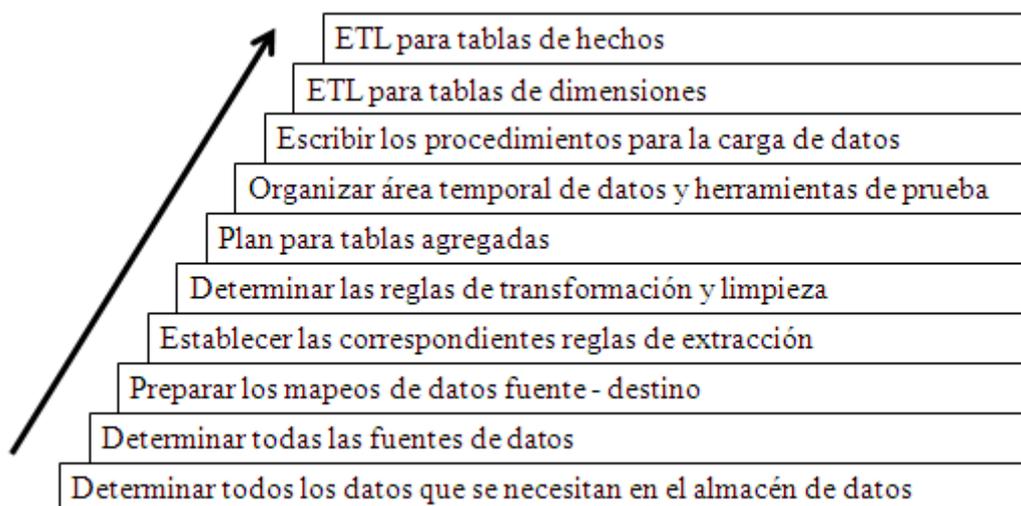


Figura 2. 4 Principales pasos en el proceso de ETL.

Los especialistas de ETL tienen una misión de alto grado de complejidad que consiste de manera general en construir la trastienda del almacén de datos (Kimball, y otros, 2004). Específicamente:

- ✓ Proporcionar los datos de manera más eficiente para las herramientas de usuarios finales.
- ✓ Adicionar valor a los datos en los pasos de limpieza y consolidación.
- ✓ Proteger y documentar el linaje de los datos.
- ✓ Asegurar la calidad y limpieza de los datos.

- ✓ Conformar etiquetas y medidas en los datos para lograr la consistencia entre las fuentes originales.
- ✓ Entregar los datos en un formato físico que pueda ser usado por las herramientas de consulta, redactores de informe y cuadro de mando.

2.6 Análisis multidimensional de datos

Hoy en día, la información es considerada como un factor productivo, fundamental para la toma de decisiones. Procesos de negocio que llevaban días, hoy toman segundos y generan un gran volumen de datos. Las estadísticas muestran que la cantidad de datos en el mundo crece exponencialmente año tras año.

Uno de los objetivos principales de diseñar un almacén de datos es proveer a los analistas de las herramientas necesarias para generar reportes que permitan a los gerentes o empresarios tomar las decisiones correctas en el momento correcto, de manera tal que las empresas o instituciones asciendan competitivamente.

Todos los reportes tributan siempre a una línea central de análisis que se puede ver marcada por tres fases. ¿Qué sucedió? ¿Qué está sucediendo? ¿Qué debe suceder?

El análisis multidimensional está orientado a buscar patrones de comportamiento, tendencia e incluso predicción, todo esto, siempre sobre una base de datos multidimensional. Cuestiones como: ¿Quiénes son mis mejores clientes? ¿Dónde se ubican? ¿Qué compran? ¿Qué los caracteriza? ¿Cuáles líneas de productos son más rentables? ¿Qué productos son mis top 10? ¿Cuáles mis peores 10? ¿Qué productos son los más vendidos durante este mes, y cuál fue la variación respecto al mes anterior? ¿Cómo van mis avances respecto a lo planificado? ¿Cómo deben comportarse mis ventas teniendo en cuenta los mismo meses de años anteriores? Son los principales requerimientos para los especialistas en inteligencia de negocio involucrados en el diseño de una capa de reportes multidimensionales.

Las aplicaciones de BI facilitan muchas actividades como son: el análisis multidimensional; análisis de *click-stream*; minería de datos; pronosticar; análisis de negocio; preparación de cuadros de mando integrales; visualización; consultas, informes y gráficas (incluso en tiempo real); análisis *geo-espacial*; gestión del conocimiento; minería para texto, contenido y voz entre otras (T. Moss, y otros, 2003).

2.7 Buenas prácticas para el mantenimiento de actualizaciones de los hechos

Lidiar con la posibilidad de que un hecho cambie su verdad y la manera en que se debe solucionar este problema desde el punto de vista de un diseñador de bases de datos multidimensionales es un tema que se encuentra en actual discusión.

Las soluciones propuestas no son triviales y presentan sus ventajas y desventajas, adicionalmente, estas soluciones están atadas a requerimientos específicos por tal caso los autores siempre sugieren analizar bien las características y requerimientos del negocio en cuestión antes de tomar una decisión.

2.7.1 Cambio lento tipo 2 en la tabla de hechos

Ralph Kimball es reconocido como uno de los padres de las soluciones que dan soporte a la toma de decisiones (sistemas de inteligencia de negocio), debido a que se ha dedicado a ofrecer buenos consejos en torno a los problemas particulares que se van volviendo comunes. En los primeros libros de Kimball no aparecen referencias del fenómeno que da lugar a esta investigación.

En su consejo de diseño número 74 (Kimball, 2005), presenta una manera muy efectiva de implementar el cambio lento tipo 2 en los hechos. Esta variante está enfocada a mostrar el comportamiento tanto de un hecho como de una dimensión ante el cambio lento y tiene como objetivo mantener trazas de las transacciones realizadas ante una posible inspección o auditoría en los datos.

La necesidad de mantener trazas de las versiones de los hechos viene siendo uno de los problemas más comunes hoy en día en el campo de la inteligencia de negocio y los sistemas para el soporte a la toma de decisiones.

Este consejo de Kimball es la base de la solución propuesta en este documento. Aplicar el cambio lento tipo dos en las tablas de hechos puede parecer riesgoso y engorroso, sin embargo, implementando otras técnicas no siempre se tiene lo que se desea desde el punto de vista de diseño.

La figura 2.5 muestra un ejemplo de diseño donde la tabla de hechos experimenta el cambio lento tipo dos.

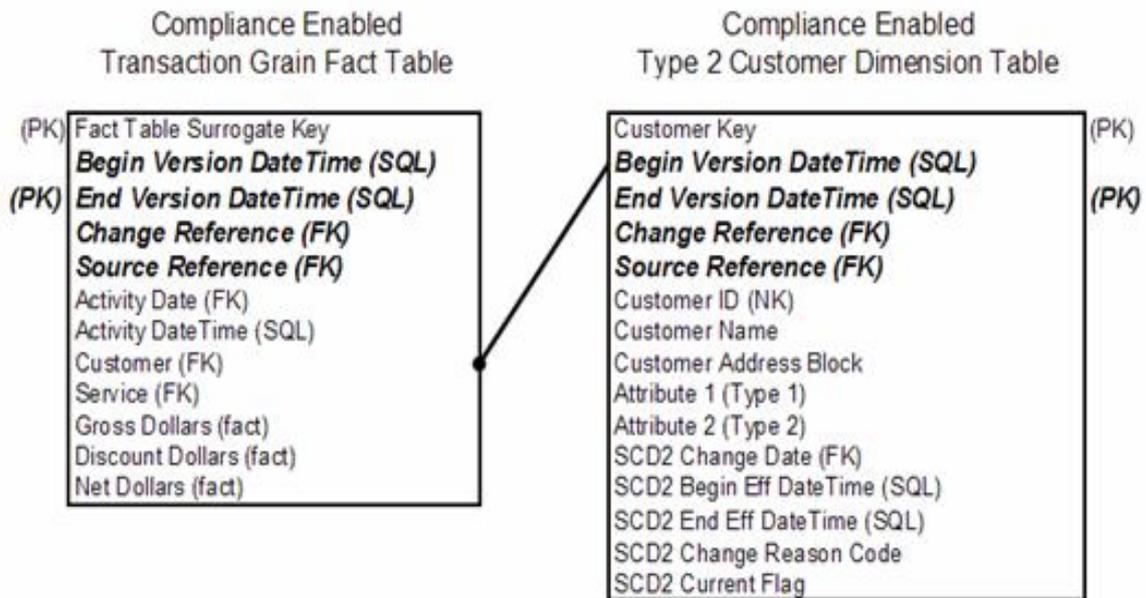


Figura 2. 5 Ejemplo de Ralph Kimball.

En este consejo de diseño, Ralph Kimball, hace especial énfasis en que solo el cambio lento de tipo 2 debe aplicarse debido a que por su implementación es la variante que permite almacenar todas las actualizaciones que pueda presentar un hecho. Para la tabla de hecho incluye una llave de sustitución (KIMBALL, 1997), la cual combinada con el atributo *End Version Date Time* conforman una llave primaria en la tabla de hechos, este ejemplo está enfocado a demostrar como el cambio lento tipo dos se administra de manera similar tanto en una tabla de hechos como en una dimensión (KIMBALL, 2005).

2.7.2 Tabla auditoria para el control de cambios

Chris Adamson, especialista en Data Warehousing, provee consultoría a través de su compañía Oakton Software LLC. Según Chris Adamson (ADAMSON, 2011), aplicar el cambio lento tipo dos en la tabla de hechos no es la mejor opción porque perjudica seriamente la capacidad de uso de esta, ya que a la hora de consultar la tabla se debe tener siempre en cuenta no mezclar los registros válidos con los inválidos, en contraposición propone manejar las actualizaciones de los hechos mediante la creación de una tabla de auditoría que almacene de manera independiente los cambios surgidos en un hecho determinado.

