

TD-0385-07

005.12

HER

TD-0385-07

**República de Cuba**



**Universidad de las Ciencias Informáticas**

**Facultad 2**

**Concepción de un sistema para el tratamiento de  
información de seguridad ciudadana**

**Trabajo de Diploma**

**Presentado para optar por el título de  
Ingeniero en Ciencias Informáticas**

**Autores:**

**Asnioby Hernández López**

**Yanet Peña Vazquez**

**Tutor: Dra. Ana María García Pérez.**

Junio del 2007

## **Resumen.**

El presente trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas describe los resultados de un estudio sobre las herramientas informáticas que corresponden al estado del arte en el tratamiento de la información sobre seguridad ciudadana.

Aborda las características esenciales de los problemas de inseguridad en la República Bolivariana de Venezuela y desarrolla la propuesta de una arquitectura Data Warehouse, para implementar un sistema de información que brinde apoyo a la toma de decisiones para un Centro de Tratamiento y Análisis de la Información de Seguridad Ciudadana (CTAISC) en este hermano país.

## Índice

Introducción.....	1
Capítulo 1: Estado del arte en el tratamiento de información sobre seguridad ciudadana.....	6
1.1 Seguridad Ciudadana.....	6
1.2 Seguridad Ciudadana en Venezuela.....	9
1.3 Antecedentes.....	12
1.4 Necesidad de un Centro de Tratamiento y Análisis de Información de Seguridad Ciudadana....	13
1.5 La información como herramienta eficaz en la seguridad ciudadana.....	15
1.6 Integración, almacenamiento y distribución de información.....	17
Conclusiones del Capítulo.....	22
Capítulo 2: Conceptualización del Data Warehouse del CTAISC.....	23
2.1 Fases para el enfrentamiento a problemas de seguridad ciudadana.....	23
2.2 Modelo para la conceptualización del Data Warehouse.....	24
2.3 Metodología a utilizar.....	26
2.4 Modelo multidimensional.....	27
2.4.1 Cubo, Dimensión y Medida.....	27
2.4.2 Tablas de Dimensiones.....	28
2.4.3 Definiendo dimensiones.....	28
2.4.4 Tablas de Hechos.....	33
2.4.5 Definiendo Tablas de Hechos.....	34
2.4.6 Cubo Multidimensional.....	35
2.4.7 Definiendo Cubos.....	35
2.5 Diseño.....	36
2.6 Construcción.....	39
2.6.1 Extracción.....	39
2.6.2 Integración.....	41
2.7 Análisis.....	42
2.7.1 Operaciones.....	43
2.8 Mantenimiento.....	44

2.9 Validación del modelo.....	44
2.10 Conclusiones del capítulo.....	45
Capítulo 3. Selección de las herramientas informáticas.....	46
Introducción.....	46
3.1 Elección.....	46
3.2 Oracle Database Enterprise Edition 10g.....	47
3.2.1 Manejo de tablas agregadas y resumizadas.....	48
3.2.2 Mecanismos de actualización y control de cambios.....	49
3.2.3 Integridad de datos.....	50
3.2.4 Funciones analíticas.....	51
3.2.5 Administración de datos.....	52
3.2.6 Replicación de Datos.....	52
3.2.7 Carga y población de datos.....	54
3.2.8 Integración de los datos en tiempo real.....	57
3.2.9 Índices.....	57
3.3 Oracle Warehouse Builder.....	57
3.3.1 OWB versus procesos manuales de programación.....	59
3.3.2 Herramientas de Limpieza de Datos.....	63
3.3.3 Administración.....	64
3.3.4 Plataforma.....	66
3.3.5 Diseño y modelado.....	67
3.3.6 Rendimiento.....	67
3.4 Oracle OLAP.....	67
3.4.1 Funcionalidad de Oracle OLAP.....	69
3.4.2 Proceso de Base de Datos Integrado.....	69
3.4.3 Almacenamiento.....	69
3.5 Oracle Spatial.....	70
3.5.1 Geocoder.....	71
3.5.2 Operadores espaciales.....	71
3.5.3 Consultas espaciales paralelas.....	72

3.6 Real Application Cluster (RAC) .....	72
3.6.1 Costos de hardware .....	73
3.6.2 Flexibilidad .....	73
3.6.3 Disponibilidad .....	74
3.7 Oracle Discoverer .....	74
3.7.1 Arquitectura del Oracle Discoverer .....	74
3.7.2 Publicación de Reportes .....	75
3.7.3 Rendimiento .....	75
3.7.4 Integración con la Base de Datos .....	76
Conclusiones del Capítulo .....	76
Conclusiones .....	77
Recomendaciones .....	78
Referencias Bibliográficas .....	79
Bibliografía .....	83
Glosario de Términos .....	84
Anexo1. Licencias .....	86

## Introducción.

**L**a seguridad ciudadana es uno de los principales problemas que se presentan en el mundo globalizado, siendo más marcada en las grandes ciudades. Los Estados deben garantizar la seguridad como un derecho ciudadano. El único respaldo que puede sentir un ciudadano cuando se siente amenazado en su instinto de seguridad, es el de sentirse apoyado por el Estado y que este haga sentir su presencia por la vía de la prevención y de la represión, pero atendiendo al principio de corresponsabilidad.

Una forma de contribuir con tal propósito es implementar políticas y estrategias de seguridad ciudadana que busque la activación de los actores sociales de las comunidades. Así se podrá satisfacer las necesidades de los ciudadanos en cuanto a la preservación de su integridad física, sus propiedades y otros bienes materiales, situaciones de emergencia y sentimientos de seguridad. De otra manera el ciudadano optará por acciones individuales para garantizarse *per se* su propia seguridad, sea adquiriendo armas de fuego, protegiendo sus propiedades sea por medios mecánicos o electrónicos, contratación de seguridad personal o familiar, limitación de movimiento, entre otras.

La ONU, (1999 9(a), p.7)<sup>1</sup> plantea los factores relacionados con la delincuencia, la violencia y la inseguridad:

- Pobreza y desempleo derivados de la exclusión social, especialmente en el caso de los jóvenes;
- Familias disfuncionales donde los padres tienen actitudes indiferentes o contradictorias, o en las que existen situaciones de violencia o de conflicto entre los padres;
- Una sociedad que acepta o promueve una cultura de la violencia;
- Discriminación y exclusión basadas en el género, la raza u otros motivos;
- Degradación de los entornos urbanos y de los vínculos sociales;

---

<sup>1</sup> ONU. 1999 (a). **Participación de la comunidad en la prevención de la delincuencia.** Documento de antecedentes para el curso práctico sobre participación de la comunidad en la prevención de la delincuencia. Documento de trabajo para el Décimo Congreso de las Naciones Unidas sobre la Prevención del delito y Tratamiento del delincuente, realizado en Viena del 10 al 17 de abril.

[www.uncjin.org/Documents/congr10/11s.pdf](http://www.uncjin.org/Documents/congr10/11s.pdf).

- Vigilancia inadecuada de los bienes y lugares públicos;
- Presencia de factores facilitadores (como armas de fuego, alcohol y drogas).

En la República Bolivariana de Venezuela, se trabaja por el Estado por disminuir los índices delictivos, aumentar el sentimiento de seguridad ciudadana a partir de lo establecido en la Constitución de la República Bolivariana de Venezuela de 1999, específicamente el tema es referido en los Artículos 55 y 332. De esos se desprende que toda persona tiene derecho a la protección por parte del Estado a través de los órganos de seguridad ciudadana regulados por ley, frente a situaciones que constituyan amenaza, vulnerabilidad o riesgo para la integridad física de las personas, sus propiedades, el disfrute de sus derechos y el cumplimiento de sus deberes. Los órganos de seguridad ciudadana son de carácter civil y respetarán la dignidad y los derechos humanos, sin discriminación alguna, siendo una función de los órganos de seguridad ciudadana constituyen una competencia concurrente con los Estados y Municipios en los términos establecidos. [Landáez, 2002]

La Encuesta de Victimización realizada en el año 2006, reportó en 10,000 encuestados que la cifra negra de criminalidad es del 66%, concentrándose en los hechos convencionales contra la propiedad, entre los cuales, robo y hurto representan el 67,8% de las ocurrencias victímales, con tasas muy superiores a las recogidas por las Estadísticas Policiales.

Esto se debe principalmente a que las personas no poseen una actitud de denunciabilidad, representando este grupo el 68.5% de la población. En estos hechos se emplean frecuentemente algún tipo de armas. Generalmente ocurre en las propias áreas de residencia de las víctimas y en horas de la tarde o en las primeras horas de la noche. [Instituto Nacional de Estadística, 2006]

Además de la Encuesta de Victimización, existen otras fuentes de información sobre Seguridad Ciudadana, algunas de ellas se soportan en sistemas de almacenamiento de datos, entre otras las correspondientes a CICPC, 171, Bomberos, Tránsito Terrestre, Sistema de Antecedentes Penales, ONIDEX. Esos sistemas almacenan determinadas informaciones, sin embargo, la toma de decisiones en la prevención y combate del delito, el enfrentamiento a las situaciones de emergencias, el apoyo a comunidades en la búsqueda de su seguridad y, en general, para lograr la Seguridad Ciudadana se requiere de la integración de la información existente. Luego, se hace necesario relacionar las

informaciones contenidas en las distintas fuentes de información empleando herramientas informáticas que propicien su integración, transformación y almacenamiento.

Después de analizar el problema en su entorno social, surge el siguiente *problema científico*: Carencia de *información integrada* que facilite el tratamiento y análisis de información de seguridad ciudadana, que apoye la toma de decisiones en la formulación de medidas, así como en el seguimiento y control de éstas, y en caso de situaciones de emergencias, desde una perspectiva proactiva y de responsabilidad ciudadana.

Esta investigación tiene como *objeto de estudio* el tratamiento y análisis de la información de Seguridad Ciudadana con vistas a poder tomar medidas de prevención y combate a los tipos de delitos identificados en la República Bolivariana de Venezuela.

El *objetivo de la investigación* es conceptualizar una solución informática que facilite el tratamiento y análisis de información de seguridad ciudadana, que apoye la toma de decisiones en la formulación de medidas, así como en el seguimiento y control de éstas, y en caso de situaciones de emergencias, desde una perspectiva proactiva y de responsabilidad ciudadana.

Dentro de los objetivos específicos que se han identificado se encuentran:

- 1 Identificar el estado del arte en el tratamiento de información sobre seguridad ciudadana en el mundo.
- 2 Conceptualizar las necesidades de información para la formulación de estrategias y referirlas a fuentes de datos de los Órganos de Seguridad Ciudadana en Venezuela.
- 3 Proponer un modelo y las herramientas para su implementación informática, que resuelva el problema científico, en particular el problema de la seguridad ciudadana en la Republica Bolivariana de Venezuela.

Al tener todos los objetivos específicos identificados surge la *hipótesis* de la investigación: Si se estructura la arquitectura de la solución informática y las herramientas de implementación de esta solución, entonces se facilitará el desarrollo y despliegue de un sistema informático que gestione la información para apoyar la toma de decisiones en la formulación de estrategias de intervención social para la seguridad ciudadana.

Para culminar una investigación con una organización satisfactoria es necesario trazarse *tareas de investigación* que guíen todo el desarrollo de la misma. Entre las que se siguieron en la construcción del presente documento se encuentran las siguientes:

1. Investigar las características actuales de los sistemas de almacenamiento de datos sobre seguridad ciudadana en Venezuela.
2. Evaluar la calidad de la información obtenida de las diferentes fuentes analizadas.
3. Estudiar la relación entre Almacenes de Datos y Procesamiento Analítico en línea con los objetivos del trabajo.
4. Comparar las tecnologías seleccionadas con otras existentes en la actualidad.
5. Evaluar la viabilidad de la solución propuesta.

Este trabajo de diploma tiene como fundamental el **aporte metodológico** ya que se justifica el uso de la tecnología de Data Warehouse para la construcción de sistemas para el tratamiento de información de seguridad ciudadana, cuestión que es novedosa en el mundo en el tratamiento de este tipo de información.

Para finalizar se describe la organización que tiene este documento, que está estructurado en:

- Un primer capítulo que describe el estado del arte en el mundo y en Venezuela sobre tratamientos de información de seguridad ciudadana, en el que se justifica el empleo de las tecnologías de almacenes de datos para resolver el problema científico,
- Un segundo capítulo que muestra el diseño del modelo de un almacén de datos para el tratamiento de dicha información y
- Un tercer capítulo que describe cómo hemos validado nuestra propuesta con los clientes de la solución informática cuya estructura se propone en este trabajo.

También contiene las conclusiones donde se realiza el resumen acerca del cumplimiento de los objetivos de esta investigación y se proponen recomendaciones para la continuidad de la misma.

Se incluye un glosario de términos definido para el sistema, en el cual se explican aquellos términos que se utilizan en el documento y son propios del objeto de estudio, las referencias bibliográficas y la

bibliografía. Por último se encuentra a modo de anexo, el conjunto de licencias de software que se deben adquirir para implantar la solución en la arquitectura de hardware diseñada para el proyecto.

## Capítulo 1: Estado del arte en el tratamiento de información sobre seguridad ciudadana.

**E**n este capítulo se analizan aspectos teóricos, que para la concepción de todo el proyecto fue necesario investigar y son importantes para una correcta interpretación del contenido que se analiza en el desarrollo de la investigación.

Se describen elementos en cuanto al estado del arte en el mundo y en Venezuela sobre tratamientos de información de seguridad ciudadana, se abordan aspectos relacionados con el uso de las nuevas tecnologías de la información asociadas a la integración, transformación y análisis de grandes volúmenes de información y por último se justifica el empleo de las tecnologías de almacenes de datos para resolver el problema científico.

### 1.1 Seguridad Ciudadana.

La *seguridad* es un concepto que solamente asume significado real cuando se preguntan algunas cuestiones adicionales sobre actores, grupos, organizaciones, sociedades o naciones. O sea:

- ¿seguridad para quién?
- ¿seguridad para qué? y,
- ¿seguridad bajo qué circunstancias o condiciones?

y puede ser definida como un *proceso social* con el propósito de alcanzar una relativa seguridad social, política y económica.[Clements, 1990]

Entendido como un proceso social supone una mirada integral, ya que no debe analizarse como una simple ocurrencia de ciertos hechos ilícitos o delitos, por lo que es necesario abordar este problema con políticas orientadas a los jóvenes, a las familias, a la comunidad, a la prevención del consumo de drogas, además de favorecer la planificación y diseño de espacios públicos para hacer el entorno más seguro. [Valdivia, 2002]

Debe entenderse como Seguridad Ciudadana el grado de respeto que se otorga al conjunto de derechos de los ciudadanos, no solo por parte del Estado, sino también de parte de las personas e instituciones públicas y privadas, que deben garantizar el bienestar de todos los componentes de las estructuras a las cuales están vinculadas, a las cuales prestan servicios o de las que depende su seguridad. [Ríos, 2006]

Se sostiene que seguridad ciudadana es la facultad que tiene toda persona, natural o jurídica, a desenvolverse cotidianamente libre de amenazas a su vida, libertad, integridad física, psíquica y cultural, lo mismo que al goce de sus bienes. Es un derecho humano, consignado en constituciones, leyes, Pacto Internacional de Derechos Civiles y Políticos de Naciones Unidas, y en el caso de América Latina en la Declaración de la OEA en Montrouis. [Martínez, 2002]

La Seguridad Ciudadana debe analizarse desde dos perspectivas fundamentales: los hechos y el sentimiento de seguridad. Ambos están unidos, pero su relación no siempre es directamente proporcional. De los primeros -las faltas, delitos y accidentalidad- se puede ser responsables, víctimas, testigos, aparecen en los medios de comunicación, en las estadísticas; mientras el sentimiento de seguridad es un producto cultural, donde interviene el pasado, el ahora y el futuro de cada persona y, dependiendo de los resultados de esa mezcla, cada cual percibe el grado de seguridad de su vida.

El universo de la seguridad ciudadana comprende todo aquello que la amenaza como lo que la protege. Comprende lo conectado con la violencia, criminalidad nacional e internacional y, con la accidentalidad vial; igualmente se puede afirmar que cubre tanto los delitos tradicionales como los nuevos delitos, producto de la globalización y el desarrollo tecnológico mundial. Por otro lado, también abarca el quehacer de las diversas instituciones estatales y de la sociedad civil relacionada con su promoción y protección.

Muchos países de todo el mundo se han percatado de la demanda creciente de servicios de atención a emergencias y hoy dan sus primeros pasos en la creación de herramientas que permitan contrarrestar estas demandas con información. Pueden mencionarse algunos proyectos exitosos como por ejemplo:

Sistema Unificado de Información sobre Violencia y Delincuencia (*SUIVD*) en Bogotá, Colombia. Este sistema integra información proveniente del Instituto Nacional de Medicina Legal, Policía Metropolitana de Bogotá y Secretarías de Gobierno, Salud y Tránsito. Su principal objetivo consiste en analizar el comportamiento de los delitos de interés combinando información sobre frecuencia, perfil sobre la víctima,

el comportamiento del delito en el tiempo y lugar de ocurrencia. El SUIVD es una herramienta para la toma de decisiones e investigaciones.

En la ciudad de Londres, Inglaterra existe un centro de comando y control en el área de planificación contra incendios y emergencias, que funcionan con un sistema *ProCAD* con la capacidad de evaluar y manejar las situaciones de emergencia permitiendo tomar decisiones rápidamente. El centro cuenta además con un sistema de información geográfica, y tiene entrada a varias bases de datos del Departamento de Bomberos que están directamente conectados a varios órganos relacionados con la atención a emergencias.

En Madrid, España se está desarrollando un modelo de seguridad ciudadana orientado a encaminar las actuaciones de carácter preventivo y a aumentar el sentimiento de seguridad de los ciudadanos. El Centro Integrado de Seguridad y Emergencias (CISEM) se constituye como un centro de comando y control del Ayuntamiento de Madrid, que atenderá todas las demandas que tienen como destinatarios los servicios municipales de emergencias, y que permitirá el tratamiento único y centralizado de las solicitudes.

Buenos Aires, Argentina cuentan con un “*Sistema de Recolección, Procesamiento y Análisis de la información Delictiva*”, soportado en herramientas de Análisis Criminal y Sistemas de Información Geográfica (SIG). Este sistema permite recolectar información referida a los incidentes y hechos delictivos cometidos en jurisdicción de las Comisarías, Destacamentos y otras dependencias policiales y su transmisión a un *Centro de Procesamiento y Análisis de la Información Delictiva dependiente de cada Jefatura Departamental*<sup>2</sup> donde es procesada en una base de datos y en un Sistema de Información Geográfica (SIG). [Montenegro, 2003].