En esta situación en autor parte de la idea de que solo hay una versión de la verdad y solo esta es la que debe ser almacenada en la tabla de hechos, consecuentemente

propone crear una tabla de auditoría con la cual se lleve una traza de los cambios ocurridos en los hechos.

Según Chris Adamson la tabla auditoria debe contener el identificador de la transacción, la fecha y el valor de cada hecho. De esta manera se almacena el registro original en la tabla de hechos y cada una de las actualizaciones en la tabla de auditoría. **La figura 2.6** muestra el diseño para almacenar las actualizaciones de los hechos en una tabla auditoría independiente.

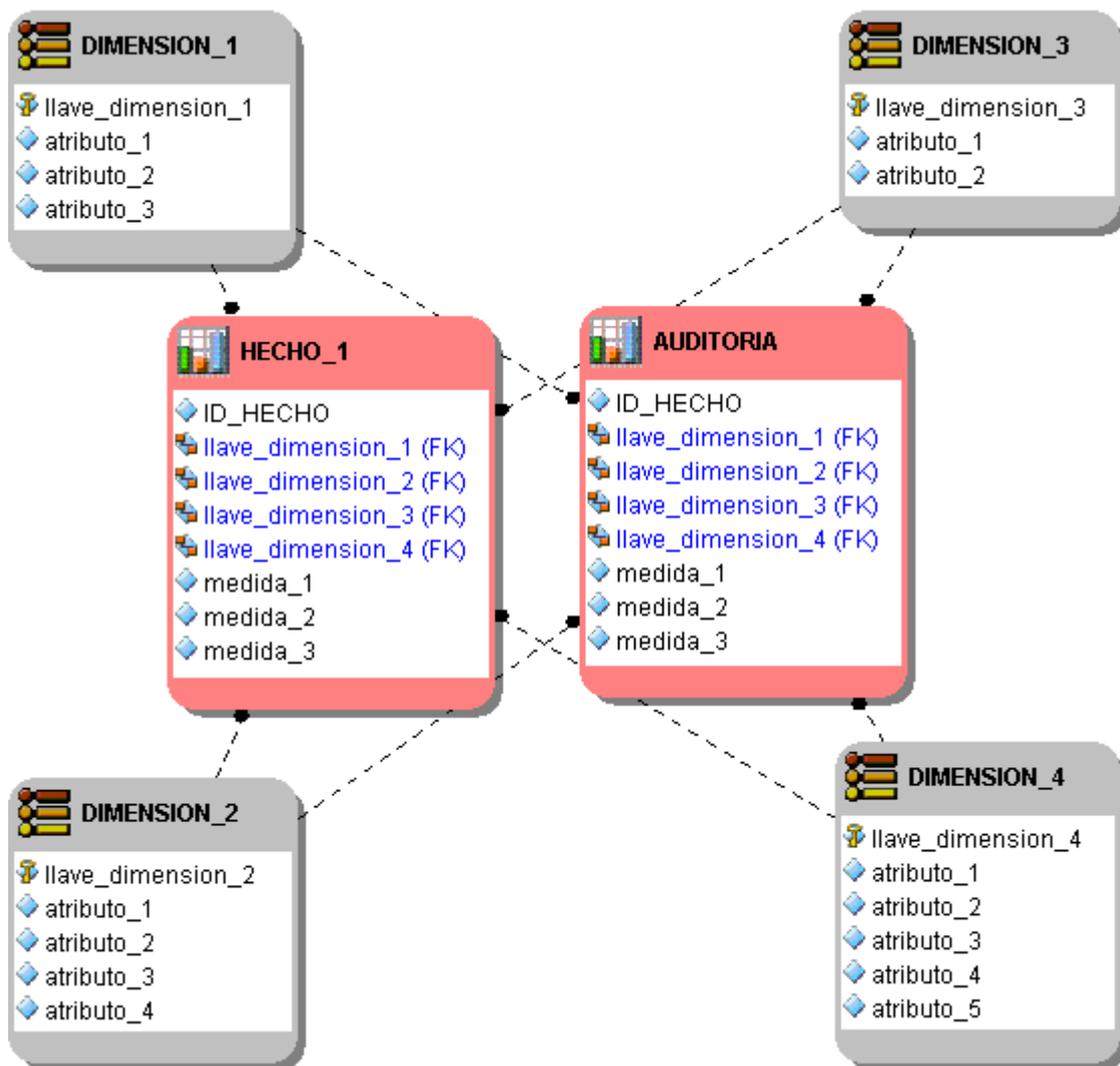


Figura 2. 6 Tabla de hechos con tabla auditoria para Control de cambio.

Finalmente el autor propone que dicha tabla de auditoría puede verse como una tabla de hechos con sus respectivas dimensiones asociadas, respetando así los principios del modelo multidimensional.

Si bien es cierto que la tabla de hecho original siempre estará limpia de registros

inválidos u obsoletos y crear una tabla para mantener la traza de los cambios se ve como una propuesta interesante, hay casos en los que esta variante puede provocar muchos dolores de cabeza a la hora de realizar consultas.

Supóngase que la última versión de cada hecho es la que realmente vale, o sea cuando en cada análisis lo que siempre se desee es la última versión de la verdad de cada hecho involucrado en la consulta. En este caso habría que incluir constantemente las dos tablas de hechos, original y auditoria, en la consulta.

Las uniones entre tablas es una de las tareas que mayor esfuerzo requiere del gestor de base de datos y más en estos casos donde el volumen de datos se vuelve exageradamente grande.

Por esta cuestión es importante analizar a profundidad las características del negocio y decidir si es conveniente mantener el histórico de actualizaciones en tablas separadas. El propio autor de este consejo sugiere fuertemente valorar bien las características del negocio y los requerimientos del mismo.

En el ejercicio de diseño de una solución de inteligencia de negocio se debe dedicar el tiempo necesario al análisis y entendimiento del problema en cuestión, esto al principio puede parecer pérdida de tiempo, pero a la hora de desarrollar la solución se tiene un camino iluminado, se puede saber con antelación los pasos a seguir en la implementación de la solución. Esto es factor clave en el éxito de una solución para el soporte a la toma de decisiones.

2.7.3 Tablas snapshots

Periodic Snapshot Fact Tables: Son tablas de hecho usadas para recoger información de forma periódica a intervalos de tiempo regulares. Dependiendo de la situación medida o de la necesidad de negocio este tipo de tablas de hecho son una agregación de las tablas de hechos transaccionales (CURTO, 2008), (MUNDY, 2012).

El diseño de cubos ***snapshot*** provee un estado de los datos en un momento determinado (al final de cada mes, al final de cada semana, al final de cada día) (MUNDY, 2012), (KIMBALL, Y OTROS, 2002).

Este tipo de diseño para tablas de hechos es muy práctico en casos donde se desee tener en cierto momento el estado general de una actividad o proceso, por ejemplo los

balances de cuenta, el libro mayor entre otras actividades en el ámbito de la gestión económica, se puede aplicar también cuando interesa analizar datos socio demográficos en un país o región, entendiéndose cuando se desea almacenar los datos de un censo de población y vivienda lo cual se realiza cada cierto periodo de tiempo y los análisis siempre están enfocados a determinar cuánto han variado las poblaciones de un tiempo a otro.

Con esta variante se estaría almacenando cada bloque de registros con una marca de fecha en la cual se integraron. En este caso las medidas de un bloque no se pueden sumarizar con las medidas de otro bloque debido a que cada bloque solo refleja el estado actual de todos los hechos. En cada integración se almacena una foto general de todos los hechos independientemente de si sufrieron algún cambio o no. **La figura 2.7** muestra el diseño de un cubo snapshot para manejar las actualizaciones de los hechos.

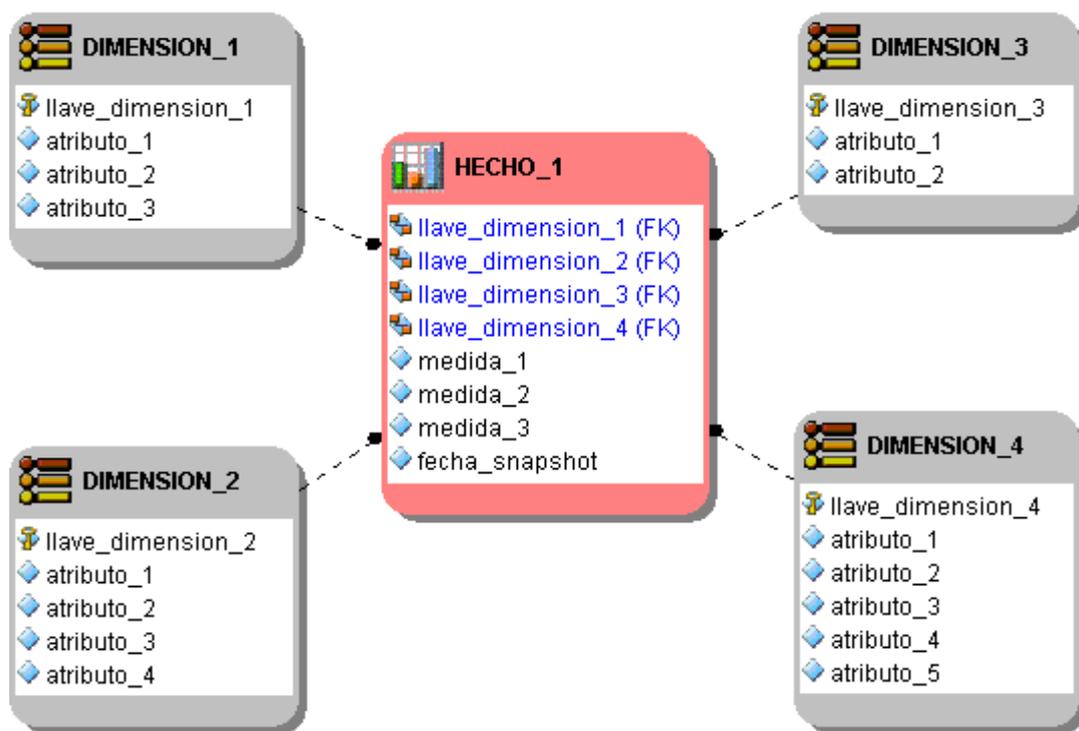


Figura 2. 7 Tabla snapshot para el almacenamiento de actualizaciones de los hechos.

Si bien se puede tener almacenado un nuevo estado de los datos del sistema cada un período de tiempo determinado, con esta variante se estarían almacenando registros duplicados por la sencilla razón de que existan transacciones que no han presentado una nueva versión de la verdad, esto se resume en que se corre el riesgo de que se genere un volumen de datos exageradamente grande y redundante (Adamson, 2011).

2.8 Herramientas de desarrollo

En este epígrafe se hace una descripción de las herramientas de desarrollo utilizadas para construir la solución completa que apoya al mercado de datos diseñado.

2.8.1 Pentaho Data Integration

Pentaho Data Integration (*PDI*) es la herramienta usada para el proceso de recopilación de los datos. Extraer, transformar y cargar los datos al almacén es una tarea de alto grado de complejidad. PDI provee al desarrollador las funciones necesarias para el cumplimiento de estas tareas.

- ✓ Simple diseñador visual para mejorar la productividad del desarrollador.
- ✓ Conectividad con los principales gestores de base de datos tales como (Oracle, PostgreSQL, MySQL, incluso para base de datos NoSQL como MongoDB, Cassandra y Hbase).
- ✓ Procesamiento con gran velocidad de grandes volúmenes y variedad de datos.
- ✓ Perfilado y calidad de datos.

Vale aclarar que la versión usada para solucionar el problema es *Pentaho Data Integration 4.1 community*.

2.8.2 Oracle Database 10g R2 Enterprise Edition

Por la confiabilidad y responsabilidad de los proveedores de los productos Oracle, el Observatorio usa como repositorio central de datos el producto Oracle Database 10g R2 en clúster de dos servidores.

Este producto provee además funciones para el particionado y la minería de datos. Con Oracle Database se cubren todas las necesidades de almacenamiento que requiere una solución de inteligencia de negocio. Entre otras cosas Oracle provee:

- ✓ Escalabilidad y rendimiento.
- ✓ Usabilidad.
- ✓ Resguardo y restauración.
- ✓ Alta disponibilidad.

- ✓ Gestión de contenidos.
- ✓ Seguridad.
- ✓ Integridad.
- ✓ Particionado.

Todas estas características se explican con mayor profundidad en los sitios oficiales de Oracle.

2.8.3 Oracle Warehouse Builder 10g

Oracle Warehouse Builder es una herramienta única y completa para todos los aspectos de la gestión de datos. Warehouse Builder aprovecha la base de datos Oracle para transformar los datos en información de alta calidad.

Proporciona calidad de datos, auditoría de datos, total integración relacional y modelado dimensional, en un ciclo completo de vida de los datos y metadatos. Warehouse Builder le permite crear almacenes de datos, migración de datos desde sistemas heredados, la consolidación de los datos de fuentes de datos diversos, limpiar y transformar los datos para proporcionar información de calidad y gestionar los metadatos de las empresas.

2.8.4 PL / SQL Developer

PL / SQL Developer es un entorno de desarrollo integrado que está específicamente dirigido a la creación de unidades de programa almacenadas en bases de datos Oracle. Con el tiempo se ha visto que más lógica de negocio y de aplicación se ha trasladado al servidor de Oracle, por lo que la programación en **PL / SQL** se ha convertido en una parte importante del proceso de desarrollo. **PL / SQL Developer** se centra en la facilidad de uso, código de calidad y productividad, aspectos claves durante el desarrollo de aplicaciones de Oracle.

2.8.5 Oracle Discoverer

Oracle Business Intelligence Discoverer es un juego de herramientas de consulta ad hoc intuitiva, análisis, informes y publicación en Web que proporciona a los usuarios de negocios acceso inmediato a la información de las bases de datos.

Oracle Business Intelligence Discoverer permite a los usuarios de negocios en todos los niveles de la organización tomar decisiones de negocio más rápidas y mejor fundadas. Con cualquier explorador Web estándar se tiene acceso seguro e inmediato a los datos de orígenes de datos relacionales y multidimensionales.

Oracle Business Intelligence Discoverer proporciona una vista del negocio para ocultar la complejidad de las estructuras de datos subyacentes y, por tanto, permite centrarse en resolver los problemas de los negocios.

2.9 Conclusiones Parciales

En este capítulo se presentaron los principales conceptos que giran en torno a la inteligencia de negocios, se describieron las herramientas utilizadas para el desarrollo de la solución y se exponen las buenas prácticas para el diseño de un almacén de datos en situaciones donde se requiera almacenar las actualizaciones de los hechos.

CAPÍTULO III “Mercado de datos para el seguimiento de hechos delictivos”.

En el presente capítulo se describen las principales características del mercado de datos para hechos delictivos, así como de la capa de análisis y de integración de datos. Esta arquitectura puede ser aplicada a todos los entornos donde los hechos estén sujetos a cambios que por cuestiones del negocio se requieran almacenar.

Complementariamente se mostrarán y explicarán imágenes de las capas de integración de datos y análisis multidimensional, describiendo las principales particularidades que presentan estas dos importantes capas.

Respecto a esto, la capa de integración debe estar concebida para gestionar los hechos nuevos a integrar más los hechos que cambian, así mismo la capa de consultas y reportes debe tener siempre en cuenta los filtros para acceder a la información actual, o sea la última actualización de cada hecho, además de poder analizar la evolución de uno o varios hechos.

3.1 Arquitectura del Sistema de Información Nacional de Seguridad Ciudadana

A continuación se muestra una imagen de la arquitectura general de SINSEC la solución de inteligencia de negocio implementada en el CTAISC. De izquierda a derecha la imagen muestra el conjunto de fuentes de datos (órganos de seguridad del estado), los repositorios temporales de extracción, transformación y carga (*ETL*), el repositorio central para el almacenamiento histórico de los hechos y las actualizaciones y finalmente la capa de análisis y visualización a través de reportes de texto, gráficos y mapas. **La figura 3.1** muestra la arquitectura de software implementada en CTAISC.

La mayoría de las fuentes de datos están en formato *MICROSOFT OFFICE EXCEL*². En el caso de los datos provenientes de CICPC la conexión es entre sistemas de base de datos. Como se puede apreciar todo el modelo de datos de SINSEC-DI2 está soportado sobre bases de datos *Oracle Database 10g R2 Enterprise Edition*.

² MICROSOFT OFFICE EXCEL: Producto del paquete de herramientas de oficina que provee MICROSOFT.

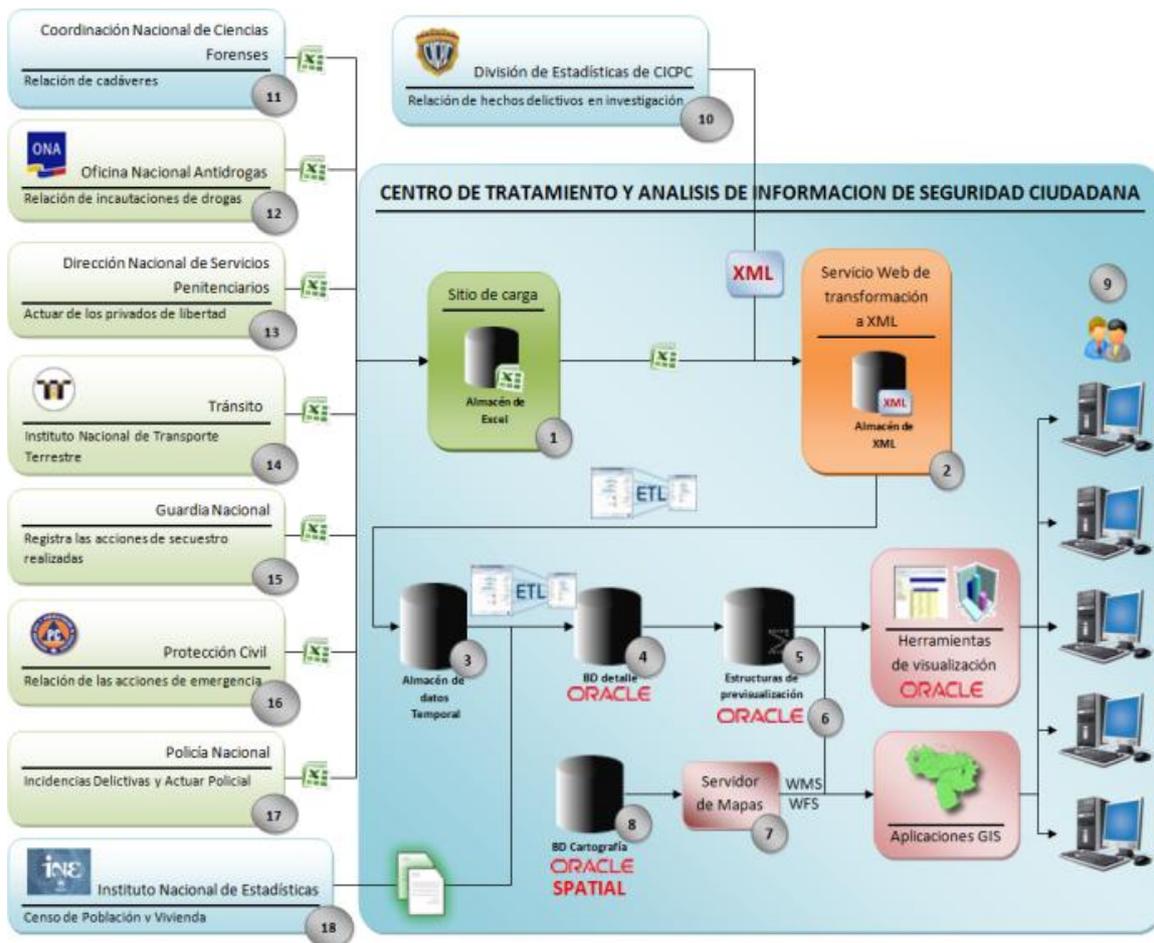


Figura 3. 1 Arquitectura de la solución de software.