En New York, EUA la policía se auxilia de un sistema llamado *Compstat*, a través del cual ubican en un mapa electrónico todos los crímenes que se comenten en la ciudad, desde violaciones, robos, homicidios hasta el número de personas que cometen indisciplinas en vía pública. Esta herramienta informática permite identificar las áreas en las que se comete cada tipo de crimen, señalar los patrones en que se desarrollan y se distribuyen las fuerzas policiales. [Aizen, 1997]

---

<sup>2</sup> Cada Jefatura comprenden esencialmente un conjunto de partidos que se corresponden con los Departamentos Judiciales. **Dieciocho Jefaturas Departamentales cubren los 134 partidos de la Provincia de Buenos Aires.** De acuerdo a un criterio general, **7 de éstas 18 corresponden al conurbano bonaerense.**

## **1.2 Seguridad Ciudadana en Venezuela.**

En la República de Venezuela la seguridad ciudadana se ha convertido, luego de la aprobación de la Constitución de 1999, en una cuestión de políticas públicas y de corresponsabilidad entre el Estado y la Sociedad. El art. 55 de la Constitución [Asamblea Nacional Constituyente, 2000] establece como derecho la protección frente a situaciones que constituyan amenaza, vulnerabilidad o riesgo para la integridad física, la propiedad, el disfrute de los derechos y el cumplimiento de los deberes, agregando que la participación ciudadana será regulada por la ley. El art. 332 confiere relevancia especial a los órganos de policía en materia de seguridad ciudadana, estableciendo una competencia concurrente entre el ejecutivo nacional, los estados y los municipios.

Si bien el concepto de seguridad, dentro del marco constitucional, rebasa la cuestión delictiva, esta última incidencia representa un aspecto importante de la seguridad y un desafío para la sociedad venezolana actual, tanto en términos de costos de vidas humanas, lesividad económica y disminución de la calidad de vida, como en términos de erosión de la confianza ciudadana en las instituciones del Estado, cuando no hay respuestas efectivas y adecuadas para enfrentarla. La percepción de desbordamiento de la criminalidad genera muchas veces alternativas autoritarias que, lejos de mejorar la situación, promueven la inseguridad, la injusticia y el miedo entre la población.

La apreciación general de la sociedad y el gobierno nacional tiende a coincidir en las graves limitaciones de la policía en el país para garantizar la seguridad ciudadana, el cumplimiento de la ley y el libre ejercicio de los derechos y libertades fundamentales. En ese sentido, se ha planteado una reforma que, tomando en consideración las normas constitucionales y los instrumentos internacionales suscritos y ratificados por la República, faciliten la función del Estado de contribuir significativamente, a través de la policía, a la seguridad ciudadana y la prevención del delito, dentro de un marco de desempeño garante de la legalidad, del libre ejercicio de los derechos humanos, de la eficacia y la eficiencia administrativa y del respeto a la dignidad humana.

La seguridad ciudadana es compleja por su propia naturaleza social. A lo que se le añade otras dificultades; por una parte, la tendencia de la delincuencia organizada a adoptar, al igual que el terrorismo, estructuras en red, dotadas de una mayor flexibilidad y anonimato, constituyendo auténticas "redes

criminales” [Arquilla & Ronfeldt, 2003]. Y por la otra, todo lo que debilite las relaciones, los compromisos entre las personas crea inseguridad, luego la participación ciudadana es un elemento central en la idea de crear cohesión, un espacio público, una comunidad que vincule a las personas mediante redes vecinales e institucionales de una comunidad.

Por todo ello, en última instancia, las políticas y estrategias de seguridad ciudadana tienen el reto de crear o fortalecer las redes vecinales e institucionales de la comunidad en cuestión, pero a su vez debilitar o desbaratar las redes delincuenciales que operan en la comunidad. El mayor reto que tiene la seguridad ciudadana es fortalecer los vínculos de la comunidad en una sociedad que tiende a debilitarlos, el delincuente no crea problemas sociales; más bien son los problemas sociales los que crean delincuentes [Hraba, 1998].

En ambos casos el máximo interés está en conocer la estructura y funcionabilidad de esa red, es especial situar a los líderes de las redes, los que adoptan las decisiones tanto a nivel organizativo, como funcional. En las redes delincuenciales, esos no participan por lo general en la comisión de los delitos, lo cual representa no pocas dificultades a la hora de su incriminación en los hechos delictivos, de ahí la imperiosa necesidad de una buena información extraída a través de la inteligencia, máxime cuando en demasiadas ocasiones, la imagen pública de los dirigentes delincuenciales goza de prestigio y reconocimiento social. Mientras que en las redes vecinales son los que convocan y articulan a la comunidad, pero a su vez, son los más expuestos a represalias por los propios delincuentes, sea sobre su persona o sobre su familia.

Las intervenciones en redes delincuenciales entraña a largo plazo un deterioro más sensiblemente de la capacidad operativa, las razones principales son dos: por un lado permite conseguir información de gran calidad sobre la composición, modo de actuar y planes de estos grupos y por otro, genera psicosis de infiltración en los grupos de delincuentes, obligándoles a aumentar las medidas de seguridad internas y a cerrarse sobre ellos mismos, con lo que sus oportunidades de reclutar nuevos integrantes y de coordinarse con otras organizaciones se ven considerablemente mermadas. A partir de la determinación de objetivos fundamentales: ¿se quiere destruir la red, disminuir simplemente su capacidad para realizar acciones delictivas o desgajarla de su aparato de apoyo en el mundo de la legalidad?

Pudiendo valerse de operaciones encubiertas que transmitan información dentro de la red delincencial para que puedan variar o dificultar la realización de operaciones provocando inestabilidad en el entorno, el mero manejo de información, la creación de relaciones disfuncionales que descompongan gravemente las capacidades de la grupo delincencial, acciones destinadas a destruir la confianza mediante información y a crear desconfianza y acritud internas, intoxicación informativa, hasta ataques más directos a la propia organización con el fin de interrumpir sus actividades, dislocar o degradar sus capacidades o incluso destruirla completamente.

Esto podría tener no sólo un efecto corrosivo, sino además forzar a los delincuentes a efectuar movimientos que aumentaran su vulnerabilidad. Aunque todos los objetivos resulten legítimos, resulta esencial que exista claridad sobre la delimitación legal del marco de actuación aplicable. [Sansó-Rubert, 2006]

La Comisión Nacional para la Reforma Policial, constituida mediante Resolución del Ministerio del Interior y Justicia de fecha 10 de abril de 2006, tiene como mandato la proposición, a través de un proceso de diagnóstico y consulta amplio y participativo, de un nuevo modelo de policía en el contexto de la sociedad venezolana actual, mediante un marco jurídico institucional y de gestión que permita concebirla como un servicio público general orientado por los principios de permanencia, eficiencia, extensión, democracia y participación, control de desempeño y evaluación de acuerdo con procesos y estándares definidos y planificación y desarrollo en función de las necesidades nacionales, estatales y municipales, dentro del marco de la Constitución de la República Bolivariana de Venezuela y de los Tratados y Principios Internacionales sobre protección de los derechos humanos.

Dentro de este mandato general, se ha previsto un proceso amplio de consulta nacional y un diagnóstico de seguridad, criminalidad y percepción de la policía, que constituye un insumo valioso para determinar tendencias de criminalidad, victimización, expectativas y percepciones públicas sobre la función policial y mecanismos de contraloría social.

### 1.3 Antecedentes

En el artículo 29 con fecha 06 de noviembre del año 2001 se crea la Ley de Coordinación de Seguridad Ciudadana y en la misma en el Título IV Capítulo I establece el objeto, organización y funcionamiento del *Sistema Nacional de Registro Delictivo, Emergencia y Desastres (SINARDED)* con la finalidad de que los órganos de seguridad ciudadana dispongan de un sistema de información que facilite la debida planificación, formulación y ejecución integral de los planes, estrategias y acciones de Seguridad Ciudadana. Este sistema centralizado, es utilizado como un gran repositorio de datos, extraídos de Bases de Datos externas, tales como: ONIDEX, SETRA, CICPC, DICOSEVIP y ANTECEDENTES PENALES.

Este intento por integrar información de los diferentes órganos de seguridad ciudadana presenta algunas limitantes:

1. No se logró automatizar el proceso de carga y transformación de los datos provenientes de las diferentes fuentes.
2. El diseño relacional no está normalizado lo que provoca inconsistencia en los datos.
3. Los tiempos de respuesta no son adecuados por la complejidad de las consultas solicitadas.

Es válido destacar que SINARDED no es la única aplicación informática en explotación utilizada por los Órganos de Seguridad Ciudadana, cuentan además con:

1. *Sistema Integrado de Información Policial*, que permite gestionar el control de armas, reseñas de personas, objetos, hechos delictivos, denuncias, etc. Perteneciente al Cuerpo de Investigaciones Científicas, Penales y Criminalísticas. (CICPC)
2. *Sistema del Instituto Nacional de Tránsito y Transporte Terrestre* que se encarga de almacenar información relacionada con los propietarios y registros de vehículos, placas, multas y accidentes.
3. *Sistemas de Gestión para Centros de Emergencias 171*, encargados de almacenar los datos referentes a los incidentes de emergencia ocurridos en el área monitoreada que puede comprender un municipio o estado.

#### **1.4 Necesidad de un Centro de Tratamiento y Análisis de Información de Seguridad Ciudadana.**

Para promover medidas y estrategias eficaces es necesario analizar un número ingente de datos asociados a los escenarios sociales donde ocurren los incidentes, permitiendo detectar tendencias y comportamientos análogos, que sirvan de base a la preparación de programas de intervención social que estén dirigidos a las causas que los provocan.

Para el análisis de la violencia y el delito se recurre a los factores asociados. Este es un punto de vista que tiene como base una visión probabilística, flexible y multidimensional. Reconoce la importancia e incidencia de las estructuras económica, social y de control formal, pero las trasciende reconociendo y proyectando dinámicamente la acción de las personas en las dimensiones educativa, cultural e histórica, al tiempo que ubica la familia como factor central, seguida de las amistades, escuela y, barrio. También reconoce que los medios de comunicación son factores que inciden transversalmente en la vida nacional, tanto en el área urbana como rural. Persigue una comprensión mucho más ajustada a la realidad que el enfoque casual.

Para obtener la información proveniente del entorno social y de las características individuales de las personas, se hace necesario consultar los datos que manejan los diferentes órganos de Seguridad del Estado e Instituciones Públicas de la República Bolivariana de Venezuela. Se reconoce que la nación cuenta con algún grado de informatización dentro del Ministerio Público, pero los datos se recogen por diferentes instituciones (Policía, bomberos, hospitales y otras), y es muy complejo integrarlos y analizarlos de forma centralizada, quedando gran parte de esta información extraviada o inaccesible.

Ante esta problemática se propone el *diseño e implementación de un Centro de Tratamiento y Análisis de Información* (Ver Figura 1.1) en la estructura del Ministerio del Poder Popular para las Relaciones Interiores y Justicia de la República Bolivariana de Venezuela (MPPRIJ), que permita integrar la información relacionada con la seguridad ciudadana a nivel nacional a fin de elevar la efectividad de las estrategias y políticas para garantizar el orden y la tranquilidad ciudadana, elevar el seguimiento y control sobre prioridades establecidas en la actuación de los principales órganos policiales y brindar seguimiento en tiempo real a hechos y situaciones extraordinarias. [ALBET, 2006]

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, desarrollo territorial equilibrado y la estabilidad de la nación, requiere la creación del Centro de Tratamiento y Análisis de Información de Seguridad Ciudadana, el cual cumplirá 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.

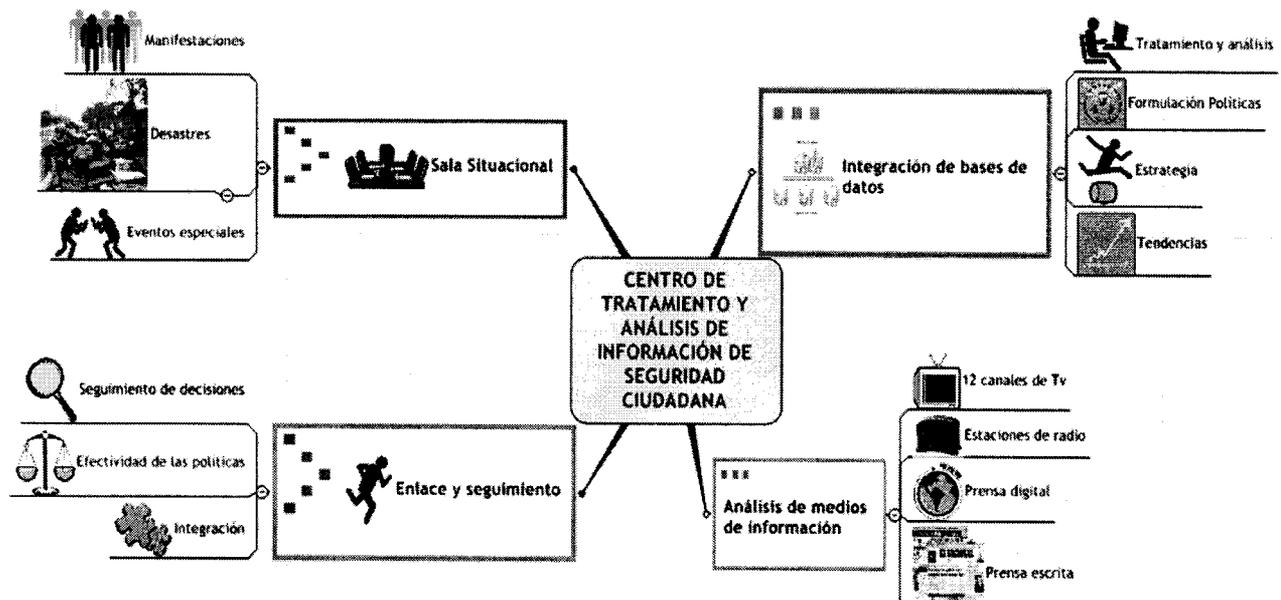


Figura 1.1: Esquema de la concepción general del CTAISC [ALBET, 2006]

Para este centro se propone la siguiente arquitectura:

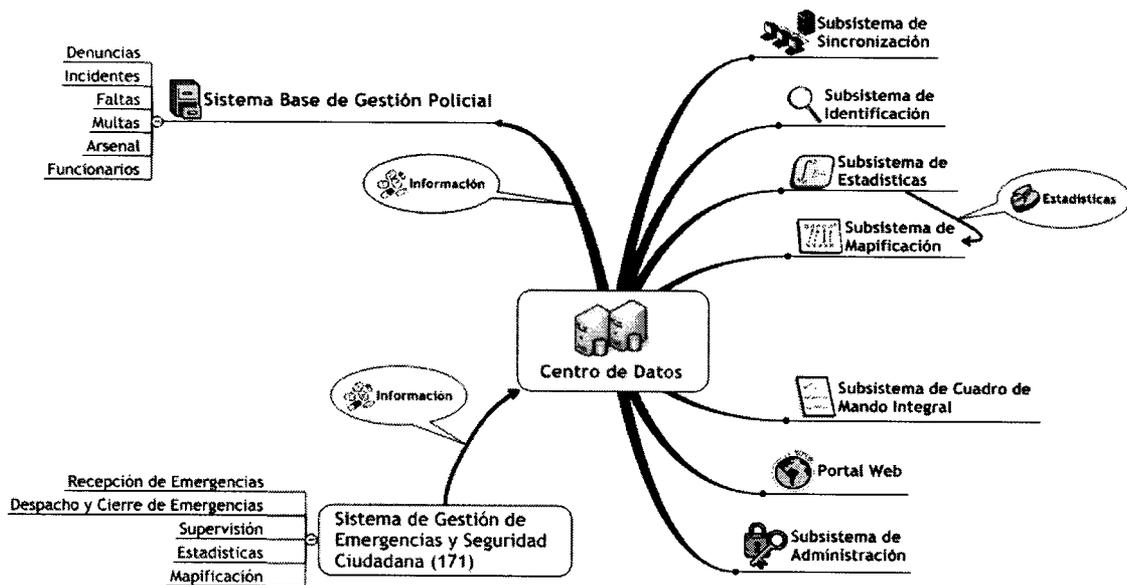


Figura 1.2 Arquitectura del Sistema. [ALBET, 2006]

### 1.5 La información como herramienta eficaz en la seguridad ciudadana.

La información tiene dos perspectivas de análisis en la seguridad ciudadana:

1. La primera, para conocer la realidad social.
2. La segunda, para intervenir en esa realidad social.

Para conocer la realidad social es necesario realizar múltiples análisis que aboguen por la búsqueda de variadas fuentes de obtención de información, que una vez procesadas sirvan para la elaboración de inteligencia que sustente estrategias y políticas de seguridad ciudadana. Mientras que la intervención en esa realidad social debe de conllevar a la explotación de nuevos mecanismos o de perfección de los ya existentes, con el fin de adecuar las capacidades estatales al nuevo escenario operativo y estratégico. [Sansó-Rubert, 2006]

La utilización efectiva de información para fundamentar los procesos de toma de decisiones en cualquier ámbito espacial (Estados, Municipios, Alcaldías y Parroquias) y temporal (época del año, meses,

semanas, días, horas del día) constituye una constante histórica, hasta el punto de que resulta incuestionable que, en el dominio del conocimiento sobre seguridad ciudadana, ha proporcionado garantías de éxito.

El dominio de la información ha cobrado paulatinamente trascendental proyección como elemento de prevención e identificación de riesgos ciudadanos. Defenderse de las amenazas y aprovecharse de las oportunidades que el entorno puede ofrecer, mediante el empleo de las capacidades de inteligencia, es una necesidad que ha permanecido inmutable a lo largo de los siglos. Sin embargo, lo que ha cambiado enormemente han sido las formas y los medios a través de los cuales se obtiene y procesa la información.

En la actualidad, una dificultad añadida a la hora de abordar este objeto de estudio de por sí complejo, lo constituye la falta de consenso sobre todas aquellas manifestaciones que entran en la categoría conceptual de lo que puede ser considerado como “inteligencia”<sup>3</sup>. A este respecto, evitando entrar en disertaciones y conflictos doctrinales, al no existir una definición comúnmente aceptada con validez internacional, se ha optado por adoptar como referencia una definición avalada por diversos autores expertos en la materia y contemplada a su vez en documentos oficiales.