Con esta solución el CTAISC puede desempeñar efectivamente todas las tareas que está llamado a cumplir y con las cuales esta comprometidas ante el gobierno venezolano, los órganos de seguridad del estado y el país.

El diseño de la solución de inteligencia de negocio para el CTAISC donde los hechos presentan cambio lento fue implementado para todos los casos de uso. Este documento se centra en los hechos delictivos que envía diariamente el CICPC al CTAISC.

3.2 Selección del proceso de negocio

El proceso de negocio consiste en el seguimiento a los hechos delictivos.

Cada día el sistema operacional de CICPC publica para el consumo exclusivo de CTAISC los datos relativos a los hechos delictivos ocurridos. En cada uno de estos bloques hay hechos nuevos y actualizaciones. Los sub procesos para convertir estos datos en información que brinde soporte a la toma de decisiones tienen el siguiente orden:

3.2.1 Publicación de datos

Cada noche en horas de la madrugada se ejecuta una tarea programada que hace correr un script de código SQL. Este script hace una consulta al sistema operacional de CICPC e inserta el resultado en una tabla la cual tiene permisos de acceso desde el servidor de ETL localizado en CTAISC.

3.2.2 Extracción, transformación y carga de datos

El flujo de ETL se dispara unas horas después de ejecutarse el script SQL anteriormente mencionado. Este proceso consiste en conectarse al sistema fuente, extraer los datos, transformar los que requieran serlo y finalmente cargar el almacén con los nuevos hechos y las actualizaciones de los que ya existen.

3.2.3 Elaboración de los reportes

Los reportes están diseñados a la medida de los requerimientos, tanto del despacho del vice ministerio como de los Órganos de Seguridad del Estado. Al día siguiente en horas de la mañana se actualizan los reportes, lo cual recargar los datos frescos.

3.3 Granularidad

Se desea llevar un control de los hechos delictivos que consisten en la **cantidad de lesionados, fallecidos y detenidos** en un hecho **por fecha** en que ocurrió, **ubicación geográfica, perfil de la persona** que cometió el delito, **lugar** en que ocurrió, **órgano de seguridad** que atiende el caso delictivo, **ley** que se viola en el caso y el **expediente policial** del mismo.



Column Name	Relationship
dim_exp	(FK)
dim_per_prf	(FK)
dim_ley	(FK)
dim_fch	(FK)
dim_geo	(FK)
dim_lug	(FK)
dim_org_seg	(FK)
dim_aud	(FK)
nro_fallecidos	
nro_lesionados	
nro_detenidos	
fecha_desde	
fecha_hasta	

Figura 3. 2 Tabla de hechos Delitos de violación a la ley.

La dimensión auditoría se crea para propósitos de trazabilidad en los flujos de integración. Los campos *fecha_desde* y *fecha_hasta* son usados conjuntamente con el campo llave foránea *dim_exp* para llevar el control y almacenamiento de las actualizaciones.

3.4 Dimensiones

Ley: Almacena el tesoro delictivo de Venezuela. Todos los delitos agrupados por *clase* y *tipo*, incluso mantiene una marca en el atributo *plan_seguridad* para los delitos que están considerados como los que requieren mayor atención. En orden jerárquico la dimensión está organizada por los niveles *clase*, *tipo* y *delito*.



Figura 3. 3 Dimensión Ley.

Fecha: Almacena la fecha en la que ocurren los hechos. Cada uno de los atributos representa un nivel dentro de la dimensión. La dimensión presenta dos jerarquías, *mes* (que agrupa los niveles *año*, *semestre*, *trimestre*, *mes* y *día*) y *semana* (que agrupa los niveles *año*, *semana* y *día*). Los registros que contiene esta estructura datan desde el 01/01/1990 hasta la fecha, la intención de esto es poder tener almacenados los hechos ocurridos antes del gobierno bolivariano y después de este.



Figura 3. 4 Dimensión Fecha.

Geografía: Almacena la información geográfica de un hecho en cuanto a estado, municipio y parroquia. Con esta dimensión se pueden ubicar geográficamente las áreas de mayor incidencia delictiva.

Esta estructura de cambio lento tipo 2 ya que almacena todas las distribuciones político territoriales que tienen lugar en Venezuela. Mantener el historial de estos registros da la posibilidad de analizar los índices de seguridad ciudadana de cierto lugar a pesar de haber cambiado de nombre. Está organizada jerárquicamente por los niveles Estado, Municipio y Parroquia.



Figura 3. 5 Dimensión Geografía.

Lugar: Almacena de manera general los lugares y zonas donde ocurren los hechos. Esta dimensión existe por la necesidad de tener un mayor nivel de detalle en cuanto a la localización de un hecho delictivo, si bien la dimensión *Geografía* aporta la ge localización de un hecho, esta dimensión *Lugar* almacena datos adicionales que brindan información más exacta en cuanto al comportamiento y las tendencias de los hechos delictivos.

Los atributos de esta dimensión no guardan relación entre sí, son atributos de baja cardinalidad por lo cual se agrupan bajo una misma estructura. Dada estas características se resume que la dimensión *Lugar* es de tipo *Junk*.



Figura 3. 6 Dimensión Lugar.

Perfil de persona: Almacena los datos referentes al perfil físico social de una persona natural. Los niveles en esta dimensión no guardan relación directa, sin embargo, es práctico mantenerlos bajo una misma estructura para evitar una excesiva cantidad de dimensiones en el diseño dimensional. Dadas las características de esta estructura se presenta como una dimensión de tipo *Junk*.



Figura 3. 7 Dimensión Perfil de Persona.

Órgano de Seguridad: Almacena los atributos referentes a las características de los Órganos de Seguridad de Estado que intervienen en los hechos. Dada la ocurrencia de un delito son las sub-delegaciones geográficamente distribuidas por todo el país quienes acuden inmediatamente a atender el caso. Las autoridades encargadas registran en el sistema operacional los detalles de la declaración y al mismo tiempo asocian el caso con la sub-delegación y delegación que le da seguimiento.

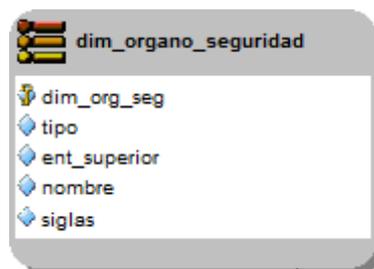


Figura 3. 8 Dimensión Órgano de Seguridad.

Expediente: Almacena los atributos referentes a los expedientes de los hechos ocurridos, incluyendo el identificador del expediente. Este identificador es el campo por el cual se buscan los hechos que presentan cambios. Además funge como dimensión conformada en el modelo dimensional y presenta cambio lento tipo 2.

El atributo relevancia (*alta, media, baja*) es de gran utilidad pues conjuntamente con el atributo *plan_seguridad* de la dimensión *Delito* se puede establecer una tabla de clasificación de hechos delictivos en cuanto a impacto directo en la seguridad ciudadana.



Figura 3. 9 Dimensión Expediente.

Auditoría: Almacena los metadatos de los registros insertados (fuente, hora de integración, vía por la cual se integra, tipo de documento desde el cual se integra y nombre del documento desde el cual se integra). Esta dimensión juega un papel importante a la hora de hacer una revisión de los volúmenes de datos integrados diariamente. Para propósitos del negocio esta dimensión no tiene implicación, solo existe por el hecho de mantener un traza de la gestión de los datos.



Figura 3. 10 Dimensión Auditoría.

3.5 Tabla de hechos

Los campos de datos cuantitativos son:

- ✓ Cantidad de fallecidos.
- ✓ Cantidad de lesionados.
- ✓ Cantidad de detenidos.

Las tres medidas son aditivas. Estos campos están igualmente sujetos a cambios. Si ocurre un robo en el que no se ha detenido a alguien, el valor de la cantidad de detenidos será cero, luego de un proceso de investigación este valor puede variar en

dependencia de si se capturan al victimario. Igualmente la cantidad de fallecidos y lesionados reportados a priori puede resultar alterada en una investigación donde se descubra que cierta cantidad de estos no resultaron fallecidos o lesionados por causa del hechos delictivo en cuestión.

A continuación se muestra en **la figura 3.11** el diseño de la tabla de hechos que almacena las violaciones a la ley (hechos delitos), garantizando además el almacenamiento de las posibles actualizaciones de los hechos que presentan cambios.