*La inteligencia* es el resultado de un proceso cíclico que consta de varias fases. Hay distintas versiones del ciclo de inteligencia. Las fases reciben nombres diferentes dependiendo de las agencias y de la bibliografía utilizada. Básicamente la secuencia de un modelo simplificado e ideal de ciclo de inteligencia es la siguiente: [Sansó-Rubert, 2006]

1. Fase de Dirección: Los “consumidores” de la inteligencia indican la clase de información que necesitan y esas necesidades generales son convertidas en requerimientos específicos por los responsables de alto nivel de inteligencia. Los requerimientos indican cómo se deben distribuir los recursos y marcan las directrices de los que “obtienen” información.

---

<sup>3</sup> A pesar de la abundancia de definiciones aportadas hasta el momento, el término “inteligencia” sigue estando precariamente definido y envuelto en una continua discusión con respecto a su contenido. Sí existe al menos, una postura unánime en torno a su naturaleza: todos coinciden en que no se trata de una simple suma de datos recopilados a partir de diversas fuentes, sino que es el producto de un determinado modo de conectar y analizar estos datos, por parte de los analistas de los Servicios de Inteligencia, para descubrir, comprender y valorar los hechos a los que se remiten y prever su posible evolución, con el fin de suministrar conocimiento especializado y estructurado, que permita al Estado tomar decisiones adecuadas y reducir los riesgos inherentes a toda acción.

2. Fase de Obtención: Los que trabajan en la obtención consiguen información bruta.
3. Fase de Elaboración: Esa información bruta es la materia prima con la que trabaja el analista y acaba convirtiéndose en inteligencia.
4. Fase de Difusión: La inteligencia es distribuida a los consumidores, que marcan nuevas necesidades o hacen ajustes en los programas de inteligencia con el fin de mejorar la eficacia y eficiencia.

En el campo concreto de la inteligencia, contar con fuentes de obtención de inteligencia humana, de procesamiento de información, o con suficientes analistas especializados, requiere siempre varios años de preparación e inversión. Se trata de capacidades que no se improvisan. La inversión en capacidades de inteligencia y de captación de información, constituye una elección estratégica de política de seguridad ciudadana a medio y largo plazo.

Las fuentes de obtención de inteligencia son múltiples y de muy variada índole: [Sansó-Rubert, 2006]

- Inteligencia humana.
- Inteligencia técnica (inteligencia de imágenes, inteligencia de señales, captación de comunicaciones canalizadas a través de diversos medios: radio, teléfonos, fax o Internet)
- Inteligencia de fuentes abiertas.

## **1.6 Integración, almacenamiento y distribución de información.**

En la actualidad, por los grandes volúmenes de información existentes, no es posible pensar que los analistas recolecten y analicen la información existente en las diferentes instituciones del ministerio público de manera manual. Las tecnologías de la información han evolucionado y en la actualidad existen tecnologías capaces de integrar, almacenar y distribuir información con tiempos de respuesta aceptable.

Una solución altamente eficiente y validada son los Almacenes de Datos que convierten los datos operacionales de una organización en una herramienta profesional, haciéndolos disponibles a los usuarios que lo necesiten para el análisis y toma de decisiones [Thomsen, 2002]. Según Bill H. Inmon, padre reconocido del Data Warehousing, "El Data Warehouse es una colección de datos orientados al tema, integrados, no volátiles e historizados que facilitan la toma de decisiones". [Inmon, 2005]

Debe ser *Integrado* porque los datos almacenados en el Data Warehouse deben unirse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

Es *Temático* porque los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Las peticiones de información serán más fáciles de responder dado que toda la información reside en el mismo lugar.

El *Tiempo* es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de actividades en el momento presente. Esta información almacenada sirve, entre otras cosas, para realizar análisis de tendencias. El Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

*No volátil* porque la información existe para ser leída, y no modificada. Es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

Los almacenes de datos cumplen el objetivo de acumular, homogenizar y brindar todos los datos relevantes para una actividad determinada, dígame información histórica de una entidad, resultados de una actividad o simplemente el valor de una determinada variable en el tiempo. [Giménez, 2003]

Un Data Warehouse es simplemente una base de datos, generalmente con un esquema específico que facilita la consulta de los datos. Esto no implica que la implantación de un Data Warehouse sea sencilla, ni mucho menos. Un Data Warehouse puede almacenar la información en tablas relacionales, o bien en bases de datos multidimensionales.

En primer lugar un almacén de datos no es un producto que pueda ser comprado en el mercado, sino más bien un concepto que debe ser construido. Es una combinación de conceptos y tecnología que cambian significativamente la manera en que es entregada la información a las personas encargadas de tomar las decisiones estratégicas. Ofrece acceso fácil y rápido a información vital para la gestión y operación de una institución.

Debe tener las condiciones necesarias para asegurar su funcionamiento, digase respaldo eléctrico adecuado, conectividad estable, poseer un plan de realización de salvadas de seguridad, y una adecuada seguridad.

El objetivo del DW es el de satisfacer los requerimientos de información interna de la organización para una mejor gestión. El contenido de los datos, la distribución y estructura son dirigidos a satisfacer las necesidades de información de los analistas y usuarios tomadores de decisiones

Los sistemas transaccionales son dinámicos, constantemente se encuentran actualizando datos. Analizar esta información puede presentar resultados distintos en cuestión de minutos, por lo que se deben extraer y almacenar fotografías de datos<sup>4</sup>, para estos efectos, con la implicancia de un consumo adicional de recursos de cómputo. Llevar a cabo un análisis complejo sobre un sistema transaccional, puede afectar al sistema.

Los almacenes de datos generan bases de datos tangibles con una perspectiva histórica, utilizando datos de múltiples fuentes que se fusionan en forma congruente. Estos datos se mantienen actualizados, pero no cambian al ritmo de los sistemas transaccionales. Muchos de estos almacenes se diseñan para contener un nivel de detalle hasta el nivel de transacción, con la intención de hacer disponible todo tipo de datos y características, para reportar y analizar.

Así un almacén de datos resulta ser un recipiente de datos transaccionales para proporcionar consultas operativas, y la información para poder llevar a cabo análisis multidimensionales. De esta forma, dentro de él existen dos tecnologías que se pueden ver como complementarias, una relacional para consultas y una multidimensional para análisis. [Sperley, 1999]

Puede considerarse que el modelo relacional en el cual se basa el Procesamiento Transaccional en Línea<sup>5</sup>, tiene como objetivo mantener la integridad de la información (relaciones entre los datos) necesaria para operar de forma eficiente los datos almacenados. Sin embargo, este modelo no corresponde a la forma en que el usuario percibe la información que necesita.

---

<sup>4</sup> Del inglés "snapshots".

<sup>5</sup> Del inglés "OnLine Transactional Processing" con las siglas OLTP.

Un DW está basado en un procesamiento distinto al utilizado por los sistemas operacionales, es decir, éste se basa en Procesos de Análisis en Línea<sup>6</sup>, usado en el análisis de información en las organizaciones. Para ampliar los conceptos anteriores, en la tabla 1.1 se exponen las principales diferencias entre los sistemas Transaccionales y los Data Warehouse. [Nader, 2003]

Transaccionales	Data Warehouse
Admiten el acceso simultáneo de muchos usuarios -miles- que agregan y modifican datos.	Admiten el acceso simultáneo de muchos usuarios -cientos- que consultan datos.
Representan el estado, en cambio constante de la organización, pero no guardan su historial.	Guardan la historia de una organización.
Contiene grandes volúmenes de datos, incluidos los datos asociados a la configuración y administración.	Contiene grandes volúmenes de datos, sumariados, consolidados y transformados.
Tienen estructuras de bases de datos complejas porque se debe garantizar la integridad referencial mediante la normalización.	Tienen estructuras de bases de datos simples y los datos se encuentran desnormalizados para lograr un mayor entendimiento de los usuarios.
Se ajustan para dar respuesta a la actividad transaccional que exigen tiempos de respuestas mínimos.	Se ajustan para dar respuesta a las consultas analíticas que no requieren tiempos de respuesta instantáneos.
Proporcionan la infraestructura tecnológica necesaria para realizar las operaciones diarias.	Proporcionan la infraestructura tecnológica necesaria para realizar análisis de los datos.
Se requiere de una amplia experiencia técnica para crear consultas a la medida por la compleja estructura de datos.	Pueden combinar datos de orígenes heterogéneos en una única estructura homogénea y simple, facilitando la creación de informes y consultas.

<sup>6</sup> Del inglés " OnLine Analysis Process " con las siglas OLAP.

Las consultas analíticas que resumen grandes volúmenes de datos afectan negativamente la capacidad del sistema para responder a las transacciones en línea.	Organizan los datos en estructuras simplificadas buscando la eficiencia de las consultas analíticas.
El rendimiento del sistema cuando está respondiendo a consultas analíticas complejas puede ser lento o impredecible, lo que causa un servicio poco eficiente a los usuarios del proceso analítico en línea.	Contienen datos transformados que son válidos, coherentes, consolidados y con el formato adecuado para realizar el análisis sin interferir en las operaciones transaccionales diarias.
Los datos que se modifican con frecuencia interfieren en la coherencia de la información analítica.	Proporcionan datos estables que representan el historial de la empresa. Se actualizan periódicamente con datos adicionales, no con transacciones frecuentes.
La seguridad se complica cuando se combina el análisis en línea con el proceso de transacciones en línea.	Simplifican los requisitos de seguridad debido a que los usuarios solo consultan información.

**Tabla 1.1 Diferencias entre sistemas transaccionales y basados en Data Warehouse [Nader, 2003]**

El acceso a todos los datos a veces no es conveniente o necesario para determinados usuarios que solo necesitan un subconjunto de éstos. En estos casos se utilizan los Datamart. [Poe *et al.*, 1998]. El concepto Datamart es una especialización del Data Warehouse, y está enfocado a un departamento o área específica, permitiendo así un mejor control de la información que se está abarcando.

## **Conclusiones del Capítulo.**

Por todo lo antes expuesto, se puede llegar a las siguientes conclusiones parciales:

- La gestión de la Seguridad Ciudadana debe sustentarse en el manejo de grandes volúmenes de información desde múltiples fuentes de información.
- La Seguridad Ciudadana es una variable compleja lo que supone el tratamiento de datos multidimensionales y a diferentes niveles y en una relación espacio-temporal.
- Los Almacenes de Datos responden a las exigencias del CTAISC, esta tecnología permite integrar y analizar grandes volúmenes de datos de diversas fuentes.
- Existen fuentes de información en sistemas transaccionales en la República de Venezuela que permiten alimentar al DW del CTAISC.
- La revisión mostrada constituye el marco de referencia necesario para dar cumplimiento al objetivo propuesto en este trabajo.

## Capítulo 2: Conceptualización del Data Warehouse del CTAISC.

**E**n este capítulo se describe la conceptualización del sistema informático como herramienta de integración, transformación y almacenamiento de datos para el tratamiento y análisis de la información de Seguridad Ciudadana por el CTAISC adscrito al Ministerio del Poder Popular para Relaciones Interiores y Justicia (MPPRIJ) de la República Bolivariana de Venezuela.

En el mismo se expone la metodología a emplear en el desarrollo del sistema informático basado en DWH que apoya el trabajo a realizar en los análisis de seguridad ciudadana. Se definen fases para la construcción del DWH sobre la concepción de diferentes estructuras multidimensionales que conformarán el almacén, las fases del flujo de desarrollo del DWH y las fuentes u orígenes de datos que poblarán el almacén. Por último se exponen las conclusiones del capítulo.

### 2.1 Fases para el enfrentamiento a problemas de seguridad ciudadana.

En la prevención y ataque a los problemas de seguridad ciudadana es importante analizar 4 etapas: [Valdivia, 2002]

- **Diagnóstico:** Tiene como resultado el conocimiento de los problemas que afectan la Seguridad Ciudadana.
- **Elaborar el plan de acción:** Plan de acción que contiene una estrategia con objetivos, metas claras y precisas, y actividades concretas.
- **Ejecutar el plan de acción:** Ejecución del Plan de acción.
- **Evaluar el plan y su impacto:** Evaluación de productos, resultados e impacto del plan en la población.

El análisis de las estadísticas policiales, desempeño policial y de la justicia y la percepción de la comunidad deben ser usados para definir los problemas de seguridad ciudadana, establecer prioridades y describir cuáles son sus elementos y características principales. Los resultados que provengan de tal

análisis permitirán detectar cuáles son aquellos problemas que se deben enfrentar primero. [Valdivia, 2002]

Para realizar un análisis de Seguridad Ciudadana adecuado, es imprescindible tener, entre otras informaciones, las siguientes: [Valdivia, 2002]

1. Características de las víctimas de los delitos;
2. Características de los victimarios;
3. Presencia de alcohol y drogas en la comisión de los delitos;
4. Características de los hechos delictivos;
5. Factores que inciden en la comisión de los delitos;
6. Características físicas y sociales de la comunidad.

La tarea de identificar los problemas de la comunidad no sólo involucra la recolección de información estadística y la que le pueda entregar la misma comunidad, sino que además es necesario establecer qué información es importante y cuál no.

La realización de esta labor debe entenderse como una tarea compartida entre los distintos departamentos de la Municipalidad y actores relevantes en el ámbito de la seguridad ciudadana de la comuna, como lo pueden ser las instituciones policiales y gubernamentales, los colegios, comercios y demás organizaciones presentes en la misma comuna. [Valdivia, 2002]

## **2.2 Modelo para la conceptualización del Data Warehouse.**

La conceptualización del modelo del DW contextualizado a la gestión de datos de seguridad ciudadana se efectúa relacionando los factores claves de éxito del ciclo de inteligencia que se operacionaliza en fases, unido a las concebidas para el desarrollo de la tecnología informática para la gestión de bases de datos: [Hobbs, Hillson & Lawande, 2005] Diseño, Construcción, Análisis y Mantenimiento.

Así, se construye el modelo distinguiendo cuatro pilares interrelacionados: dirección, obtención, elaboración y difusión e impactos, basados en el principio de corresponsabilidad ciudadana de los actores del sistema. A continuación, se comentará la operacionalización de los ejes del modelo.

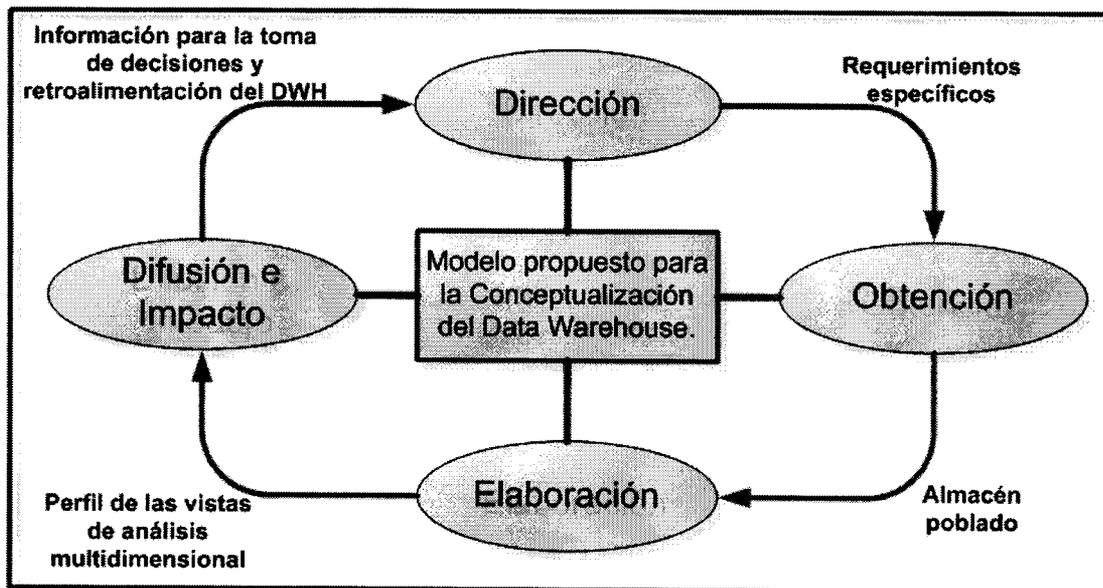


Figura . Modelo para la conceptualización del Data Warehouse.

Fuente: Elaborado por Jorge González Ferrer, Asnioby Hernández López y Yanet Peña Vazquez.

La perspectiva de dirección refiere a la arquitectura del DW en función de las necesidades y alcance de la gestión de información que requiere el tratamiento y análisis de la información de seguridad ciudadana. Es decir, las necesidades de información a satisfacer de los usuarios finales, donde se debe definir claramente su alcance y la estrategia, las que son convertidas en requerimientos específicos para el diseño del DWH.

El eje de obtención, se trata de la obtención de la información bruta para lo cual se requiere documentar las fuentes y evaluar las brechas entre las necesidades de información y su disponibilidad real en esas. Si es factible, se construye el DWH a partir de los procesos de extracción, transformación y carga de los datos de las fuentes, resultando el almacén poblado que responde al diseño propuesto.

En la elaboración, sobre el almacén poblado se perfilan las vistas multidimensionales que respondan a las necesidades de información, a las que le anteceden los precálculos, sumalizaciones y agregaciones de los datos, contenidas en la fase de análisis en el desarrollo del DWH. Obteniéndose el perfil de la vistas en función de las necesidades de información a satisfacer.

Por último, en la difusión e impacto se especifica los niveles de acceso a la información generada por el proceso de análisis de los perfiles multidimensionales, sus formas de presentación y reportes. Además, se evalúa la pertinencia de la información brindada en función de las necesidades, este proceso se ubica en el mantenimiento del DWH. Esta perspectiva actúa como retroalimentación natural del DWH para su desarrollo y adaptación a las necesidades de información de seguridad ciudadana para la toma de decisiones y retroalimentación del DWH en el entorno en el cual está ubicado.

### 2.3 Metodología a utilizar.

En el capítulo anterior quedó justificado el empleo de la tecnología de Data Warehouse (DW) para dar tratamiento a la información derivada de las fuentes existentes en la Republica Bolivariana de Venezuela, con el fin de ayudar a derivar estrategias de prevención e intervención en materia de seguridad ciudadana.

Un DW puede ser construido siguiendo un proceso iterativo e incremental de cuatro fases. Ellas son: [Hobbs, Hillson & Lawande, 2005]

1. **Diseño:** Se refiere al proceso de elegir y diseñar la estructura interna, la que soportará la manipulación y recuperación de los datos.

2. **Construcción:** Se refiere al diseño e implementación de las herramientas encargadas de llevar los datos de las fuentes al repositorio, y ya una vez dentro de este los homogenizará. En base a la construcción existen tres aspectos a considerar:

- Conectar las fuentes mediante extractores, se refiere a la implementación de los extractores que deben conocer el formato, protocolo y ubicación tanto de las fuentes como del DW.
- Integrar las fuentes, se refiere a reunir la información presentada en las diversas fuentes en un mismo sitio con un mismo formato.
- Poblar el repositorio con los esquemas y vistas definidos: una vez que la fase de diseño fue cumplida en base al análisis realizado, se procede a poblar el DW con la información de las fuentes en los esquemas definidos.

3. **Análisis:** Es la explotación que se realiza de los datos almacenados dentro del DW a través de técnicas que facilitan y hacen más eficiente la consulta de los mismos. Ya con el DW poblado lo último es diseñar e implementar una interfaz que le permita al usuario final interactuar con el sistema, brindándole todas las ventajas del análisis dinámico de la información.

4. **Mantenimiento:** Proceso encargado de llevar los datos nuevos que se ingresen al repositorio y también se encarga de actualizar los datos previamente calculados en las vistas. [Hobbs, Hillson & Lawande, 2005]

## **2.4 Modelo multidimensional.**

Para el Diseño del Almacén de Datos se emplea la representación de un Modelo Multidimensional [Inmon, 2002]. El modelado multidimensional es una técnica de diseño lógico que busca presentar los datos en un estándar que permite la recuperación adecuada de éstos. [Nader, 2003]

El modelo multidimensional es además un modelo adecuado que provee un camino viable para agregar hechos a lo largo de múltiples atributos, llamados dimensiones. Para entender la definición presentada, es preciso definir los conceptos de: *Cubo*, *Dimensión* y *Medida*.

### **2.4.1 Cubo, Dimensión y Medida.**

Un modelo multidimensional soporta el manejo de una vasta cantidad de datos institucionales y temporales. De esta forma surge la instancia del modelo multidimensional, también conocido como cubo o hipercubo. Para clarificarlo un poco se puede imaginar un cubo con tres dimensiones, donde cada una de ellas tiene diferentes niveles o hechos, que cuando se intersectan se obtiene una medida. [Palomar & Trujillo, 2001-2002]

La *dimensión* es un concepto esencial dentro del modelo multidimensional porque es usada para seleccionar y agregar datos a un cierto nivel deseado de detalle. Cada instancia de una dimensión o valor de una dimensión corresponde a un nivel particular. [Nader, 2003]

Por otra parte, *la medida o hecho* es un dato numérico que representa una actividad específica de una organización o empresa. Se obtiene al intersectar varias dimensiones. Pueden tomar valores atómicos o derivados que dependen de fórmulas para calcularse. [Nader, 2003]

Viendo los datos dentro de un cubo se tiene la ventaja de que se puede manejar cualquier número de dimensiones definidas para ese Cubo. Sin embargo, usualmente un cubo tiene entre 4 y 12 dimensiones. Generalmente, un cubo soporta una vista de dos o tres dimensiones simultáneamente. [Nader, 2003]

#### **2.4.2 Tablas de Dimensiones.**

Las tablas de dimensiones describen los *objetos relevantes para la organización*. Cada tupla de la tabla se compone del identificador de un objeto y de sus atributos descriptores. Los datos de las tablas de dimensiones son generalmente de tipo alfanumérico y para un buen diseño se orienta tener los datos desnormalizados, lo que aporta ventajas para el acceso a los datos. [Windom, 1995] Entre ellas se encuentran:

- Evitar la concatenación de tablas durante las consultas.
- Sencillez de uso y mejor entendimiento de la estructura de los datos. [Windom, 1995]

La desnormalización también tiene inconvenientes, como lo es la redundancia de los datos, lo que requiere mayores capacidades de almacenamiento.

Otra de las orientaciones en el diseño de las Tablas del Dimensiones, es la inclusión de la dimensión *tiempo*, porque muchas consultas son restringidas por criterios relativos a períodos de tiempo como puede ser el “último mes” y “días festivos”. [Nader, 2003] El autor sólo refiere el tiempo pues en su estudio el espacio no tiene sentido práctico, sin embargo, en aplicaciones más extensivas el espacio cobra su real interés tal es el caso de la seguridad ciudadana en un país.

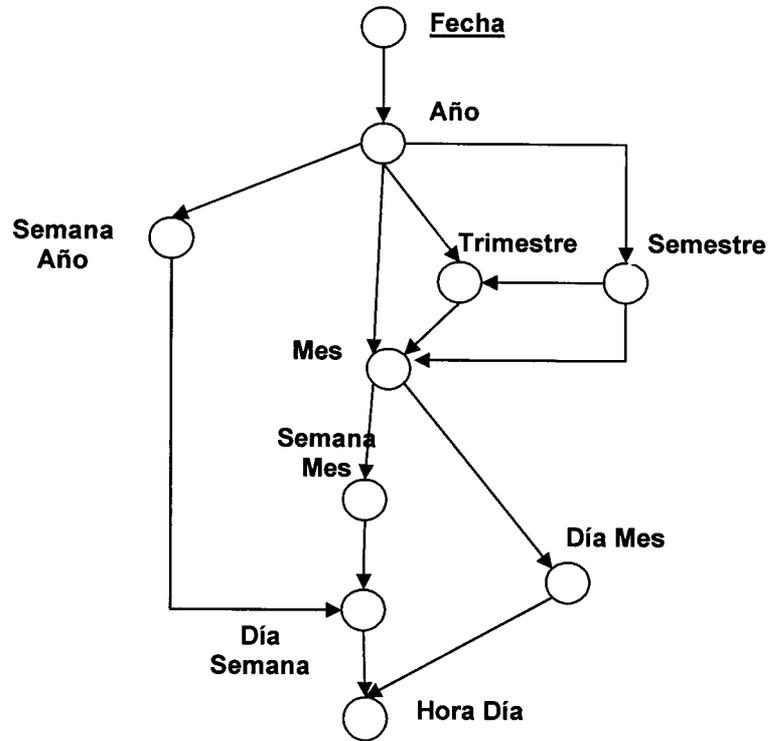
#### **2.4.3 Definiendo dimensiones.**

Entre las dimensiones que se han identificado dentro del DW del CTAISC se encuentran las siguientes:

### Dimensión DIM\_Temporal.

<b>Nombre</b>		Dimensión Temporal	
<b>Descripción</b>		Esta dimensión da una visión temporal de los hechos delictivos.	
<b>Llave</b>	<b>Atributo</b>	<b>Tipo de dato</b>	<b>Descripción.</b>
PK	IDNumérico	INTEGER	Llave primaria de la tabla.
	Fecha	DATE	Fecha
	Anno	INTEGER	Año
	Semestre	INTEGER	Periodo de 6 meses (1er o 2do).
	Trimestre	INTEGER	Periodo de 3 meses(1er, 2do, 3ro, 4to).
	Mes	INTEGER	Mes del año.
	SemanaAnno	INTEGER	Semana del año (1...48).
	Semanales	INTEGER	Semana del mes (1...5).
	DíaMes	INTEGER	Día del mes (1...31).
	DiaSemana	INTEGER	Día de la semana (1...7)
	<i>VisperaDiaFestivo</i>	<i>BIT</i>	<i>Vispera de día festivo (Si o No).</i>
	<i>DiaFestivo</i>	<i>BIT</i>	<i>Día Festivo (Si o No).</i>
	HoraDia	VARCHAR(45)	Hora del día (mañana, tarde, primeras horas de la noche, madrugada).

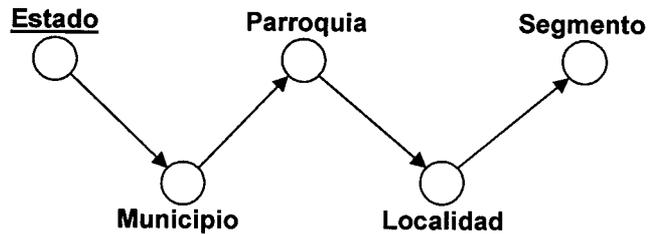
**Jerarquía DIM\_Temporal.**



**Dimensión DIM\_Espacial.**

<b>Nombre</b>	Dimensión Espacial		
<b>Descripción</b>	Esta dimensión define el área geográfica en que se efectuó el hecho delictivo.		
<b>Llave</b>	<b>Atributo</b>	<b>Tipo de dato</b>	<b>Descripción.</b>
PK	idEspacial	INTEGER	Llave primaria de la tabla.
	Estado	VARCHAR(45)	Nombre del Estado.
	Municipio	VARCHAR(45)	Nombre del Municipio.
	Parroquia	VARCHAR(45)	Nombre de la Parroquia.
	Localidad	VARCHAR(45)	Nombre de la Localidad.
	Segmento	VARCHAR(45)	Nombre del Segmento.

## Jerarquía DIM\_Espacial



## Dimensión DIM\_Lugar

<b>Nombre</b>	Dimensión Lugar		
<b>Descripción</b>	Esta dimensión ubica el escenario social en que ocurre el hecho delictivo.		
<b>Llave</b>	<b>Atributo</b>	<b>Tipo de dato</b>	<b>Descripción.</b>
PK	IDLugar	INTEGER	Llave primaria de la tabla.
	EntornoFisico	VARCHAR(45)	Breve descripción del entorno en que se realizó el hecho.
	AmbienteSocial	VARCHAR(45)	Breve descripción de la
	Droga	BIT	Presencia de drogas en el lugar. Si o No.
	PresenciaPolicial	BIT	Presencia de policías en el lugar. Si o No.
	Iluminación	VARCHAR(45)	Si el lugar tiene iluminación o no. (Buena, Aceptable, Mala, Inexistente)
	Graffiti	BIT	Presencia de mensajes obscenos en las paredes. (Si o No)
	Suciedad	BIT	Poca higiene en la zona, basura en las calles. etc. (Si o No)
	Deterioro	BIT	Zona deteriorada, con derrumbes, salideros de agua, vías dañadas, etc. (Si o No)
	Pandillas	BIT	Existencia de pandillas en la zona. (Si o No)

### Dimensión DIM\_Víctima

Nombre		Dimensión Víctima	
Descripción		Esta dimensión describe las características individuales de las víctimas.	
Llave	Atributo	Tipo de dato	Descripción.
PK	idVictima	INTEGER	Llave primaria de la tabla.
	Edad	INTEGER	Valor que indica la edad de la víctima.
	Genero	BIT	Sexo de la víctima (F o M)
	InclinacionSexual	VARCHAR(45)	Indica la inclinación sexual de la victima (homosexual o heterosexual).
	ClaseSocial	VARCHAR(45)	Define la clase social de la víctima (alta, media o baja).
	Habitat	VARCHAR(45)	Condiciones de la vivienda.
	GrupoEtnico	VARCHAR(45)	Grupo étnico al que pertenece la victima. (Blanco, Mestizo, Indio, Amarillo, Negro)
	RutinaDiaria	VARCHAR(45)	Actividades frecuentes.
	Denunciabilidad	BIT	Expresa si la víctima denunció al victimario o no.
	Multivictimizacion	BIT	Expresa si la víctima ha sido atacada en varias ocasiones.

### Dimensión DIM\_Victimario

Nombre		Dimensión Victimario	
Descripción		Esta dimensión describe las características individuales de los victimarios.	
Llave	Atributo	Tipo de dato	Descripción.
PK	idVictimario	INTEGER	Llave primaria de la tabla.
	EstadoCivil	VARCHAR(45)	Estado civil del victimario (Casado, Soltero, Divorciado o Viudo).
	Nacionalidad	VARCHAR(45)	Nacionalidad del Victimario.

	Empleado	BIT	Dice si es empleado o desempleado.
	NivelIngreso	VARCHAR(45)	Breve descripción del nivel de ingreso del victimario.
	SectorLaboral	VARCHAR(45)	Define el sector laboral del victimario.
	MedioFamiliar	VARCHAR(45)	
	NivelEducativo	VARCHAR(45)	Nivel educacional del victimario.
	Creencia	VARCHAR(45)	Tipo de creencia.
	Religion	VARCHAR(45)	Tipo de religión.
	Edad	INTEGER	Valor numérico que expresa la edad del victimario.
	Genero	BIT	Sexo de la víctima (F o M).
	Habitat	VARCHAR(45)	Condiciones de la vivienda.
	GrupoEtnico	VARCHAR(45)	Grupo étnico al que pertenece la víctima. (Blanco, Mestizo, Indio, Amarillo, Negro)
	RutinaDiaria	VARCHAR(45)	Actividades frecuentes.
	AdiccionBebida	BIT	Dice si es adicto a la bebida o no.
	AdiccionDroga	BIT	Dice si es adicto a la droga o no.
	AparienciaFisica	VARCHAR(45)	Breve descripción de la apariencia física del victimario.

#### 2.4.4 Tablas de Hechos.

Las tablas de hechos contienen datos sobre las *actividades básicas de la organización*. Cada tupla de la tabla se compone de los datos observables de la actividad y de las referencias a las dimensiones que los caracterizan. Estas últimas se denominan claves ajenas. [Giménez, 2003]

Para lograr un buen diseño de las Tablas de Hechos, se orienta el uso de claves ajenas, que pueden ser agrupadas en dos tipos principales:

- Claves ajenas con significado, en las que no es necesario acceder a la tabla de dimensiones para algunas consultas.

- Claves ajenas sin significado (identificadores internos del sistema) que permiten la actualización del identificador real de un objeto sin implicar actualizaciones en las tablas de hechos. [Giménez, 2003]

El tipo de clave más recomendado es la clave ajena sin significado porque evita inconsistencias a la hora de adicionar datos al almacén, por la diversidad de las fuentes de la que se extrae la información. Por ejemplo, en dos sistemas diferentes, podría almacenarse información de una persona con el número de cédula solamente y en otros sería este mismo número de cédula con un prefijo para identificar su procedencia (nacional o extranjero). También permite almacenar información histórica de la organización pues se pueden identificar unívocamente hechos con claves naturales diferentes en el tiempo.

Los datos de las tablas de hechos son, en la mayoría de los casos, de tipo numérico lo que permite hacer cálculos estadísticos y agregación, representando generalmente una relación muchos a muchos entre varias entidades (dimensiones). [Giménez 2003]

### 2.4.5 Definiendo Tablas de Hechos.

Entre las Tablas de Hechos que se han identificado, se tienen las siguientes:

**Tabla: FACT\_Hechos**

<b>Nombre</b>	Tabla de medidas: Hechos.		
<b>Descripción</b>	Esta tabla contiene información asociada a hechos delictivos.		
<b>Llave</b>	<b>Atributo</b>	<b>Tipo de dato</b>	<b>Descripción.</b>
PK	idHecho	INTEGER	Llave primaria de la tabla.
FK	idVictimario	INTEGER	Llave foránea de DIM_Victimario.
FK	idVictima	INTEGER	Llave foránea de DIM_Victima.
FK	idLugar	INTEGER	Llave foránea de DIM_Lugar.
FK	idEspacial	INTEGER	Llave foránea de DIM_Espacial.
FK	Intemporal	INTEGER	Llave foránea de DIM_Temporal.
	TipoHecho	VARCHAR(45)	Clasificación del hecho.

	Clasificacion	VARCHAR(45)	Término policial.
	Arma	VARCHAR(45)	Tipo de arma usada.
	Motivo	VARCHAR(45)	Motivo del hecho delictivo.
	Gravedad	VARCHAR(45)	Gravedad del hecho.
	RelVictimaVictimario	VARCHAR(45)	Relación víctima-victimario.
	TipoDenunciabilidad	VARCHAR(45)	Quien informa el incidente.
	CantVictima	INTEGER	Cantidad de víctimas.
	CantVictimario	INTEGER	Cantidad de victimarios.
	Fuente	VARCHAR(45)	Fuente de la que se obtuvo la información.

#### 2.4.6 Cubo Multidimensional.

También conocido como Cubo de Decisión, es una potente herramienta que permite distribuir las métricas y dimensiones en filas, columnas y páginas, mostrando agrupaciones de los resultados y permitiendo reorganizar la representación “en línea”. [Action Ware S.L, 2004]

Los cubos pueden organizar la información que poseen en esquemas de estrella o de copo de nieve. El esquema estrella se basa en una tabla de hechos central (que contiene las medidas) la cual se enlaza a las tablas de dimensiones relacionadas (las categorías descriptivas de las medidas), mientras que el esquema copo de nieve, enlaza una tabla de hechos central a las tablas de dimensiones relacionadas, pero estas a su vez se enlazan a otras tablas dimensionales. [Inmon, 2002]

Con este tipo de esquemas se simplifica el entendimiento de los datos por parte del usuario, a la vez que maximiza el desempeño de las consultas de la base de datos para aplicaciones de soporte de decisiones. [Inmon, 2002]

#### 2.4.7 Definiendo Cubos.

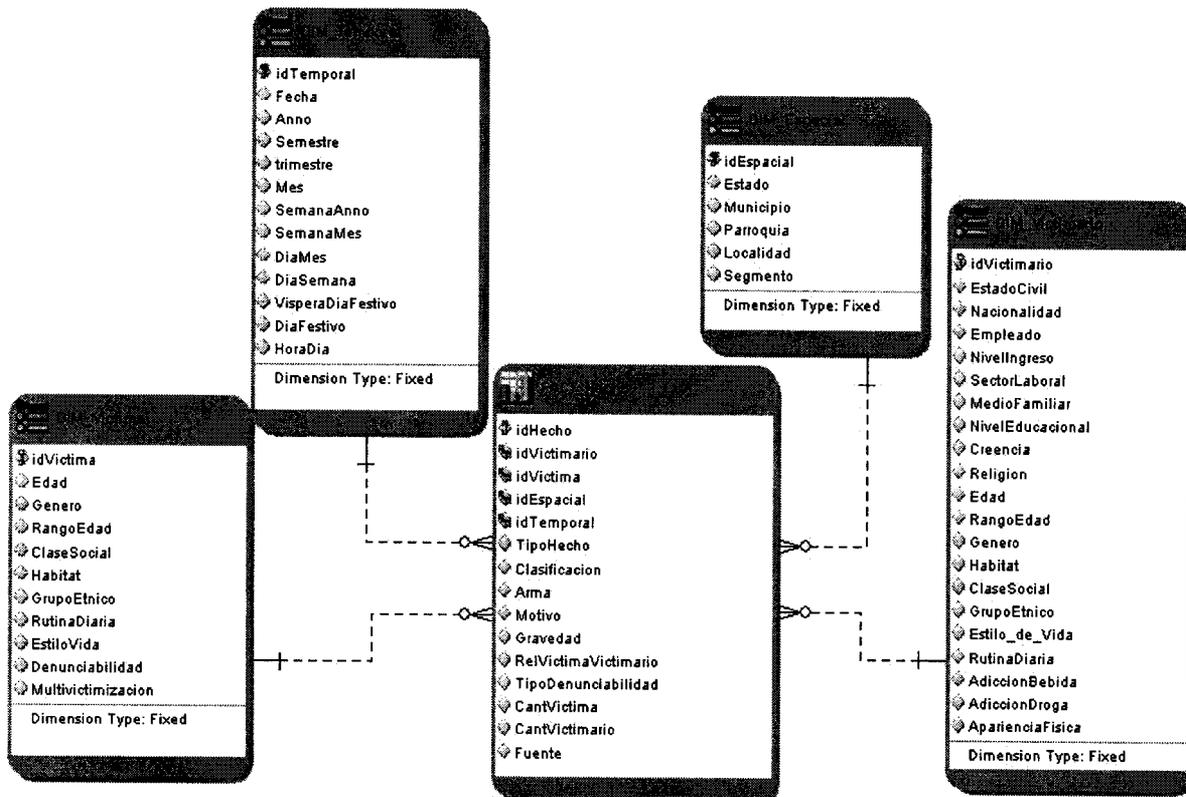
Para analizar y diagnosticar los problemas de Seguridad Ciudadana, se han identificado tres líneas de investigación: [Thomé, 2004]

- La Percepción Ciudadana.