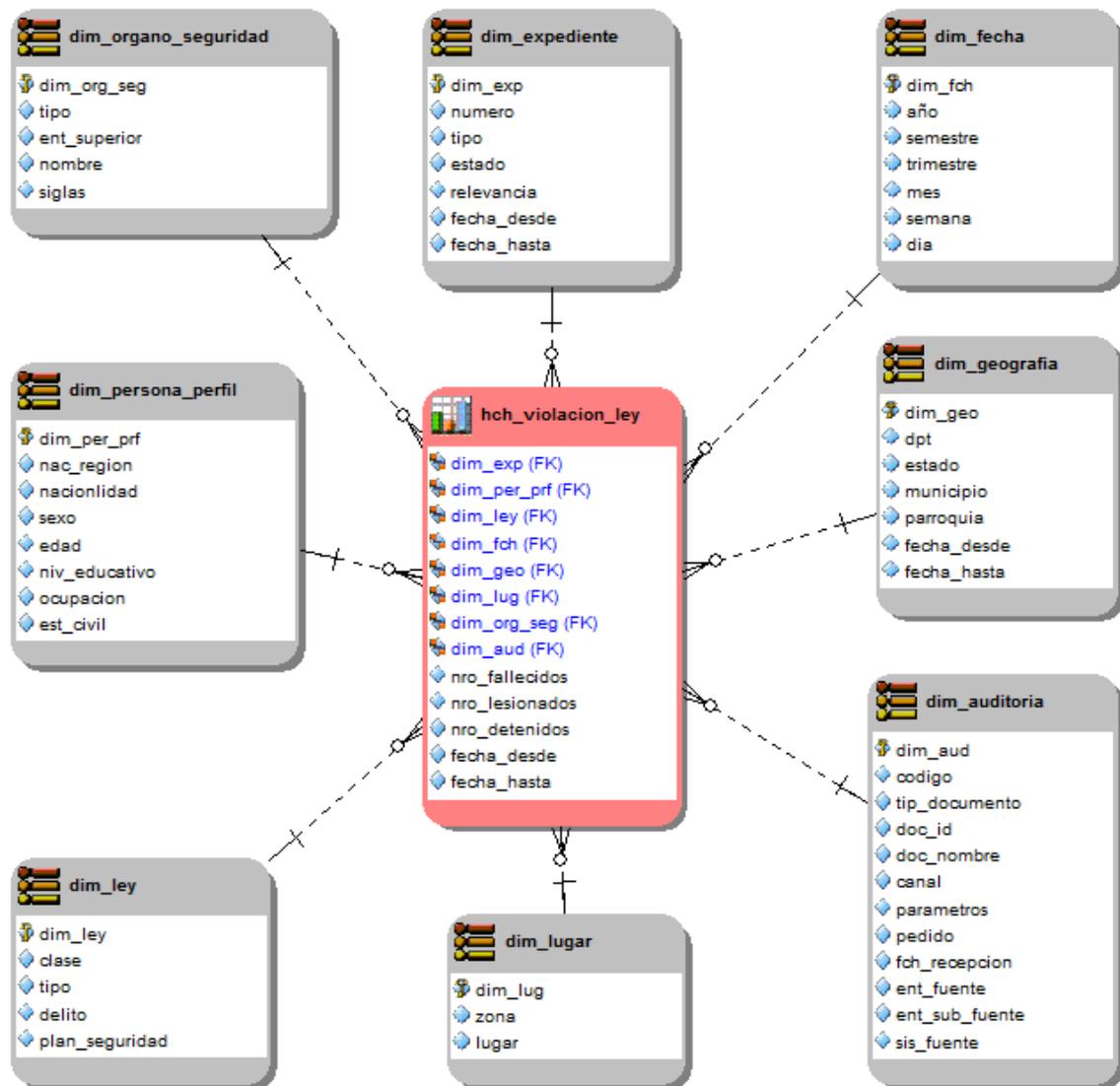


Figura 3. 11 Tabla de hechos para almacenar violaciones a la ley (delitos).

El diseño de una solución de *B/DW* en entornos donde los hechos están sujetos a cambios es una ardua tarea y se debe enfrentar con todo el dominio del negocio en cuestión, lo cual aumentara las posibilidades de concebir una robusta y escalable

arquitectura que responda eficazmente a las necesidades del cliente, aportándole un plus de competitividad y desempeño laboral.

El Centro de Tratamiento y análisis de información de seguridad ciudadana es el ente rector en cuanto a trazar estrategias en la política de seguridad ciudadana se trata, su misión es orientar las acciones de los órganos de seguridad del estado en aras de aumentar la efectividad en la lucha contra la violencia.

3.5.1 Propuesta de solución para manejo de actualizaciones en la tabla de hechos

La propuesta de diseño del mercado de datos es simple y cómoda en su implementación, se describe el modelo dimensional que se debe concebir en casos donde los hechos sufren actualizaciones.

Esta propuesta hereda de la variante propuesta por Ralph Kimball descrita en el anterior capítulo, tomando las ventajas y simplificando su implementación. En este consejo Kimball deja claro que es sumamente importante analizar las características y los requerimientos del negocio en cuestión antes de tomar una decisión en la manera de implementar la solución. El fenómeno de cambio lento en los hechos es un tema que está en actual y profunda discusión entre los especialistas de soluciones para el soporte a la toma de decisiones.

Vale aclarar que las características físicas que presenta la tabla de hechos se pueden ver como un patrón de diseño, en entornos donde los hechos presenten cambios, extensible a todas las tablas de hechos con naturaleza similar.

La propuesta de solución se presenta de 3 formas diferentes en forma de recomendaciones. La selección de una u otra variante descritas a continuación es decisión del diseñador.

3.5.1.1 Variante 1: Identificador del hecho en dimensión degenerada

A continuación se presenta **la figura 3.12** que muestra el diseño de una tabla de hechos con sus dimensiones y medidas provisto para la gestión de los cambios en los hechos.

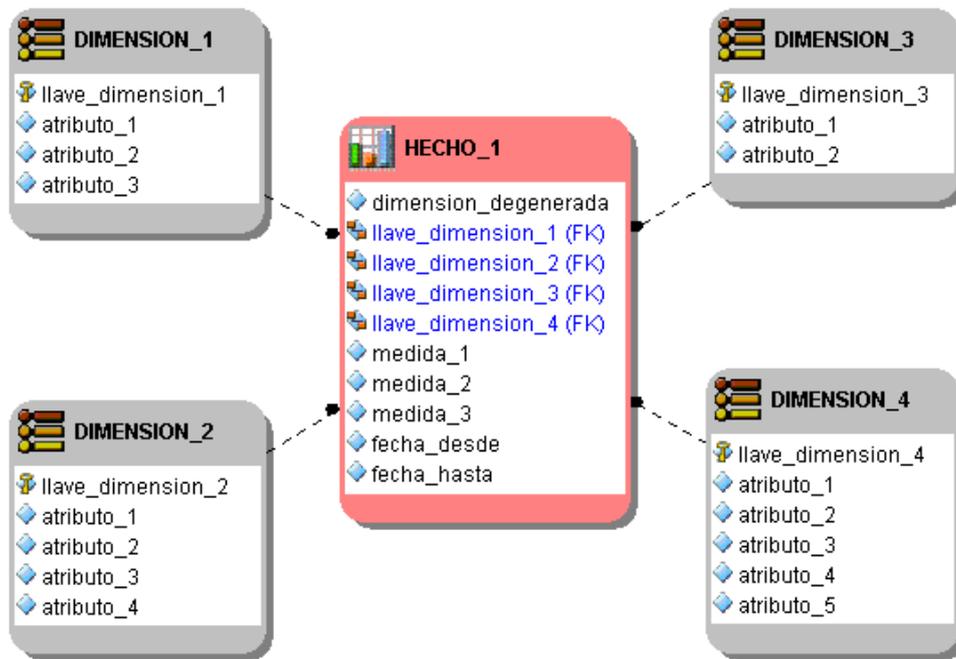


Figura 3. 12 Tabla de hechos con identificador en dimensión degenerada.

La tabla de hechos tiene un campo que se comporta como dimensión degenerada (dimensión con un único campo), y dos campos de tipo fecha con los que se enmarca el rango de tiempo en que un hecho se considera válido.

La lógica de almacenamiento para esta variante hace conjugar los campos *dimension_degenerada*, *fecha_desde* y *fecha_hasta*:

CAMPOS	VERSIÓN 1
dimension_degenerada	ABCDE-12345
llave_dimension_1 (FK)	500
llave_dimension_2 (FK)	66
llave_dimension_3 (FK)	98
llave_dimension_4 (FK)	1
medida_1	12
medida_2	10
medida_3	15
fecha_desde	01/10/2012
fecha_hasta	null

Figura 3. 13 Tabla de hechos con un registro.

La primera vez que se registra el hecho “X” (versión 1).

- ✓ El campo *dimension_degenerada* almacena el identificador del hecho.

- ✓ Los campos *medida_1*, *medida_2* y *medida_3* almacenan los datos cuantitativos.
- ✓ Los campos *llave_dimension_1*, *llave_dimension_2*, *llave_dimension_3* y *llave_dimension_4* almacenan los valores numéricos que hacen referencia a los datos cualitativos de las tablas de dimensión.
- ✓ El campo *fecha_desde* almacena la fecha de validez del hecho.
- ✓ El campo *fecha_hasta* guarda un valor nulo, indicando que esta versión del hecho es la habilitada como verdadera.

En una posible actualización de este hecho, la tabla quedaría de la siguiente manera:

La figura 3.14 muestra un juego de valores posibles para una actualización del hecho previamente registrado.