- Actuar Policial.
- Estadísticas Policiales.

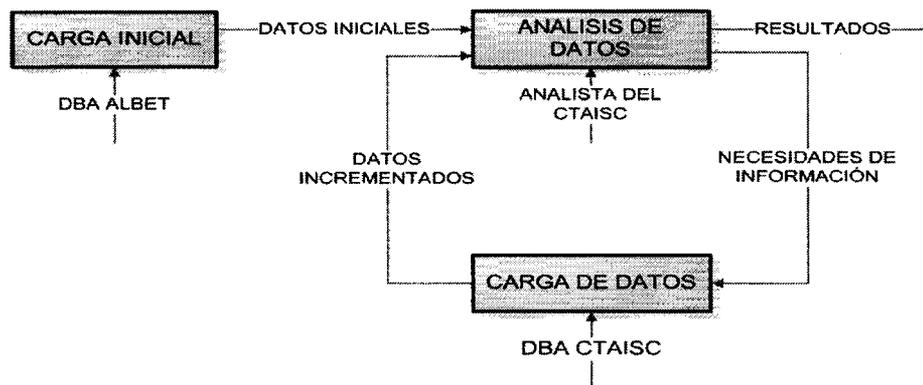
Por el momento se ha conceptualizado la instancia Hechos Delictivos, que corresponde a la línea de las Estadísticas Policiales.

### Cubo 1. Hechos delictivos.



### 2.5 Diseño.

Para comprender los procesos del Data Warehouse del CTAISC se ha elaborado el siguiente diagrama IDEF, representativo del tratamiento y análisis de la información. Ver Figura 2.1.



**Figura 2.1: Diagrama IDEF. Procesos Elementales de Negocio: [García, 2007]**

La primera fase es la *carga inicial de los datos* por los administradores de bases de datos ALBET. Esta información inicial debe ser procesada por los analistas del CTAISC, quienes obtendrán resultados e identificarán las necesidades de información. Estas nuevas necesidades serán atendidas por los informáticos del CTAISC, y guardadas en el almacén de datos, siendo actualizadas e incrementadas. [García, 2007]

La figura 2.2 muestra los diferentes componentes agrupados en paquetes que conforman el sistema desde el punto de vista del diseño de la arquitectura del Software del Sistema. También muestra el sentido del flujo de la información.

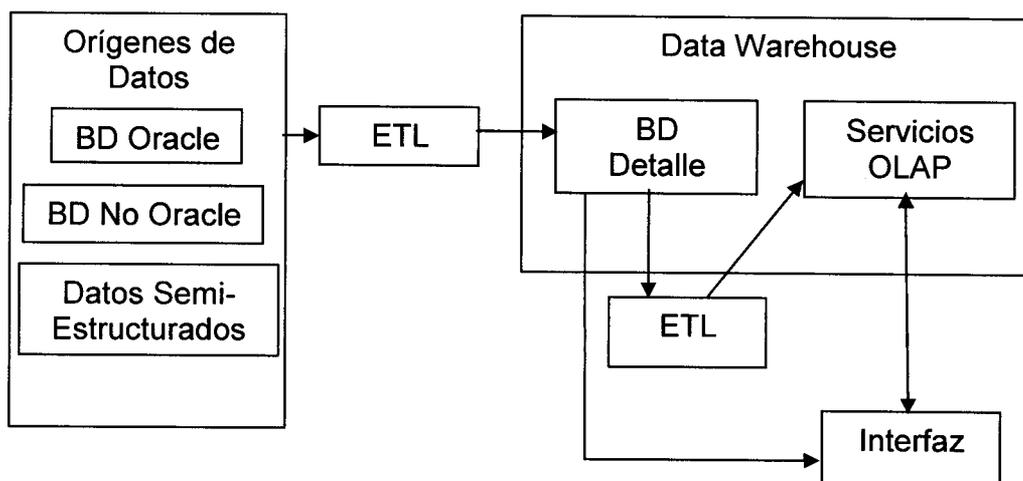
El paquete *Orígenes de datos* gestiona las conexiones e incluye la información de las bases de datos de los sistemas que poblarán el almacén. Estas fuentes pueden ser *Base de Datos Oracle*, *Bases de Datos No Oracle* y *Datos Semi-estructurados*.

El paquete ETL<sup>7</sup>, implementa las funcionalidades de Extracción, Transformación y Carga de Datos desde los sistemas fuentes a la Base de Datos Detalle. El paquete Data Warehouse contiene la Base de Datos Detalle y los Servicios OLAP, los cuales se implementan sobre estructuras de cubos multidimensionales utilizando los datos existentes, y se encargan de almacenar los datos integrados y las vistas multidimensionales.

<sup>7</sup> Del inglés "Extraction, Transformation and Load" con las siglas ETL.

Además, el paquete ETL implementa la funcionalidad de extracción y carga de datos desde la Base de Datos Detalle a los Cubos OLAP, realizando todos los precálculos y agregaciones necesarias.

Finalmente, para completar el diseño lógico del sistema, el paquete *Interfaz* permite al usuario final consultar información de los Cubos y en caso de ser necesario, visualizar datos, con un menor grado de sumariación, de la *Base de Datos Detalle*.



**Figura 2.2** Diseño de la arquitectura del Software del Sistema.

La topología que se decidió usar es Topología de estrella ya que una sola tabla de hechos está relacionada con cada tabla de dimensión. Las tablas de dimensiones son enlazadas a la tabla de hechos mediante referencias a una llave foránea. La llave primaria en la tabla de hechos se compone de la relación de las llaves primarias de las tablas de dimensiones.

En caso de ser necesario, el esquema de estrella puede ser redefinido al esquema de copo de nieve con un soporte de jerarquía de atributos, permitiendo que las tablas de dimensiones tengan tablas de subdimensiones. Sin embargo se recomienda utilizar esquemas tipo estrella, ya que este tipo de estructura está diseñada para optimizar los ambientes de consulta, y muchas de las funcionalidades que optimizan el desempeño en este tipo de ambientes están orientadas a trabajar sobre este tipo de modelado.

La herramientas definidas para la construcción de data warehouses y procesos de ETL, facilitan la creación de este tipo de modelos con asistentes de creación de dimensiones y tablas de hechos. Todo este proceso permite que la carga de los datos del almacén favorezca la construcción del DW.

## **2.6 Construcción.**

El proceso de construcción es el encargado de llevar la información de las fuentes hacia el DW. Inicialmente es posible descomponer la construcción en cuatro pasos: extraer los datos de las fuentes, transformarlos al formato interno del DW, integrarlos y almacenar los resultados integrados. Los pasos de extracción y transformación pueden ser agrupados en un mismo componente, siendo denominado éste el *módulo del extractor*. El paso integración es usualmente unido junto con las capacidades de almacenamiento en un mismo componente.

### **2.6.1 Extracción.**

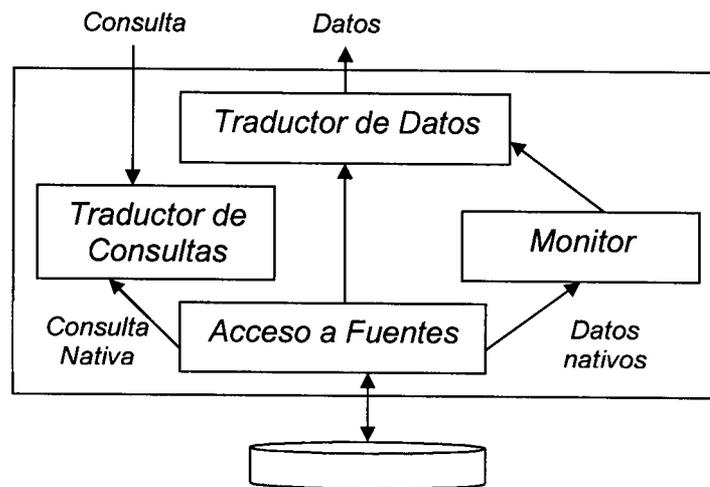
El proceso de extracción implica la recuperación de datos de las fuentes. Esto requiere de un programa que conecte las fuentes de interés con el DW, limpie los datos y escriba los datos seleccionados en el formato específico del DW. Entre las fuentes u orígenes de datos seleccionados para poblar el almacén del CTAISC, se tienen los siguientes sistemas [ALBET, 2006]:

- **SIGESC:** Sistema de Gestión de Emergencias 171. Esta solución proporciona aplicaciones informáticas, que permiten gestionar las actividades relacionadas con la atención de las emergencias en un Centro 171. Entre sus principales funciones se encuentran la recepción de llamadas, el registro de incidencias y despacho de las emergencias.
- **SIGEPOL:** Sistema de gestión policial, que permite almacenar la información generada por los comandos de las policías estatales y municipales referente a denuncias, hechos delictivos, faltas y operativos policiales.
- **TARSYS:** Sistema de Monitoreo y catalogación de contenido audiovisual. Solución para el desarrollo de sistemas de información que permite la grabación digital desasistida de canales de televisión y radio.

- **SIIPOL:** Sistema Integrado de Información Policial que permite gestionar el control de reseñas de personas, objetos, hechos delictivos, denuncias, etc. Perteneció al Cuerpo de Investigaciones Científicas, Penales y Criminalísticas.
- **SINARDED:** Sistema Nacional de Registro Delictivo, Emergencias y Desastres. Tiene como misión centralizar la información y almacenar datos de los órganos de Seguridad del país en un solo repositorio, manejando la situación de cada una de las personas residentes en el país, delitos, faltas, objetos robados y/o solicitados, etc.
- **INTTT:** Sistema del Instituto Nacional de Tránsito y Transporte Terrestre. Este sistema maneja información relacionada con los propietarios y registros de vehículos, placas, multas y accidentes.
- **Encuesta Nacional de Victimización y Percepción Policial:** Constituye un instrumento para determinar el volumen y las características de los hechos delictivos cometidos durante un período de tiempo. Aporta información valiosa sobre las características personales y perfiles de las víctimas, su vinculación con el victimario, “modus operandi” de los autores, alcance del daño y condiciones espaciales y temporales de los delitos. [ALBET, 2006]

Cada una de estas fuentes aporta información relevante y de interés para el CTAISC, permitiendo modelar los diversos escenarios que afectan a la población de Venezuela.

El extractor tiene acceso directo a las fuentes de las cuales recupera la información que es transportada a un traductor de datos, o bien a un traductor de consultas. También es importante resaltar la presencia de un módulo de monitoreo encargado de detectar los cambios ocurridos en las fuentes. (Ver figura 2.3). [Thomsen, 2002]



**Figura 2.3 Arquitectura de un extractor [Thomsen, 2002]**

El extractor conoce el formato de las fuentes así como el formato de representación de datos de DW, conoce el protocolo de comunicación y la ubicación de ambos. Esta estructura varía según las necesidades y tipos de datos que se extraigan de las fuentes. [Thomsen, 2002]

### 2.6.2 Integración.

El objetivo es integrar los datos provenientes de las múltiples fuentes para finalmente integrarlos en el DW. En general, la integración es un proceso de dos pasos: [Corey, 1998].

1. Homogenización: Datos de fuentes individuales son transformados conforme a un esquema global de datos. Dicho proceso es realizado por los extractores.
2. Integración: Durante este paso, las piezas homogenizadas son puestas juntas. Es decir, se agrupa la información y se deposita en el DW, según el esquema definido.

Las tareas principales del integrador son: descomponer la consulta ejecutada por el DW en subconsultas dirigidas a las fuentes, fusionar los datos provenientes de las diversas fuentes y finalmente cargar estos ya fusionados en el DW. Los datos son llevados al integrador por los extractores. Este tiene un módulo de

proceso de consultas, el cual las envía al administrador de consultas, encargado de actualizar la información en el repositorio.

El trabajo del integrador consiste en recibir notificaciones de cambios por parte del extractor y reflejar dichos cambios en el DW. Con un alto nivel de abstracción, el DW puede ser visto como un conjunto de vistas materializadas, desarrollando el mantenimiento de las vistas. Debe a su vez realizar la recolección de todos los datos recuperados por los extractores, de forma que la información no sea duplicada.

Otra tarea del integrador es la apertura de los datos dentro del DW, lo cual consiste en poder revisar los niveles de agregación y ordenamiento, así como la construcción de índices y la partición en áreas de almacenamiento. [Inmon, 2005]

## **2.7 Análisis.**

Según el diccionario de la Real Academia de la Lengua Española, *análisis* es la acción de descomponer algo en sus partes constitutivas con el fin de poder estudiarlo. De esta forma, al hablar de análisis y visualización, se refiere a la manipulación o explotación que tiene el usuario del DW, la forma en que éste consultará el sistema y la parte de información a la que tendrán acceso los diversos usuarios.

Una solución para el análisis de los datos almacenados en el DW son las técnicas OLAP. El objetivo de las herramientas OLAP es proveer a los usuarios el análisis multidimensional de la información. Para lograr esta meta, dichas herramientas emplean un modelo multidimensional para el almacenamiento y presentación de los datos.

Las características de las técnicas OLAP son: [Thomsen, 2002]

- Recuperación rápida de datos, ya que la recuperación se puede realizar en una sola operación de entrada y salida,
- El almacenamiento es muy eficiente, ya que los bloques sólo contienen datos y los índices son simples.

### 2.7.1 Operaciones.

Las tareas de modelado en el mundo de las bases de datos, se realizan desde tres perspectivas diferentes. *La conceptual*, que lidia con un alto nivel de representación del mundo; *la física*, que lidia con los detalles de representación de la información; y *la lógica* que actúa como un intermediario entre los dos niveles anteriores, tratando de balancear el paradigma de almacenamiento independiente y una representación natural. [Pence & Creeth, 2002]

Los operadores relacionados con las técnicas OLAP son:

- Rotar (Swap): Alterar las filas por columnas (permutar dos dimensiones de análisis).
- Bajar (Down): Bajar el nivel de visualización en las filas a una jerarquía inferior.
- Detallar (Drill-down): Informar para una fila en concreto de datos a un nivel inferior. El usuario navega entre los niveles de los datos desde los más resumidos hasta los más detallados.
- Expandir (Expand): Ídem anterior sin perder la información a nivel superior para éste y el resto de los valores.
- Agregar (Roll-up): Puede ser simple o expresiones complejas que afecten a la consolidación de datos de varias dimensiones.
- Rebanar y cortar (Slicing y Dicing): Capacidad de ver la base de datos desde diferentes puntos de vista.
- Pivotar (Pivoting): Consiste en reorientar la visión multidimensional de los datos.
- Anidamiento (Nesting): Técnica de visualización utilizada para mostrar el resultado de una consulta multidimensional que devuelve un sub-cubo.
- Reach through: Medio de extender los datos accesibles al usuario final más allá de lo que se encuentra almacenado en el servidor OLAP, consultando y recuperando de forma automática datos de un almacén o sistema OLTP. [Pence & Creeth, 2002]

## **2.8 Mantenimiento.**

La posibilidad de tener “datos frescos”, es importante para las aplicaciones de Data Warehouse. Existen dos formas de refrescar datos: la primera es llevando los datos al DW segundos después de que las fuentes fueron actualizadas, un ejemplo claro de esto son las transacciones de un banco. La segunda es acumulando y almacenando los datos ya integrados y transformados, en un sitio intermedio para finalmente de forma periódica pasar dicha información al DW.

Refrescar un DW consiste en propagar las actualizaciones de las fuentes. Hablando del refrescado, hay dos cuestiones que se deben considerar: ¿Cuándo refrescar? y ¿Cómo refrescar? Respondiendo al cómo refrescar, se puede realizar un refrescado incremental o bien un re-calculado de los datos; y respondiendo al cuándo, se puede hacer a solicitud explícita del administrador, o periódicamente con un tiempo determinado.

Uno de los métodos más empleados es el refrescado periódico, pero esto depende mucho de las cualidades de los datos que maneje el DW. Así pues, el refrescado de un DW es considerado como un problema crítico y difícil debido a tres principales razones.[Inmon, 2005]

Primero, el volumen de datos almacenados en el DW es muy grande y crece cada vez más. La segunda razón, es que el refrescado debe ser accesible a los diferentes cambios de desempeño o ejecución del DW. Y por último, el refrescado envuelve transacciones que acceden múltiples datos, lo que implica cálculos complejos si se produce un alto nivel de agregación. El componente de extracción debe ser capaz de recuperar y guardar todos los cambios que ocurren en las fuentes.

## **2.9 Validación del modelo.**

El modelo propuesto fue evaluado por especialistas del MPPRIJ, entre los que se encontraban Ingenieros Informáticos, Administradores de bases de datos y Policías, quienes validaron y aprobaron la propuesta para su futura implementación en el CTAISC.

## 2.10 Conclusiones del capítulo.

Se puede concluir que:

- La metodología para el desarrollo de un DW a emplear será un proceso iterativo e incremental de cuatro fases: Diseño, Construcción, Análisis y Mantenimiento.
- La topología que se decidió usar es Topología de Estrella ya que una sola tabla de hechos está relacionada con cada tabla de dimensión, y puede migrar fácilmente a copo de nieve en caso de ser necesario.
- El modelo multidimensional es un modelo adecuado que provee un camino viable para agregar hechos a lo largo de múltiples atributos, llamados dimensiones.
- Las dimensiones propuestas en este Capítulo, reflejan las necesidades de información del CTAISC.
- Deben definirse los atributos de los cubos a un nivel de detalle mayor.

## Capítulo 3. Selección de las herramientas informáticas.

### Introducción.

**E**n este Capítulo se analizarán las diferentes herramientas propuestas para el desarrollo del Data Warehouse del CTAISC. Se realizará una descripción técnica general de cada componente a utilizar, y la explicación de la combinación e integración entre ellos. Se entiende que la elección de una arquitectura de software adecuada es fundamental para garantizar los niveles de servicio y funcionalidad necesarios.