CAMPOS	VERSIÓN 1	VERSIÓN 2
<i>dimension_degenerada</i>	ABCDE-12345	ABCDE-12345
<i>llave_dimension_1</i> (FK)	500	500
<i>llave_dimension_2</i> (FK)	66	66
<i>llave_dimension_3</i> (FK)	98	56
<i>llave_dimension_4</i> (FK)	1	1
<i>medida_1</i>	12	25
<i>medida_2</i>	10	10
<i>medida_3</i>	15	15
<i>fecha_desde</i>	01/10/2012	31/12/2012
<i>fecha_hasta</i>	31/12/2012	null

Figura 3. 14 Tabla de hechos con un registro, más una actualización del mismo.

El campo *medida_1* ha cambiado su valor, igualmente lo ha hecho el campo *llave_dimensión_3*. Esto consiste en insertar el mismo registro con los nuevos valores de *medida_1* y *llave_dimensión_3*, actualizando el campo *fecha_hasta* del registro anteriormente válido e insertando este mismo valor de fecha en el campo *fecha_desde* del hecho que se convierte en la nueva verdad.

3.5.1.2 Variante 2: Identificador del hecho en dimensión conformada

A continuación se presenta la figura 3.15 que muestra el diseño de una tabla de hechos con sus dimensiones y medidas provisto para la gestión de los cambios en los hechos.

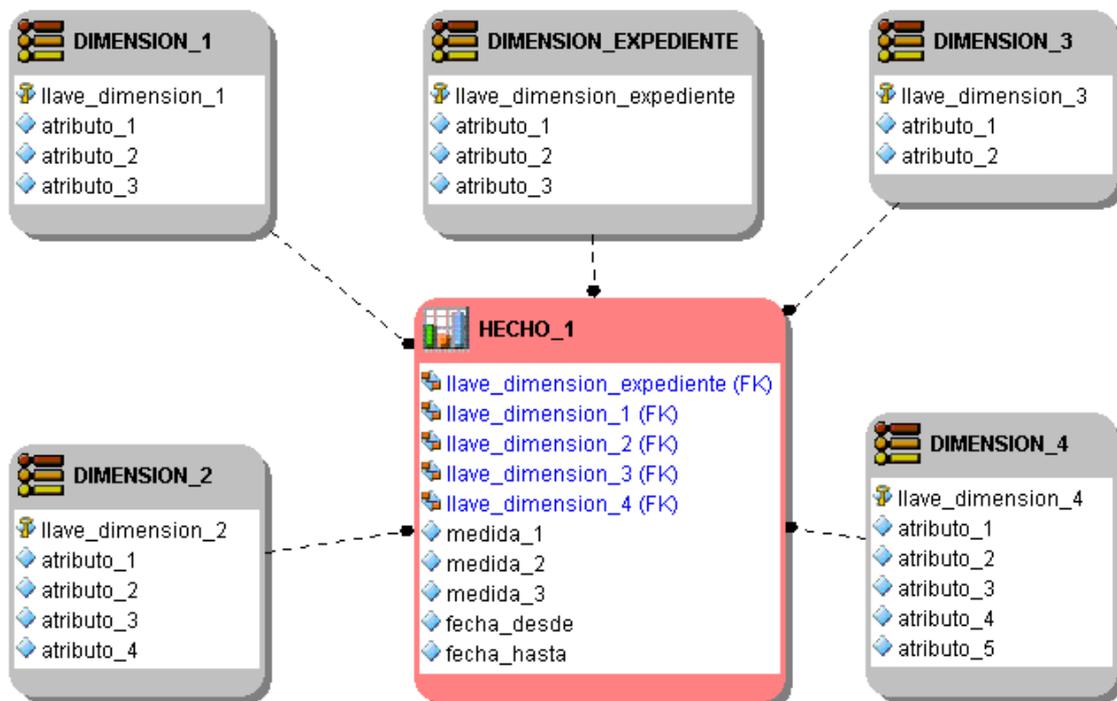


Figura 3. 15 Tabla de hechos con identificador en dimensión conformada.

La lógica de almacenamiento para esta variante hace conjugar los campos *llave_dimension_expediente*, *fecha_desde* y *fecha_hasta*:

La figura 3.16 muestra un juego de valores para la variante con identificador en una dimensión.

CAMPOS	VERSIÓN 1
<i>llave_dimension_expediente</i>	123
<i>llave_dimension_1 (FK)</i>	500
<i>llave_dimension_2 (FK)</i>	66
<i>llave_dimension_3 (FK)</i>	98
<i>llave_dimension_4 (FK)</i>	1
<i>medida_1</i>	12
<i>medida_2</i>	10
<i>medida_3</i>	15
<i>fecha_desde</i>	01/10/2012
<i>fecha_hasta</i>	null

Figura 3. 16 Tabla de hechos con un registro.

La primera vez que se registra el hecho “X” (versión 1).

- ✓ El campo *llave_dimension_expediente* almacena el valor numérico que hace referencia al identificador del hecho en la tabla *dimension_expediente*.

- ✓ Los campos *medida_1*, *medida_2* y *medida_3* almacenan los datos cuantitativos.
- ✓ Los campos *llave_dimension_1*, *llave_dimension_2*, *llave_dimension_3* y *llave_dimension_4* almacenan los valores numéricos que hacen referencia a los datos cualitativos de las tablas de dimensión.
- ✓ El campo *fecha_desde* almacena la fecha de validez del hecho.
- ✓ El campo *fecha_hasta* guarda un valor nulo, indicando que esta versión del hecho es la habilitada como verdadera.

En una posible actualización de este hecho, la tabla quedaría de la siguiente manera:

La figura 3.17 muestra un juego de valores posibles para una actualización del hecho previamente registrado.

CAMPOS	VERSIÓN 1	VERSIÓN 2
llave_dimension_expediente	123	123
llave_dimension_1 (FK)	500	500
llave_dimension_2 (FK)	66	66
llave_dimension_3 (FK)	98	56
llave_dimension_4 (FK)	1	1
medida_1	12	25
medida_2	10	10
medida_3	15	15
fecha_desde	01/10/2012	31/12/2012
fecha_hasta	31/12/2012	null

Figura 3. 17 Tabla de hechos con un registro, más una actualización del mismo.

Como se muestra en la figura anterior, el atributo identificador por el cual se busca el hecho que ha cambiado, forma parte de la *dimensión expediente* y solo es referenciado en la tabla de hechos.

El campo *medida_1* ha cambiado su valor, igualmente lo ha hecho el campo *llave_dimensión_3*. Esto consiste en insertar el mismo registro con los nuevos valores de *medida_1* y *llave_dimensión_3*, actualizando el campo *fecha_hasta* del registro anteriormente válido e insertando este mismo valor de fecha en el campo *fecha_desde* del hecho que se convierte en la nueva verdad.

3.5.1.3 Variante 3: Identificador del hecho como la composición de las llaves foráneas de las dimensiones

En esta variante se asume que los hechos no tienen un identificador explícito, más bien consiste en la combinación de las llaves foráneas de las tablas de dimensión asociadas a la tabla de hechos. En este caso la búsqueda a realizar para encontrar el registro que debe ser actualizado debe hacerse asumiendo que los campos foráneos de la tabla de hechos conforman una llave primaria.

A continuación se presenta la **figura 3.18** que muestra el diseño de una tabla de hechos con sus dimensiones y medidas provisto para la gestión de los cambios en los hechos.

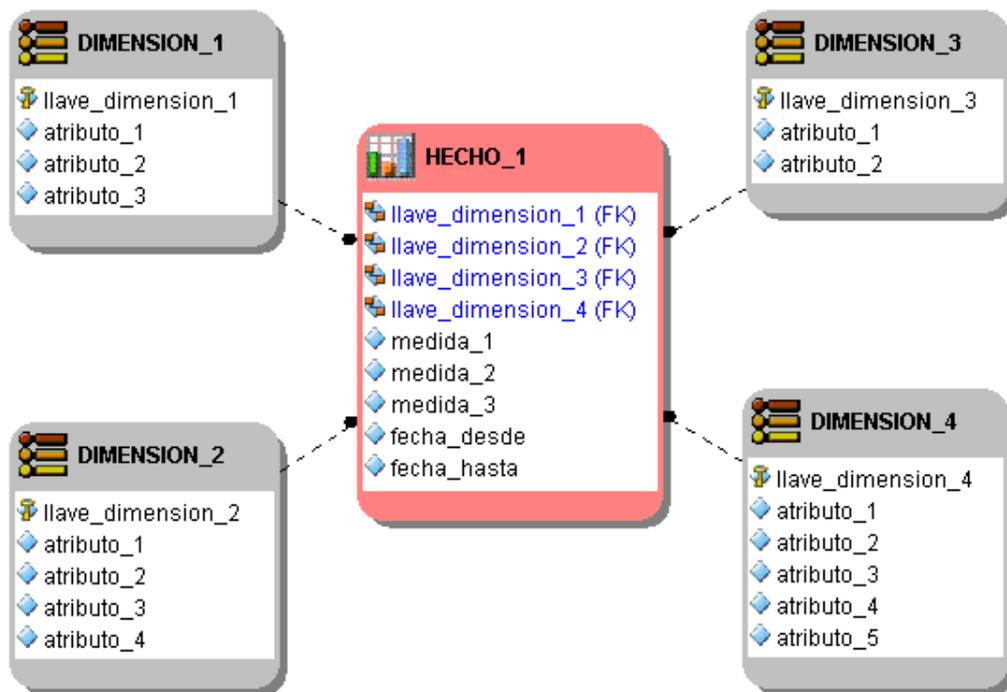


Figura 3. 18 Tabla de hechos con identificador conformado por la combinación de las llaves foráneas de las tablas de dimensión asociadas a la tabla de hechos.

La lógica de almacenamiento para esta variante hace conjugar los campos *llave_dimension_1*, *llave_dimension_2*, *llave_dimension_3*, *llave_dimesnion_4*, *fecha_desde* y *fecha_hasta*:

La **figura 3.19** muestra un juego de valores para el caso donde el identificador del hecho esté compuesto por las llaves foráneas de las dimensiones relacionadas a la tabla de hechos.

	CAMPOS	VERSIÓN 1
Identificador	llave_dimension_1 (FK)	500
	llave_dimension_2 (FK)	66
	llave_dimension_3 (FK)	98
	llave_dimension_4 (FK)	1
	medida_1	12
	medida_2	10
	medida_3	15
	fecha_desde	01/10/2012
	fecha_hasta	null

Figura 3. 19 Tabla de hechos con un registro.

La primera vez que se registra el hecho “X” (versión 1).

- ✓ Los campos *medida_1*, *medida_2* y *media_3* almacenan los datos cuantitativos.
- ✓ Los campos *llave_dimension_1*, *llave_dimension_2*, *llave_dimension_3* y *llave_dimension_4* almacenan los valores numéricos que hacen referencia a los datos cualitativos de las tablas de dimensión.
- ✓ El campo *fecha_desde* almacena la fecha de validez del hecho.
- ✓ El campo *fecha_hasta* guarda un valor nulo, indicando que esta versión del hecho es la habilitada como verdadera.

En una posible actualización de este hecho, la tabla quedaría de la siguiente manera:

La figura 3.20 muestra un juego de valores posibles para una actualización del hecho previamente registrado.

	CAMPOS	VERSIÓN 1	VERSIÓN 2
Identificador	llave_dimension_1 (FK)	500	500
	llave_dimension_2 (FK)	66	66
	llave_dimension_3 (FK)	98	98
	llave_dimension_4 (FK)	1	1
	medida_1	12	25
	medida_2	10	10
	medida_3	15	15
	fecha_desde	01/10/2012	31/12/2012
	fecha_hasta	31/12/2012	null

Figura 3. 20 Tabla de hechos con un registro, más una actualización del mismo.

Como se muestra en la figura anterior, el atributo identificador por el cual se busca el hecho que ha cambiado consiste en la composición de las 4 llaves de tablas de dimensión.

Vale aclarar que si uno de los campos foráneos cambia se asume como un nuevo hecho y no como una actualización. Los hechos que sugieren actualización solo manifiestan cambios en los valores de las medidas que lo componen.

El campo *medida_1* ha cambiado su valor. Esto consiste en insertar el mismo registro con el nuevo valor del campo *medida_1*, actualizando el campo *fecha_hasta* del registro anteriormente válido e insertando este mismo valor de fecha en el campo *fecha_desde* del hecho que se convierte en la nueva verdad..

3.5.2 Técnica para integrar los datos

En el presente epígrafe se describe el procedimiento para realizar la extracción, transformación y carga de los datos. Según los especialistas esta capa de integración de datos consume alrededor del setenta por ciento del tiempo y el esfuerzo en el ciclo de vida de una solución de *BI/DW*.

Para una mejor comprensión del proceso de integración de datos se ha ilustran flujos de trabajo creados con la herramienta *Pentaho Data Integration*. Los pasos y transformaciones a continuación presentados constituyen el eje central de las actividades de extracción, transformación y carga de los datos hacia la base de datos multidimensional. **La figura 3.21** muestra el procedimiento general para procesar los

datos en situaciones donde los hechos recopilados están propensos a sufrir una o varias actualizaciones.

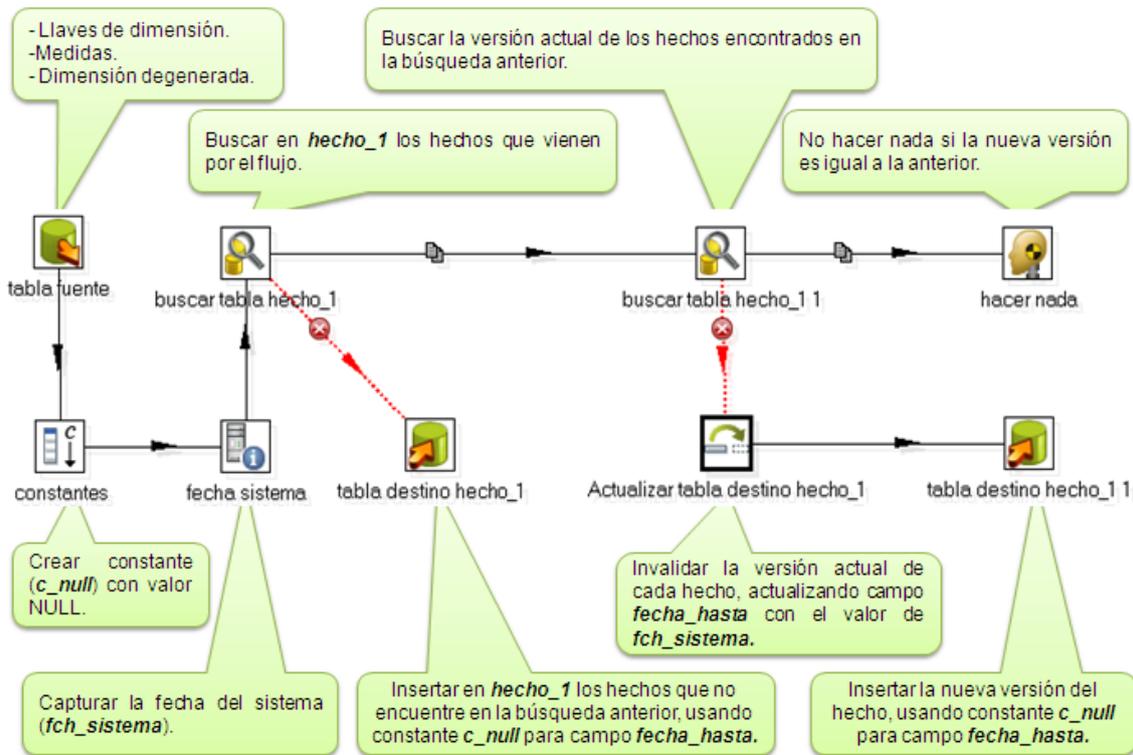


Figura 3. 21 Flujo de etl para el almacenamiento de las actualizaciones de los hechos.

El flujo funciona de manera general para cualquiera de los tres casos de modelado presentados en la sección anterior. Lo importante aquí es saber que el cambio lento tipo dos conocido sobre las dimensiones ahora se aplica a los hechos.

La imagen muestra una entrada de tabla que en este caso contiene los campos que coinciden con los de la tabla de hecho destino, seguidamente se crea un constante con valor *NULL* y se captura la fecha del sistema, valores que serán usados para los campos *fecha_desde* y *fecha_hasta* de la tabla destino.

En este punto ya se puede realizar una primera búsqueda en la tabla de hechos insertando en la misma los registros que no coincidan en cuanto al identificador del hecho (esto quiere decir que son nuevos hechos), el resto de los registros se asumen como nuevas actualizaciones y se procede a una segunda búsqueda en donde se identifican los registros que hayan cambiado los cuales pasan a ser actualizados, esta segunda búsqueda se hace por el resto de los campos no identifican el hecho.

Vale aclarar que para el caso en que el identificador del hecho es la composición de las

llaves foráneas que están referenciadas en la tabla de hechos, la segunda búsqueda estaría enfocada a buscar los valores de las medidas que cambiaron.

3.5.3 Pauta para el análisis sobre los datos

En los dos epígrafes anteriores de este capítulo se ha descrito las características técnicas del diseño del almacén de datos y los flujos de trabajo de los mecanismos de *ETL*. Esto es solo una parte importante de la propuesta, la capa de análisis complementa las dos anteriormente descritas. En el presente epígrafe se exponen las particularidades de la capa de consulta a la base de datos multidimensional en entornos donde los hechos presentan cambio lento tipo dos.

Las consultas en la capa de análisis requieren de un aspecto técnico a tener siempre en cuenta a la hora de escribirlas, como se explico en el primer epígrafe de este capítulo el diseño físico del almacén se puede modelar de tres formas diferentes en dependencia de las particularidades del negocio en cuestión.

Las consultas que se muestran a continuación están orientadas a demostrar cómo obtener los hechos validos, o sea, la última versión de la verdad de cada hecho.

En el primer caso donde la tabla de hechos contiene como dimensión degenerada el código identificador de negocio del hecho, la consulta en la capa de *BI* debe presentar el siguiente filtro:

```
select (Campos seleccionados)  
        from hecho_1  
        ... (Más cualquier otra tabla adicional)  
        where fecha_hasta = null  
        ... (Más cualquier otro filtro adicional)
```

Vale aclarar que el campo *id_negocio* almacena el identificador de negocio de los hechos que componen el cada bloque a integrar al almacén de datos.

En el segundo caso donde el identificador de negocio está contenido en la dimensión expediente y el campo *llave_dimensión_expediente* actúa como llave foránea en la tabla de hechos, la consulta anterior varía un poco, quedando de esta manera:

```

        select (Campos seleccionados)
        from hecho_1 h, dimension_expediente e
        ... (Más cualquier otra tabla adicional)
    on h.llave_dimension_expediente = e.llave_dimension
        where fecha_hasta = null
        ... (Más cualquier otro filtro adicional)

```

En el tercer caso donde el identificador de negocio para los hechos está compuesto por las llaves foráneas de las dimensiones, la consulta para obtener el mismo resultado de las dos anteriores sería de la siguiente manera:

```

        select (Campos seleccionados)
        from hecho_1
        ... (Más cualquier otra tabla adicional)
        where fecha_hasta = null
        ... (Más cualquier otro filtro adicional)

```

En este último caso se puede apreciar que la consulta no difiere de la del primer caso, sin embargo la estructura del almacén es diferente, lo cual implica cambios en la capa de integración.