### 3.1 Elección.

Para seleccionar la tecnología adecuada a la solución propuesta se han evaluado diferentes criterios. El más importante, aclarando que no fue el único analizado, lo constituye que el gestor de base de datos Oracle es líder mundial en procesamiento de grandes volúmenes de datos, con capacidades de administrar y manejar eficientemente varios Terabytes.

Otro criterio analizado es que actualmente están en despliegue o desarrollo sistemas sobre Bases de Datos Oracle, específicamente SINARDED, CICPC, SIGEPOL, SIGESC, etc., que se han identificado como posibles fuentes para poblar el almacén del CTAISC. Al utilizarse las herramientas Oracle en la solución Data Warehouse, se favorece la evolución tecnológica futura y aumentan las facilidades de integración con los sistemas existentes. Las transacciones serían más sencillas entre origen (sistemas fuentes) y destino (*data warehouse*).

En consulta con los especialistas de Oracle Venezuela, fueron identificadas herramientas agrupadas en tres capas: ETL, Data Warehouse y Visualización; y apoyadas por *Oracle Warehouse Builder 10g*, *Oracle Database Enterprise Edition 10g* y *Business Intelligence Standard Edition* respectivamente.

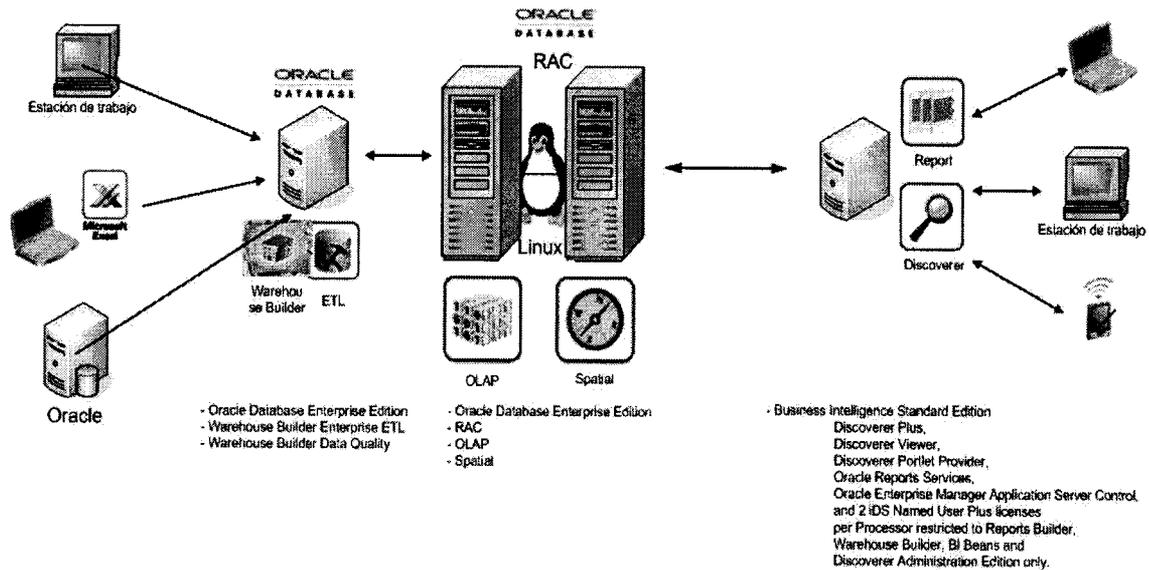


Figura 3.1 Propuesta de arquitectura tecnológica. [Nariño, 2007]

### 3.2 Oracle Database Enterprise Edition 10g.

Oracle Database Enterprise Edition es un gestor de base de datos que almacena información en estructuras relacionales, objeto-relacionales y multidimensionales. Está diseñado para soportar soluciones de tipo OLTP, DSS<sup>8</sup>-OLAP y Data Warehouse, contando con numerosas características de desempeño optimizadas para OLAP relacional, entre las que se encuentran:

- Extensiones *ANSI SQL*<sup>9</sup> para manejo de consultas OLAP.
- Definición de dimensiones y niveles de dimensión dentro del motor relacional.
- Vistas materializadas y Administración de Sumarizaciones. Las vistas materializadas permiten un manejo inteligente de la información resumizada y precálculan operaciones de agregación, unión u otras. [Nariño, 2007]

<sup>8</sup> Del inglés "Decision Support System" con las siglas DSS.

<sup>9</sup> Lenguaje declarativo de acceso a bases de datos relacionales.

### 3.2.1 Manejo de tablas agregadas y sumariadas.

El tamaño de un data warehouse abarca entre decenas de Gigabytes hasta unos cuantos Terabytes, en el cual la mayoría de los datos son almacenados en tablas de hechos que llegan a ser muy grandes. Para mejorar el rendimiento de los sistemas y disminuir los tiempos de respuesta de las consultas ejecutadas sobre estas tablas de hechos, Oracle proporciona Vistas Materializadas, que no son más que objetos de la base de datos.

La definición de una vista materializada puede incluir cualquier número de agregaciones: (SUM, COUNT(x), COUNT (\*), COUNT (DISTINCT x), AVG, VARIANCE, STDDEV, MIN, y MAX) y cualquier número de uniones (JOIN).

Las Vistas Materializadas reducen considerablemente el tiempo de respuesta para obtener los resultados de la consulta solicitada, estas realizan las operaciones de agregación y unión en la base de datos antes de la ejecución de dichas consultas.

Para lograr un mayor entendimiento, se expone el siguiente ejemplo de vista materializada con la que se obtiene la cantidad de homicidios ocurridos en las diferentes parroquias de los estados, siendo las víctimas de sexo masculino:

```
CREATE MATERIALIZED VIEW Homicidio_Parroquia
BUILD IMMEDIATE          <- Cuando construirla
REFRESH FORCE             <- Cómo refrescarla
ON DEMAND                <- Cuando refrescarla
ENABLE QUERY REWRITE     <- Si se puede reescribir la vista o no
AS                        <- Consulta resultante contiene

SELECT de.estado, de.parroquia, dt.anno, dt.mes, COUNT (*) as Total_Homicidios
FROM DIM_Temporal dt, DIM_Espacial de, DIM_Victima dv, FACT_Hecho fh
WHERE dt.IdTemporal = fh.IdTemporal and de.IdEspacial = fh.IdEspacial and
      dv.IdVictima = fh.IdVictima and dv.sexo = 'M' and fh.TipoHecho = 'Homicidio'
GROUP BY de.estado, de.parroquia, dt.anno, dt.mes;
```

### 3.2.2 Mecanismos de actualización y control de cambios.

Oracle provee diversos mecanismos para actualizar y controlar adecuadamente los cambios de los datos. Algunas de las opciones implementadas se enuncian a continuación:

- Establecer diferencias entre la Base de Datos fuente con una copia remota, por medio de la sentencia MINUS. Esta opción es válida cuando la copia de las tablas debe ser exacta.
- Implementación del proceso manualmente con la utilización de disparadores, que controlen los cambios entre columnas cuando son tablas con diferentes valores.

Estas técnicas presentan algunas limitaciones:

- En los datos que han cambiado en más de una oportunidad, solo será posible recuperar los valores actuales.
- Aumentan la carga de trabajo y pueden afectar el rendimiento de los sistemas transaccionales.
- No se podrán recuperar los valores que tienen una corta duración. [Hobbs, Hillson & Lawande, 2005]

Para reducir estas limitaciones se incorpora una nueva opción denominada "*Change Data Capture*<sup>10</sup>" que se encarga de publicar y administrar los datos que han cambiado en el tiempo de manera sincrónica o asincrónica.

Es la solución propuesta, se hace uso de la Opción "*Change Data Capture*" debido a que se requiere información actualizada y el refrescamiento de las dimensiones será incremental para disminuir la complejidad de las sumalizaciones y agregaciones necesarias. Este mecanismo permite trabajar de manera asíncrona facilitando la recuperación de datos utilizando los archivos *Redo logs*<sup>11</sup>, sin aumentar la complejidad de las transacciones de los sistemas fuentes.

---

<sup>10</sup> Control de los datos cambiados en el tiempo.

<sup>11</sup> Archivo que contienen las trazas de las transacciones realizadas sobre una Base de Datos.

### 3.2.3 Integridad de datos.

La base de datos de Oracle provee mecanismos para garantizar la integridad referencial de los datos, para lograr este fin implementa restricciones que garantizan el cumplimiento de reglas definidas para los datos. Las más utilizadas son:

- **UNIQUE:** Garantiza que el valor de una columna sea único.
- **NOT NULL:** Impide que la columna tome valores nulos.
- **FOREING KEY:** Valida la relación entre las tablas de dimensiones y las tablas de hechos.

Las restricciones pueden ser utilizadas en un almacén de datos para los siguientes propósitos:

- **Limpieza de datos:** Verifican que los datos tengan un nivel básico de consistencia y precisión en el Data Warehouse, evitando la introducción de datos erróneos o incompletos.
- **Optimización de las consultas:** La base datos utiliza las restricciones para optimizar las consultas, importantes para la reescritura de las consultas SQL y de las vistas materializadas.

El almacén de datos del CTAISC debe contener información validada, consistente, integrada y fidedigna, que apoye la formulación de políticas y estrategias para mejorar los niveles de seguridad ciudadana. Por estas razones se deben realizar una adecuada limpieza de datos y estandarización de las escalas de análisis.

Ejemplos de tales problemas de limpieza y necesidad de transformaciones lo constituyen los siguientes:

- Los formatos de fecha pueden ser diferentes en los distintos sistemas transaccionales que sirvan de fuente y para guardar la fecha se requiere estandarizar dicho formato,
- Pueden estar omitidos el día del mes y la operación de limpieza puede consistir en grabar el día 1ro,
- Se puede requerir reflejar cada hecho delictivo una sola vez si el mismo apareciera repetido en una base de datos de denuncias y otra de reseñas

- En general, en materia de seguridad ciudadana es comun la necesidad de completar datos acerca de si la fecha del hecho es víspera de día feriado, si el horario en que se produjo es la madrugada o la mañana, a partir de los datos ya presentes en las bases operacionales.

### **3.2.4 Funciones analíticas.**

Oracle incorpora funciones analíticas que pueden ser utilizadas para definir fórmulas de transformación de datos y/o vistas materializadas. Se documentan a continuación algunas de ellas:

- **Ranking:** La función puede examinar la salida completa antes de producir el resultado del requerimiento. Para esto Oracle provee funciones como: RANK, DENSE\_RANK, PERCENT\_RANK, CUME\_DIST y NTILE.
- **Ventana de Agregación:** Involucra el procesamiento acumulativo de las funciones SQL de agregación tales como: AVG, SUM, MIN, MAX, COUNT, VARIANCE y STDDEV.
- **Reportes de Agregación:** Uno de los tipos más comunes de cálculos es la comparación a diferentes niveles de agregación. Para ello se involucra el procesamiento acumulativo de las funciones SQL de agregación tales como AVG, SUM, MIN, MAX, COUNT, VARIANCE y STDDEV.
- **Regresión Lineal:** Oracle provee funciones de regresión lineal incluyendo la pendiente, intersección, coeficiente de correlación, entre otras.
- **Porcentaje inverso:** Permite realizar consultas para encontrar los datos que corresponde a valores percentiles.
- **Primera/Última Agregación:** Permite definir grupos de datos ordenados a partir de una consulta de datos y retornar el primero o el último valor de cada grupo.

Hacer uso de las funciones anteriores permite resolver problemas típicos de recuperación de datos en materia de seguridad, como son las comparaciones de las estadísticas delictivas de un periodo en comparación con el mismo periodo del trimestre anterior, hallar el delito de más incidencia en un horario y día de la semana determinados. Es posible agregar y sumarizar datos asociados a determinadas áreas geográficas como pueden ser estados, municipios o parroquias. Se pueden aplicar funciones analíticas sobre los datos que validen hipótesis científicas sobre el modo de actuación de los infractores de los delitos, entre otras.

### **3.2.5 Administración de datos.**

El almacenamiento de grandes volúmenes de datos es de vital importancia dentro de un ambiente Data Warehouse. Cuando se piensa en una solución de este tipo, es necesario que esta provea elementos como integración de datos provenientes de diferentes fuentes, con un rápido acceso a los datos, eficiente análisis, refresco de información periódico y escalable a medida que el volumen de datos lo exige, sin olvidar que la administración debe ser adecuada para este tipo de sistemas. Oracle provee estos elementos en sus soluciones, para el manejo de grandes volúmenes de datos, mediante características como: [Nariño, 2007]

- Paralelismo
- Particionamiento
- Manejo de índices avanzados
- Soporte a cluster en tiempo real
- Vistas materializadas
- Compresión de datos
- Grid control

Estas funciones son el soporte adecuado para una aplicación que necesita el manejo de grandes volúmenes de datos como ocurre en nuestro problema, ya que diariamente se espera el arribo de cientos de registros correspondientes a diferentes temas de seguridad como son los registros provenientes de los hechos reportados por CICPC y Centros 171 del país.

### **3.2.6 Replicación de Datos.**

Oracle garantiza la sincronización y replicación de datos entre varias instancias de bases de datos e incluso entre bases de datos heterogéneas. Proporciona además mecanismos para detectar los cambios sobre los datos (inserciones, actualizaciones y eliminaciones) que permite llevar a cabo movimientos regulares de datos entre bases de datos Oracle utilizando elementos que permiten obtener los últimos cambios realizados en los datos de una tabla y publicarlos a otra tabla local o de una base de datos remota.

La BD Oracle soporta estándares de conectividad como *ODBC*<sup>12</sup> y *OLE DB*<sup>13</sup> para conectarse a cualquier repositorio relacional y no relacional. Además existen componentes desarrollados por Oracle, llamados *gateways*<sup>14</sup>, los cuales permiten la conexión de manera nativa a las diferentes bases de datos existentes hoy en el mercado.

Oracle soporta replicación transparente, bidireccional, incremental, homogénea y heterogénea, sincrónica o asincrónica, de tipo maestro-esclavo o de tipo multi-maestro con resolución de conflictos. [Nariño, 2007]

### **Características de replicación de datos**

La Base de Datos Oracle tiene diferentes opciones para replicación en una o dos direcciones; el uso de las diferentes opciones depende de la necesidad de replicación y es posible combinarlas en una sola solución para lograr el esquema adecuado. [Nariño, 2007]

- Mediante la opción “Replicación Avanzada” de la Base de Datos se pueden replicar tablas, segmentos de tablas, o vistas materializadas de acuerdo con cada solución. Este mecanismo se auxilia de un ambiente gráfico para facilitar su configuración.
- *Oracle Streams* provee la funcionalidad de replicar segmentos de datos (tablas, subconjuntos de una tabla, esquemas o Bases de Datos), transacciones o eventos de manera rápida.
- *Oracle Data Guard* permite realizar copias automáticas de la Base de Datos en un sitio remoto.

De todos estos mecanismos de replicación proporcionados por Oracle, se usará solo *Data Guard*, debido a que la base de datos se ha concebido centralizada.

---

<sup>12</sup> Del inglés “Open Database Connectivity” con las siglas ODBC. Estándar de acceso a Bases de Datos, desarrollado por la Microsoft.

<sup>13</sup> Del inglés “Object Linking and Embedding for Database” con las siglas OLA DB. Estándar desarrollado por la Microsoft, que traducido al español significa: Incrustación y Enlace para Objetos de Bases de datos.

<sup>14</sup> Puertas de enlaces. Puede ser una computadora o componente que permite y controla el acceso a otras computadoras de la red.

### **3.2.7 Carga y población de datos.**

La base de datos permite, desde el *Enterprise Manager (EM)* definir tareas. En él se ejecutarían los *scripts*<sup>15</sup> que realizan la carga, definiéndose cada cuánto tiempo se repetirá este proceso.

#### **Recuperación de fallos y habilidad de empezar nuevamente**

La base de datos de Oracle provee un conjunto de mecanismos de protección, recuperación y respaldo para impedir la pérdida de datos ante fallas. El *Oracle Recovery Manager (RMAN)* es una herramienta integrada en la base de datos que satisface las demandas de rendimiento, administración de los respaldos y de los archivos de recuperación de la base de datos. *RMAN* tiene la capacidad de rastrear los cambios estructurales en la base de datos, fusionar imágenes de las bases datos y realizar respaldos incrementales para proveer una recuperación rápida y eficiente de la base de datos.

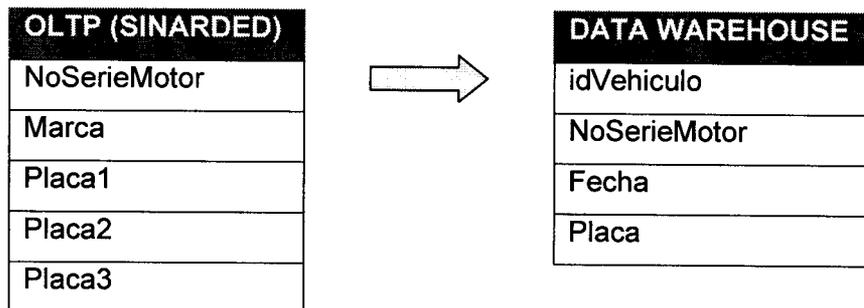
#### **Generación de llave secuencial automática en los procesos de carga.**

En la base de datos Oracle se puede crear secuencias automáticas que pueden ser utilizadas en los procesos de carga e inclusive en cualquier código PL/SQL o desde *Oracle Warehouse Builder* en el diseño del almacén.

Esta característica es de vital importancia para el proyecto, ya que en temas de seguridad, se hace necesario llevar registros históricos de los cambios sobre diferentes objetos: sobre las armas y sus diferentes propietarios en el tiempo, sobre los vehículos y sus propietarios y números de placa; información sobre los funcionarios y su historial de comportamiento en las diferentes entidades policiales donde ha brindado servicios.

---

<sup>15</sup> Fichero de texto, que contiene un conjunto de instrucciones, que puede ser ejecutado desde un intérprete de líneas de comandos.



**Figura 3.2 Histórico de placas de vehículos.**

En el ejemplo se aprecia como en el sistema SINARDED solo se permiten conservar hasta tres números de placa diferentes para un mismo vehículo, sin embargo en nuestro esquema, dado el id de un vehículo, se conservan cuantos registros de placas diferentes sea preciso.