Las consultas anteriormente presentadas están orientadas a obtener siempre la última versión de cada hecho, como se puede apreciar siempre es necesario la inclusión del filtro *fecha_hasta = null*, esto hace que se descarten todos los registros obsoletos, dejando para el análisis solo los registros validos y sobre este bloque aplicar los demás filtros que se requieran como las fechas y geografías.

3.6 Conclusiones Parciales

Concebir el diseño físico aplicando la técnica de cambio lento tipo dos en las tablas de hechos constituye la tarea primaria ya que es el almacén de datos el que debe estar preparado estructuralmente para asumir este fenómeno.

El equipo de integración de datos tiene la responsabilidad de extraer, transformar si es necesario y cargar los datos desde los sistemas fuentes hacia el almacén de datos, teniendo en cuenta la posibilidad de que un hecho cambie su versión de la verdad.

Los especialistas en análisis y generación de informes siempre son los primeros en evaluar la calidad de la solución, manteniendo especial cuidado a la hora de realizar las consultas al almacén de datos.

Los resultados obtenidos son satisfactorios. Hoy el Observatorio Venezolano de Seguridad (anteriormente CTAISC) cuenta con la herramienta necesaria para desempeñar efectivamente su rol social, las decisiones tomadas por los máximos dirigentes del gobierno venezolano tienen su base de apoyo en el sistema de inteligencia de negocios que se implantó en el CTAISC.

CONCLUSIONES

- ✓ Se describió el caso de estudio en el cual los hechos delictivos están sujetos a cambios.
- ✓ Se profundizó en el estudio de los principales conceptos, herramientas y buenas prácticas para darle solución al problema planteado.
- ✓ Se presentó la propuesta de solución, validando la misma con su implantación en el Observatorio Venezolano de Seguridad de la República Bolivariana de Venezuela.

RECOMENDACIONES

- ✓ Aplicar técnicas de inteligencia artificial, específicamente en el campo de la incertidumbre, al modelo dimensional propuesto.
- ✓ Profundizar en el estudio de las técnicas de integración y análisis de los datos en entornos donde los hechos estén sujetos a sufrir cambios.
- ✓ Profundizar en el estudio de soluciones al cambio lento de los hechos en la nueva era *BIG DATA*.

REFERENCIAS BIBLIOGRÁFICAS

- Adamson, Chris. 2011.** <http://blog.oaktonsoftware.com>. *Slowly Changing Facts?* [En línea] agosto de 2011. [Citado el: 20 de diciembre de 2012.]
<http://blog.oaktonsoftware.com/2011/08/slowly-changing-facts.html>.
- Alter, Steven. 1980.** *Decision support systems: current practice and continuing challenges*. 1980.
- Ariño, Miguel A. 2012.** <http://miguelarino.com>. *Toma de decisiones e incertidumbre*. [En línea] 19 de enero de 2012. [Citado el: 18 de diciembre de 2012.]
<http://miguelarino.com/2012/01/19/toma-de-decisiones-e-incertidumbre/>.
- Bello, Rafael. 2010.** <http://rcci.uci.cu>. *Los conjuntos aproximados en el contexto de la Soft Computing*. [En línea] 2010. [Citado el: 17 de diciembre de 2012.]
http://rcci.uci.cu/index.php/rcci/article/download/103/97&sa=U&ei=7j_PUKeNloS-qgH9moGAAw&ved=0CBQQFjAA&usg=AFQjCNHs5mdGfl4hogsyrsjSY84BsovsXQ.
- CARBONELL, JUAN JESÚS TORRES. 2009.** <http://www.revista-ays.com>. *Incertidumbre y toma de decisiones*. [En línea] 2009. [Citado el: 18 de 12 de 2012.] <http://www.revista-ays.com/DocsNum34/FirmasInvitadas/Carbonell.pdf>.
- Chaudhuri, Surajit, Dayal, Umeshwar y Narasayya, Vivek. 2011.** *An overview of business intelligence technology*. 2011. 0001-0782.
- Curto, José. 2008.** <http://informationmanagement.wordpress.com>. *Diseño de un datawarehouse: tabla de hecho*. [En línea] julio de 2008. [Citado el: 20 de diciembre de 2012.]
<http://informationmanagement.wordpress.com/2008/07/12/disenio-de-un-data-warehouse-tabla-de-hecho/>.
- Fernández, Carlos Molina. 2005.** <http://hera.ugr.es>. *Imprecisión e Incertidumbre en el Modelo Multidimensional: Aplicación a la Minería de Datos*. [En línea] julio de 2005. [Citado el: 17 de 12 de 2012.] <http://hera.ugr.es/tesisugr/15518772.pdf>.
- Herrera, Miguel. 2012.** <http://www.auladeeconomia.com>. *Toma de decisiones gerenciales*. [En línea] 16 de julio de 2012. [Citado el: 17 de 12 de 2012.]
<http://www.auladeeconomia.com/AG02b-TOMA%2520DE%2520DECISIONES.ppt&sa=U&ei=NkfpUL6eNZS42gX08IDQBg&ved=0CDgQFjAH&usg=AFQjCNHru2YjOLXB3j9rmZ6weqgFyku20Q>.
- Ibarra, Maria de los Angeles. 2005.** <http://exa.unne.edu.ar>. *Proceso Analítico en Línea*. [En línea] 2005. [Citado el: 17 de 12 de 2012.]
<http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAP.pdf>.
- Inmon, William H. 2005.** *Building the data warehouse*. 2005.
- Kimball, Ralph. 2005.** <http://www.kimballgroup.com>. *DesignTip #74 Compliance-Enabled Data Warehouse*. [En línea] diciembre de 2005. [Citado el: 20 de diciembre de 2012.]
<http://www.kimballgroup.com/2005/12/16/design-tip-74-compliance-enabled-data-warehouses/>.
- . 1997. <http://www.kimballgroup.com>. *A Dimensional Modeling Manifesto*. [En línea] agosto de 1997. [Citado el: 20 de diciembre de 2012.]
<http://www.kimballgroup.com/2003/01/01/fact-tables-and-dimension-tables/>.
- Kimball, Ralph y Joe, Caserta. 2004.** *The Data Warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. 2004. 0-764-57923-1.

- Kimball, Ralph y Ross, Margy. 2002.** *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling.* 2002.
- Kimball, Ralph, y otros. 1998.** *The Data Warehouse Life Cycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouse.* 1998. 0471255475.
- Maroto, Alicia, y otros.** <http://www.quimica.urv.es>. *INCERTIDUMBRE Y PRECISIÓN.* [En línea] [Citado el: 17 de 12 de 2012.] <http://www.quimica.urv.es/quimio/general/incert.pdf>.
- Mundy, Joe. 2012.** Design Tip #145 Time Stamping Accumulating Snapshot Fact Tables. [//www.kimballgroup.com](http://www.kimballgroup.com). [En línea] mayo de 2012. [Citado el: 20 de diciembre de 2012.] [//www.kimballgroup.com/2012/05/01/design-tip-145-time-stamping-accumulating-snapshot-fact-tables/](http://www.kimballgroup.com/2012/05/01/design-tip-145-time-stamping-accumulating-snapshot-fact-tables/).
- PALACIOS, Dr. FREDDY WILLIAM CASTILLO. 2009.** <http://blog.pucp.edu.pe>. *TOMA DE DECISIONES EN CONDICIONES DE CERTEZA, INCERTIDUMBRE Y RIESGO.* [En línea] 22 de noviembre de 2009. [Citado el: 17 de diciembre de 2012.] <http://blog.pucp.edu.pe/item/79396/toma-de-decisiones-en-condiciones-de-certeza-incertidumbre-y-riesgo>.
- Palacios, Freddy Castillo. 2009.** <http://manuelgross.bligoo.com>. *Toma de Decisiones en condiciones de Certeza, Incertidumbre y Riesgo.* [En línea] 23 de noviembre de 2009. [Citado el: 18 de 12 de 2012.] <http://manuelgross.bligoo.com/content/view/666731/Toma-de-Decisiones-en-condiciones-de-Certeza-Incertidumbre-y-Riesgo.html>.
- Pérez, Santiago. 2006.** <http://www.edutecne.utn.edu.ar>. *APOYO PARA LA TOMA DE DECISIONES.* [En línea] 2006. [Citado el: 17 de diciembre de 2012.] <http://www.edutecne.utn.edu.ar/sistemas-informacion/Data%20Mining-DataWarehouse.pdf>.
- PONNIAH, PAULRAJ. 2001.** *DATA WAREHOUSING FUNDAMENTALS: A Comprehensive Guide for IT Professionals.* 2001. 0-471-22162-7.
Propuesta de diseño multidimensional donde los hechos presentan diferentes versiones de la verdad. **Pérez Anchia, Enrique Alberto. 2013.** La Habana : s.n., 2013. 978-959-7213-02-4.
- Russell, Stuart J. y Norvig, Peter. 1995.** *Artificial Intelligence: A Modern Approach.* [En línea] 1995. [Citado el: 17 de 12 de 2012.]
- T. Moss, Larissa y Atre, Shaku. 2003.** *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications.* s.l. : Addison Wesley, 2003. 0-201-78420-3.
- Vanegas, Carlos Mora. 2012.** <http://educacion-dinamica.lacoctelera.net>. [En línea] 4 de febrero de 2012. [Citado el: 17 de diciembre de 2012.] <http://educacion-dinamica.lacoctelera.net/post/2012/02/04/lo-importante-saber-tomar-decisiones>.