#### **Procesamiento en paralelo.**

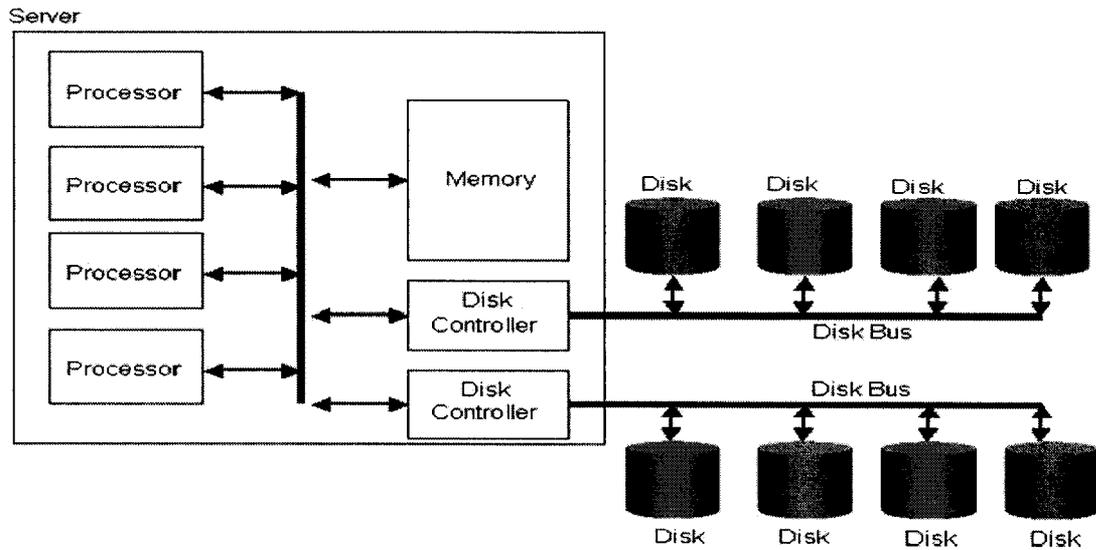
La base de datos Oracle soporta la ejecución o procesamiento en paralelo, para distribuir la carga de trabajo entre una o más tareas entre múltiples procesadores y recursos de I/O simultáneamente, garantizando la escalabilidad del sistema con la utilización adecuada de los recursos. Cuando la base de datos Oracle ejecuta una sentencia SQL en paralelo, múltiples procesos trabajan en conjunto simultáneamente para ejecutar una sola sentencia SQL; al dividir el trabajo necesario para ejecutar esa sentencia entre múltiples procesos Oracle puede ejecutar la sentencia más eficientemente que si solo la ejecutara con un solo proceso.

La base de datos Oracle soporta el procesamiento en paralelo entre múltiples procesadores de un solo sistema y entre múltiples nodos de un cluster con *Real Application Cluster (RAC)*. La base de datos Oracle soporta el procesamiento en paralelo para todo tipo de operaciones: carga de datos, creación de tablas e índices, creación de tablas temporales, particiones, recuperación, respaldo y replicación.

Las solicitudes en paralelo están presentes en la arquitectura de servidores del CTAISC. Cada uno de estos servidores *RAC* tiene cuatro procesadores que habilitan una estructura común de memoria, donde el

sistema operativo se ejecuta concurrentemente y separa los procesos en cada uno de los procesadores que lo integra. En la siguiente figura se muestran cuatro procesadores, donde cada uno accesa al mismo espacio de memoria a través de un mismo sistema de transporte (*bus*).

**Figura 3.3 Multiprocesador. [Hobbs, Hillson & Lawande, 2005]**



### Acceso a los registros de carga.

La base de datos Oracle permite mantener registros de cada una de las operaciones que se realizan en la base de datos, de los errores que pudiesen generar estas operaciones y de los procesos que se ejecutan.

Esta información es de mucha utilidad para los administradores y puede ser utilizada para detectar y corregir errores de configuración. El Almacén de Datos del CTAISC debe incorporar diariamente muchos registros y se hace necesario tener un control estricto sobre las operaciones que se han ejecutado y los resultados de las mismas.

### 3.2.8 Integración de los datos en tiempo real.

Para lograr integrar datos en tiempo real, o cercano al tiempo real, la Base de Datos Oracle cuenta con las siguientes funcionalidades:

**Colas Avanzadas de Oracle:** Esta funcionalidad permite la transformación de mensajes que puede ser utilizada para transformar, propagar, manipular y validar la comunicación entre distintas áreas.

**Captura y Control de los Cambios Oracle:** Esta funcionalidad permite llevar a cabo movimientos regulares de datos entre bases de datos oracle utilizando mecanismos que permiten obtener los últimos cambios realizados en los datos de una tabla y publicarlos a otra tabla local o de una base de datos remota.

**Servidor de Aplicaciones Oracle:** Provee componentes que permiten la integración de datos y eventos de sistemas para lograr un *DWH* con datos actuales. Incluye además componentes para la administración de flujos de procesos, integración y análisis o correlación de información agregada de eventos.

Esta característica es algo muy útil para el almacén del CTAISC, porque le aporta la capacidad de contar con datos actualizados a la hora de hacer análisis precisos de Seguridad Ciudadana, como por ejemplo los relativos a la densidad de un tipo de delito por ubicaciones geográficas, en tiempo real.

### 3.2.9 Índices.

Oracle Database 10g soporta *static bitmap indexes*<sup>7</sup> y *bitmap join indexes*<sup>8</sup> los cuales son utilizados para proveer un gran desempeño en ambientes de Datawarehousing.

### 3.3 Oracle Warehouse Builder.

Oracle Warehouse Builder (*OWB*) es la herramienta de diseño y desarrollo para Data Warehouse o Datamarts. Se utiliza para el modelado y la construcción de procesos de extracción, transformación y

carga, orientada a la construcción de modelos de datos, dando la ventaja de no necesitar herramientas adicionales para el modelado de datos. Cuando se generan físicamente los objetos del modelo destino *OWB* genera los correspondientes comandos DDL<sup>16</sup> en SQL.

Permite también registrar diversos tipos de orígenes de datos, ya sean archivos planos, bases de datos oracle y bases de datos no oracle. Los procesos de extracción y transformación de datos son diseñados gráficamente en forma simple y luego son traducidos a sentencias DDL por la herramienta. *OWB* genera solo sentencias SQL (DDL) y PL/SQL<sup>17</sup> (DML<sup>18</sup>), que son lenguajes propios del motor, no es necesario ningún tipo de compilador externo para la ejecución de estos comandos.

Todos los procesos de carga son traducidos a procedimientos almacenados PL/SQL y si existen procedimientos ya construidos, tiene la capacidad de reusarlos. *OWB* permite registrar estos procedimientos e incluirlos en la lógica de los procesos de carga. Dentro del *OWB* también se pueden definir los Flujos de Procesos que describen la dependencia entre mapeos y actividades externas tales como email, FTP y operaciones de Sistema Operativo. El código asociado a estas operaciones se crea cuando se hace el despliegue del modelo sobre la Base de Datos destino. El código es PL/SQL o comandos de SQL\*Loader, etc., dependiendo de lo que se este desplegando o implementando.

Además de la herramienta ETL, Oracle Warehouse Builder tiene la funcionalidad de implementar dimensiones que soporten el almacenamiento del histórico de cambios en una dimensión. De esta manera puede implementar la metodología de dimensiones de cambio lento, clasificadas en tres tipos: [Nariño, 2007]

Tipo 1: Sobrescribe los cambios de los valores de las dimensiones, no guardando su historia.

Tipo 2: Se crea un nuevo registro en la dimensión. Esta alternativa mantiene el histórico de cambios de un valor de dimensión utilizando una versión actual del registro y se conservan las versiones anteriores.

---

<sup>16</sup> Del ingles "Data Definition Language" con las siglas DDL. Permite a los usuarios definir las estructuras que almacenarán los datos, así como los procedimientos y funciones que permitan consultarlos.

<sup>17</sup> Lenguaje de programación que soporta consultas y manipulación de datos en los SGBD Oracle y PostgreSQL.

<sup>18</sup> Del ingles "Data Manipulation Language" con las siglas DML. Lenguaje proporcionado por los sistemas gestores de bases de datos, que le permite a los usuarios llevar tareas de consultas a la base de datos.

Tipo 3: Agrega los cambios en el mismo registro. Esta alternativa asegura que la dimensión tiene la estructura suficiente para en un mismo registro mantener el nuevo valor y los valores anteriores de la dimensión.

### 3.3.1 *OWB* versus procesos manuales de programación.

Los beneficios de emplear *OWB* versus procesos manuales de programación pueden ser vistos bajo la siguiente perspectiva: [Nariño, 2007]

- *OWB* utiliza las características de ETL de la BD Oracle haciendo innecesario un motor externo de ETL.
- Cada mejora incorporada en la BD es empleada por *OWB* e incorporada en el código generado. Por ejemplo, el comando MERGE, simplifica la ejecución por separado de comandos Update (Actualizar) e Insert (Insertar). Esta modificación es hecha por la herramienta, mientras que en un proceso manual significaría mucha inversión de tiempo identificando donde realizar esta modificación.
- Bajo costo de mantenimiento pues todo está en un solo ambiente tecnológico.
- Mejor uso de los recursos de Hardware y Software, porque el código generado esta optimizado y depurado.

La siguiente tabla resume aspectos esenciales de la comparación:

Tarea	Ventajas de usar <i>OWB</i> vs código manual.
Diseño/Código Inicial	Ganancia de tiempo a través del modelado lógico en vez de escribir directamente cientos de líneas de código.
Manejo del Error	Warehouse Builder incluye el manejo de errores.
Depurador	Se logra reducir tiempo y esfuerzo usando el Depurador gráfico de <i>OWB</i> .
Aplicar cambios de Ejecución	Mucho más rápido debido a los cambios lógicos en vez de los cambios independientes de varios lugares por Separado.

<b>Creación del EUL</b>	Automatizado usando Warehouse Builder, masivamente reduce el tiempo.
<b>Llevar a Producción</b>	Más simple y manejable con Warehouse Builder, reduce errores y repeticiones.
<b>Infraestructura de Administración</b>	Incorporada a la herramienta, no se requiere labor para instalarla y hacerla funcionar.
<b>Documentación</b>	Donde la mayoría del código de las herramientas de ETL es no documentada, al trabajar con Warehouse Builder directamente se documenta a medida que se diseña gráficamente.
<b>Balance final</b>	Reduce el esfuerzo y disminuye el tiempo de desarrollo.

Warehouse Builder reduce el tiempo de diseño a través de modelos visuales, asistentes gráficos de ayudas y librerías de transformaciones predefinidas. Adicionalmente esta herramienta permite generar el código a partir del diseño creado, el cual es libre de error y puede ser fácilmente modificado y actualizado.

Esta herramienta provee muchos asistentes y librerías de transformaciones que ayudan a la generación de código y a realizar el diseño del ETL de una forma fácil sin necesidad de generar código, sin embargo si se quisiera modificar ese código, el lenguaje utilizado sería PL-SQL. Además, el OWB tiene la capacidad de manejo y control de versión, en cualquier punto o fase del diseño, pudiendo capturar la imagen del objeto diseñado en un momento determinado. Esta imagen es almacenada en la base de datos y puede ser utilizada para propósitos de comparación con otras versiones o para restaurar en caso de cambios indeseados. Este proceso se realiza a través de una interfaz de usuario llamada Administrador de Cambios.

Oracle Warehouse Builder trabaja con un repositorio de metadata el cual puede ser exportado de un repositorio e importado en otro; también permite copiar parcialmente o completamente un proyecto en otro proyecto que maneje un único repositorio.

También tiene la capacidad de hacer depuración en tiempo de ejecución a través del editor de mapeo visual, en el cual el usuario puede definir cuáles son las tablas que desea depurar o crear, y hacer el

mapeo paso por paso; así mismo provee los puntos de seguimiento y la visualización de variables en tiempo de ejecución. Los mensajes de depuración proveen la información acerca del estado y cualquier error ocurrido durante la ejecución.

Cuenta con un componente llamado *Oracle Change Management Pack* el cual permite administrar, implementar, investigar y mantener una traza de los cambios efectuados en los objetos de la base de datos Oracle. Utilizando *Oracle Change Management Pack* se puede: [Nariño, 2007]

- a. Capturar y almacenar las definiciones de objetos.
- b. Comparar las definiciones de objetos y sus diferencias.
- c. Sincronizar definiciones de objetos.
- d. Especificar e implementar cambios complejos de definiciones de objetos.
- e. Deshacer cambios y volver al estado original.
- f. Propagar definiciones de objetos a unos o más puntos.

Cuando hablamos de cambios o actualizaciones de datos, entonces podemos emplear en *OWB* una funcionalidad de la Base de Datos 10g llamada *Change Data Capture*.

1. Análisis de dependencias (evaluación de impacto de cambios). Oracle Warehouse Builder tiene la capacidad de generar un reporte gráfico que permite analizar el impacto de los cambios efectuados tanto en los objetos fuentes como en los objetos destinos, y así determinar si se implementan o no los cambios efectuados.
2. Especificaciones de las funciones de ETL usando objetos de transformación preconstruidos, accesibles por medio de una interfaz de usuario intuitiva. Los diseñadores pueden crear procesos complejos combinando todas las representaciones gráficas y generando código basado en filas o en grupo de filas. Si el usuario ya tiene un conjunto de transformaciones que han ido construidas con PL/SQL, estas pueden ser importadas o definidas como procedimientos almacenados, funciones o paquetes para ser reutilizadas dentro de un proyecto o a lo largo de varios proyectos.
3. Agregación y cálculo incremental de agregados por la herramienta de ETL en un paso, desde los datos de las fuentes. Oracle Warehouse Builder maneja en su ambiente gráfico una serie de

iconos que permiten realizar el mapeo y transformaciones necesarias para llevar datos de la fuente al destino.

4. Soporte automatizado de dimensiones de cambio lento. (Tipo I, Tipo II y Tipo III). Las dimensiones de cambio lento son estrategias bien definidas para administrar los datos actuales e históricos a lo largo de un periodo de tiempo en un data warehouse. Es considerado e implementado como una de las tareas más críticas de ETL en rastrear la historia de los registros de una dimensión.

Los usuarios tienen que decidir cual de los tipos de cambio lento son los más adecuados para sus necesidades. Una vez que un tipo es seleccionado, los usuarios deben: [Nariño, 2007]

1. Crear una Dimensión que mantenga la historia
2. Crear un mapeo que permita extraer los datos del sistema fuente, lo transforma y lo carga en la dimensión objetivo predefinido.
3. Generar y desplegar tanto la dimensión como el mapeo en la BD Oracle.
4. Ejecutar el Mapeo.

Oracle Warehouse Builder permite a los usuarios extraer información de diferentes fuentes como: [Nariño, 2007]

- Archivos planos
- Oracle RDBMS, versión 7.3.4 en adelante
- SAP/R3
- Documentos XML
- Otras bases de datos: DB2, Sybase, Informix, SQL Server a través de Oracle.

Así mismo los datos a extraer y la parte temporal del proceso ETL (carga inicial o de mantenimiento) también impactarán en la decisión de cómo extraer los datos desde una perspectiva lógica o física.

Para la extracción lógica de los datos, existen dos métodos:

- Extracción completa: Extracción o copia completa de los datos fuentes mapeados a la parte temporal de la BD destino.

- Extracción incremental: Solamente son copiados los registros alterados o nuevos.

Por otra parte, en la extracción física de los datos se tienen otros métodos. Ellos son:

- Extracción en línea: Los datos son extraídos directamente del sistema fuente conectándose en forma directa a las tablas fuentes o a tablas especialmente creadas para el proceso de extracción.
- Extracción fuera de línea: Los datos no son extraídos directamente del sistema fuente, es más bien colocada en un área fuera del sistema fuente. Los datos poseen ya una estructura o fueron creados por una rutina de extracción.

De esta manera se ve lo conocido como "*Data Warehousing en Tiempo Real*" o "*Data Warehousing en Tiempo*" lo cual provee una solución para el mantenimiento y actualización del Data Warehouse según se requiera.

### **3.3.2 Herramientas de Limpieza de Datos.**

*Oracle Warehouse Builder* provee varias herramientas para la detección y corrección de problemas en los datos. Además de las incorporadas, también se permite al desarrollador crear sus propias funciones o procedimientos.

Las funciones incorporadas en *OWB* son:

- Perfiles de datos.
- Reglas en los datos derivadas de los perfiles o ingresadas manualmente.
- Auditores de datos basados en las reglas.
- Limpieza de Nombres y Direcciones.

Los perfiles de datos analizan los atributos de la información, los análisis referenciales y/o realizan un análisis personalizado para "perfilar" la calidad de los datos. Permite seleccionar una estrategia de limpieza (Eliminar, Comparar, Evaluar y Personalizar). [Nariño, 2007]

Los auditores de los datos son actividades de procesamiento de flujo que evalúan una o más reglas sobre un objeto determinado (ejemplo una tabla). La evaluación de estos auditores puede ser simple como un si/no o incluir límites de tolerancia expresados en porcentaje de errores. Se puede controlar los subsecuentes procesos (email, notificación, abortar, etc.) y almacenar la medida de calidad.

Para mantener un buen trabajo en los procesos de ETL es necesario realizar un conjunto de tareas que incluyen las siguientes funcionalidades:

1. Filtrado de datos, convertir códigos y calcular valores derivados. Warehouse Builder maneja estas funcionalidades a través de operadores funcionales incorporadas. Filtro, Función de transformación (Caracteres, Aritméticos, Fechas, etc.).
2. Validar datos para chequear contenido y rango de los valores de los campos. El manejador de Expresiones permitiría realizar las tareas definidas.
3. Realizar funciones de limpieza de datos. Este operador denominado "Nombre y Dirección" permite identificar y corregir los errores e inconsistencias en los datos fuentes Nombre y Dirección.
4. Cargar la información limpia, a un Datamart o Data Warehouse. OWB posee operadores para crear objetos como Tablas o Dimensiones que pueden ser pobladas usando operadores como *Splitter* o *Join* que luego de un proceso de limpieza de datos pueden redirigir el resultado a un Datamart o Data warehouse objetivo.
5. Reutilización de los procedimientos almacenados. OWB puede importar al repositorio códigos PL/SQL, Funciones, Procedimientos o Paquetes.
6. Generación automática de números de secuencia. La generación automática de números de secuencia se maneja a través del operador de secuencias. [Nariño, 2007]

### **3.3.3 Administración.**

OWB posee las herramientas para programar, coordinar y ejecutar todos los pasos para poblar un almacén de forma óptima. Estas herramientas se pueden manejar dentro de Warehouse Builder, o pueden ser programadas en el *Oracle Enterprise Manager* (OEM). El principal propósito del OWB es realizar las tareas de ETL, aprovechando las características de la BD de Oracle para Data Warehousing. Entre ellas están el particionamiento de los datos, procesos en paralelo, índices, uniones, etc.

El *Warehouse Builder* se instala con el *Oracle Universal Installer*, una interfaz gráfica que presenta un asistente que guía durante todo el proceso de instalación. Y con él se puede mantener el monitoreo y la administración de la ejecución en tiempo real de los datos que se realiza con el *Runtime Audit Viewer*. El dar soporte para la prioridad y secuencia a procesos y sesiones, por medio de una interfaz gráfica. Esto se realiza mediante la definición de un flujo de trabajo dentro de *Warehouse Builder*.

Los reportes operacionales y de auditoría que se realizan por cada carga de datos son obtenidos desde el repositorio instalado en la base de datos destino "Ambiente de Ejecución", que incluye: [Nariño, 2007]

1. Reportes del repositorio.
2. Reportes de los procesos.
3. Reportes de los procesos en ejecución.
4. Reportes de los mapeos.
5. Reportes de los mapeos en ejecución.
6. Reportes de errores de ejecución.
7. Reportes de los datos de los objetos.
8. Reportes de localización.
9. Reportes de errores de despliegue.
10. Administrar Reportes

La Metadata del *Warehouse Builder* por definición es centralizada y los programas de extracción de datos son generados de manera automática luego de realizado el mapeo (*Mapping*) en el *OWB*, que posee un componente para hacer depuraciones de los mapeos definidos con anterioridad. Se puede ir paso a paso e identificar cuál de las partes en los mapeos es la que determina el fallo.

Para la generación extensa de reportes de resultados de una sesión de ETL, incluyendo notificación automática de fallas significantes en los procesos de ETL, se tiene revisando la Metadata del *Warehouse Builder* que contiene información de los resultados de las ejecuciones de los procesos de ETL. Es posible ver el resultado de estos procesos directamente usando el cliente de *OWB* o también por un Navegador (*Browser*). Así mismo en caso de falla de un mapeo, se puede realizar otras actividades. Para este propósito se debe definir un flujo de tareas.

El cliente de *OWB* tiene la capacidad de programar sesiones de FTP basados en tiempo o eventos, es decir, traspasa vía FTP archivos planos remotos al servidor de aplicaciones: UNIX o NT. Las instituciones en general no permiten que, por temas de Seguridad, se acceda directamente a sus bases de datos. Es por eso que el traspaso de archivos vía FTP es una alternativa para obtener los cambios de las fuentes y actualizar el almacén.

Con esta herramienta se puede acceder a datos de múltiples fuentes de datos operacionales: Interfaz nativa con Oracle, interfaz con archivos de texto, fuentes XML, otras. Warehouse Builder 10g, soporta como fuentes de datos (de manera nativa) *Oracle Database 10g*, Oracle9i Versión 2 (9.2.x), Oracle8i Versión 3 (8.1.7.4.x), así como bases de datos heterogéneas como DB2, Sybase, Informix, SQL Server (vía *Oracle Transparent Gateways*) y Mainframe. [Nariño, 2007]

### **3.3.4 Plataforma.**

*Oracle Warehouse Builder* posee varios componentes: Uno sobre el Cliente (El cliente de diseño, el Navegador de Metadata y Auditoria) y el otro sobre el Servidor (Los componentes de Ejecución como objetos en la BD destino, el repositorio, Servicio de Acceso al repositorio, entre otras cosas), y trabaja tanto en ambientes Unix como Windows, donde cada uno de sus componentes puede ser desplegado indistintamente en cada sistema operativo por lo que no representa problema el “mover” el producto de un ambiente a otro.

El código PL/SQL generado por *OWB* aprovecha las bondades de *DWH* de la BD Oracle. EL procesamiento en paralelo es una de las funcionalidades que es incluida en el código para ser ejecutado en la BD. Por otra parte, el *Oracle Warehouse Builder* maneja el control de las fuentes de datos a través de denominados Módulos. Los Módulos pueden ser Bases de Datos relaciones, o archivos planos. Al identificarse una nueva fuente de datos simplemente se necesitaría agregar un nuevo módulo fuente (ya sea de base de datos relacional o archivo plano). [Nariño, 2007]

### **3.3.5 Diseño y modelado.**

Los modelos lógicos son creados en Warehouse Builder a través de la interfaz gráfica Java. Permite crear Mapeos, Dimensiones, Cubos, Vistas, Tablas, Vistas Materializadas, Tablas Externas, Colas Avanzadas, Secuencias, etc. *OWB* provee asistentes en todo momento para la creación de estos objetos.

Los índices se pueden crear usando la interfaz del *OWB* Editor. El código asociado a estas operaciones se crea cuando se hace el despliegue de modelo sobre la BD destino. El código es PL/SQL o comandos de SQL\*Loader, etc., dependiendo de lo que se manda a generar y desplegar e implantar.

### **3.3.6 Rendimiento.**

El código que se obtiene del *OWB*, puede ser ejecutado sobre plataformas UNIX o Windows. El código creado es PL/SQL que debe ser ejecutado en la Base de Datos Oracle, lo cual hace indiferente que sea Unix o Windows la plataforma. Otra característica relacionada con el rendimiento es que todo el proceso de ETL con el *OWB* se realiza a través de una interfaz gráfica (cliente Java), que presenta una serie de operadores que agrupados forman un mapeo para la transformación intermedia o una carga a una tabla o dimensión.

## **3.4 Oracle OLAP.**

Las herramientas de generación de informes y consultas dinámicas generalmente se despliegan utilizando Data Warehouse y herramientas basadas en SQL, mientras que aplicaciones más sofisticadas de análisis y planificación se despliegan utilizando otras bases de datos y herramientas.

Esta separación entre Data Warehouse y bases de datos analíticas es, sin embargo, costosa. El costo de mantenimiento de bases de datos separadas supone recursos adicionales hardware y administradores. Oracle elimina estas barreras proporcionando una base de datos preparada para el análisis. Oracle OLAP, parte de la base de datos Oracle, proporciona la funcionalidad analítica de una base de datos especializada a la vez que elimina la necesidad de gestionar bases de datos separadas.

Oracle OLAP es parte de una plataforma integrada de data warehousing e inteligencia organizacional que contiene funcionalidades ETL, OLAP y minería de datos en la base de datos. Hay muchos beneficios

derivados de tener una base de datos Relacional- Multidimensional completamente integrada en comparación con bases de datos multidimensionales y separados:

- Gestión más simple.
- Alta disponibilidad.
- Seguridad mejorada.
- Acceso abierto desde clientes SQL.
- Ciclo de información reducido.
- Mayor fiabilidad de la información.

Resulta típico, al manejar temas de seguridad, el tener que realizar consultas complejas, donde se requiere recuperar el valor de un atributo mas de una vez como consecuencia de querer analizar su variación en el tiempo, como por ejemplo:

“Recuperar la cantidad de delitos de tipo HOMICIDIO por estados y parroquias, en cada semana del año actual”

Este tipo de consulta queda optimizada cuando se emplea el modelo multidimensional de datos, pues solo se relacionan las tablas de hechos con las dimensiones definidas, que poseen datos precalculados con anterioridad, en la que se puede aumentar incluso la complejidad de las consultas. En este caso, entra en el análisis la dimensión Temporal que en el ejemplo anterior dificultaría grandemente la confección y el tiempo de respuesta de la consulta y que sin embargo facilita la redacción de la misma consulta sobre el esquema estrella del almacén:

```
SELECT de.estado, de.parroquia, dt.anno, dt.mes, COUNT (*) as Total_Homicidios
FROM DIM_Temporal dt, DIM_Espacial de, DIM_Victima dv, FACT_Hecho fh
WHERE dt.IdTemporal = fh.IdTemporal and de.IdEspacial = fh.IdEspacial and
        dv.IdVictima = fh.IdVictima and dv.sexo = 'M' and fh.TipoHecho = 'Homicidio'
GROUP BY de.estado, de.parroquia, dt.anno, dt.mes;
```

### **3.4.1 Funcionalidad de Oracle OLAP.**

La base de datos relacional-multidimensional Oracle se compone de las siguientes tecnologías: [Nariño, 2007]

- La tecnología relacional gestiona los objetos de base de datos y proporciona una interfaz SQL a los datos.
- La tecnología de objetos permite a la base de datos gestionar datos no relacionales.
- La tecnología OLAP proporciona funcionalidad analítica avanzada en el contexto de un modelo multidimensional.

### **3.4.2 Proceso de Base de Datos Integrado.**

Todas las tecnologías (relacional, objetos-relacional y OLAP) están contenidas en un solo proceso; la base de datos se trata como una sola instancia integrada. Como la tecnología OLAP es parte de un solo proceso de base de datos, proporciona la misma escalabilidad y fiabilidad que el proceso de base de datos Oracle.

### **3.4.3 Almacenamiento.**

En las BD Oracle todos los datos - relacional y multidimensional - se almacenan en los ficheros de datos. No hay ficheros multidimensionales separados que debamos gestionar. Los tipos de datos relacional y multidimensional pueden coexistir en los mismos ficheros de datos.

Los datos multidimensionales se almacenan en la base de datos en lo que se llaman áreas de trabajo analíticas (*analytic workspaces*). Oracle OLAP representa un cambio fundamental en la tecnología OLAP. Las bases de datos Relacionales y Multidimensionales se han unido para proporcionar capacidades analíticas propias de una base de datos Multidimensional en el contexto de la base de datos Oracle.

Las ventajas de un Sistema de Bases de Datos Relacionales – Multidimensionales para Soporte de Decisiones son: [Nariño, 2007]

- Gestión más simple

- Proporciona la alta disponibilidad de un servidor corporativo
- Mayor seguridad
- Ofrece el acceso más abierto
- El ciclo de información se reduce
- Se elimina la sincronización de los datos

Todas estas ventajas se obtienen a la vez que se mantiene la potencia de un servidor OLAP dedicado. Oracle OLAP proporciona un conjunto de funciones analíticas a través de la OLAP API<sup>19</sup>, un motor multidimensional y un lenguaje de manipulación de datos OLAP. El resultado es un sistema de menor costo, más seguro y fiable, que da soporte a aplicaciones analíticas, de consulta y de generación de informes.

### 3.5 Oracle Spatial.

*Oracle Spatial*, es una opción de la base de datos de *Oracle Database Enterprise Edition*. Complementa a *Oracle Locator*, y proporciona una robusta base para Sistemas de Información Geográfica<sup>20</sup> complejas, que requieren más análisis espaciales y mayor capacidad de procesamiento de la base de datos de Oracle. Esto incluye funciones espaciales (área, buffer, cálculos del centroide), soporte para sistemas de coordenadas avanzados, sistemas de referencia lineal y funciones agregadas. En *Oracle Database 10g*, la opción *Oracle Spatial* incluye nuevas capacidades importantes que cubren los requerimientos críticos del sector público, defensa, logística, exploración de energía y dominios geográficos extendiendo el poder de la base de datos geoespacial.

Esta herramienta permite asociar el almacenamiento de información de tipo “medidas” a una geometría lineal. Las funciones de agregado espaciales operan sobre un conjunto de geometrías en lugar de sobre una o dos geometrías. Una función de agregado ejecuta una operación de agregado específica sobre un conjunto de geometrías entrantes y devuelve un objeto geométrico individual.

---

<sup>19</sup> Librería utilizada por Oracle para interactuar con estructuras multidimensionales.

<sup>20</sup> Del inglés “*Geographic Information System*” con las siglas GIS, significa Sistemas de Información Geográfica.

Un nuevo tipo de datos gestiona nativamente imágenes geo-referenciadas (imágenes por satélite, datos enviados remotamente, entre otros) en Oracle Database 10g. La característica *GeoRaster* de *Oracle Spatial* proporciona georeferenciación de imágenes; cuestión que es de vital importancia para poder enfocar geográficamente cualquier estrategia de intervención o de prevención de delitos.

Se proporciona un modelo de datos para almacenar estructuras de red (gráficos) en Oracle Database 10g. Explícitamente almacena y mantiene la conectividad de las redes de nodos de enlace y proporciona a la red capacidad de análisis como el camino más corto y análisis de conectividad.

*Oracle Spatial* incluye un esquema y modelo de datos que almacena de manera persistente una topología en la base de datos de Oracle. Esto es muy útil en aplicaciones de seguridad ciudadana donde existe un alto grado de desagregación de características y necesidad de mantener la integridad de los datos en los mapas y sus capas, en particular esta herramienta es de gran utilidad para el Data Warehouse del CTAISC, porque permite almacenar información georeferenciada que puede ser utilizada en la construcción de mapas temáticos.

### **3.5.1 Geocoder.**

Con la opción de *Oracle Spatial* de *Oracle Database 10g* se proporciona un completo gestor de geocodificación. Proporciona estandarización de direcciones internacionales, geocodificación y coincidencia de POI mediante consultas de datos geocodificados almacenados en bases de datos de Oracle. La dirección analizada única soporta una gran flexibilidad y conveniencia a la aplicación de los clientes.

### **3.5.2 Operadores espaciales.**

La interacción de varias características geométricas puede determinarse a través del uso de operadores de comparación. Esto permite contestar a consultas del tipo de "lista todos los robos de vehículos en determinada autopista" o "encontrar todos los puntos de ventas de drogas dentro de una determinada área de interés".

*Locator* también proporciona una función que calcula distancias entre dos objetos geométricos. Esto es muy útil para consultas de servicios de localización como, “consulta de los 10 lugares con más riesgo de homicidios que estén más próximos al aeropuerto, y distancia en kilómetros al aeropuerto de cada uno de ellos” o “dado un determinado robo, saber cuantos robos de ese tipo han ocurrido en un radio de 5 km cercano al delito”.

### **3.5.3 Consultas espaciales paralelas.**

Las consultas espaciales pueden ejecutarse en paralelo, mejorando el rendimiento de consultas tipo “dentro de la distancia”, “delito más cercano” y “relación”. El rendimiento es directamente proporcional con el número de CPU's utilizados para ejecutar la consulta.

## **3.6 Real Application Cluster (RAC)**

*Oracle Real Application Clusters (RAC)* provee una altas disponibilidad, consolidación y escalabilidad. Este se ejecuta sobre un grupo de *Hardware (Cluster)* que no es más que un grupo de servidores independientes (nodos) que cooperan como un sistema único. Los componentes de agrupamiento principales son los nodos del procesador, una interconexión de agrupamiento y un subsistema de almacenamiento compartido.

Los nodos comparten el acceso al subsistema de almacenamiento y los recursos que administran los datos, pero no comparten físicamente la memoria. RAC combina la memoria en los nodos individuales para proporcionar una visualización única de la memoria caché distribuida en todo el sistema de la base de datos. [Nariño, 2007]

Un nodo puede estar compuesto de varios procesadores. Cada nodo tiene su propio sistema operativo, instancia de base de datos y software de aplicación. Es una solución que posibilita que un sistema funcione cuando algunos de sus componentes fallan, como sus instancias, los nodos físicos o la interconexión de agrupamiento.

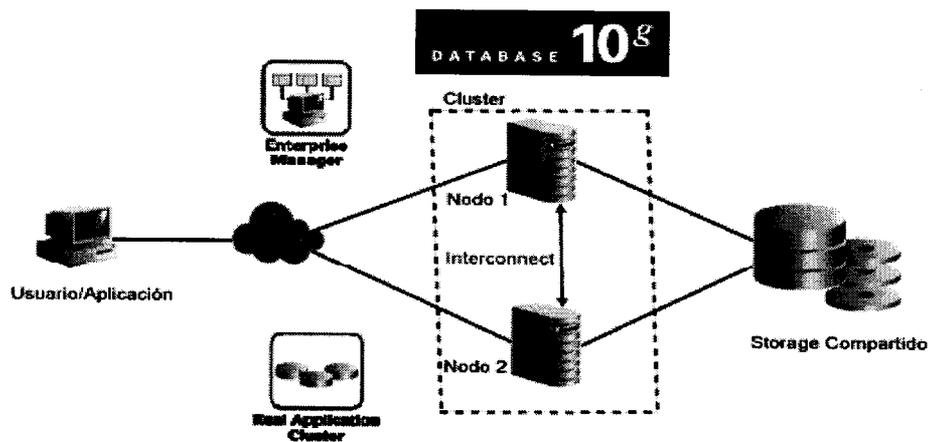


Figura 3.4 Real Application Cluster. [Nariño, 2007]

### 3.6.1 Costos de hardware.

Permite utilizar hardware de bajo costo para administrar grandes bases de datos en lugar de adquirir nuevo y costoso hardware. Se pueden confeccionar clusters de pequeños servidores, por ejemplo, se puede optar por comprar 4 servidores de 2 procesadores cada uno en lugar de comprar uno solo de 8 procesadores. Los pequeños servidores estándares del mercado son mucho más económicos que uno solo de gran capacidad.

### 3.6.2 Flexibilidad.

Permite adaptar fácilmente la infraestructura de hardware, para la base de datos. Es posible agregar nuevo hardware al cluster sin afectar las aplicaciones que se encuentran corriendo, e incorpora inmediatamente la nueva capacidad de procesamiento instalada. De igual manera el hardware puede ser removido del cluster sin afectar las aplicaciones que utilizan la base de datos y además tolera fallas de los servidores sin que las aplicaciones dejen de correr. [Nariño, 2007]

### 3.6.3 Disponibilidad

Los sistemas que implementan RAC pueden ser configurados de forma tal que no tengan un solo punto de falla; manteniendo la disponibilidad de las aplicaciones de base de datos. Las fallas del servidor principalmente son una ocurrencia común en los centros de datos, y el desafío de sus administradores es recuperarse de las fallas tan pronto como sea posible. Esto permite reducir los costos por fallas de los sistemas y mantenimiento. [Nariño, 2007]

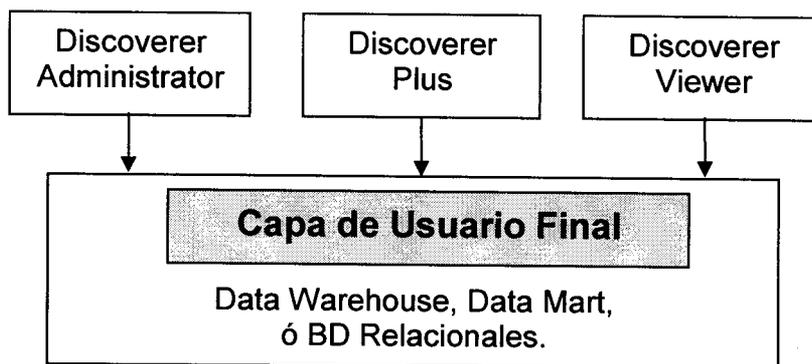
### 3.7 Oracle Discoverer.

Discoverer es una herramienta orientada a temáticas que se usa para construir consultas a la medida usando datos de un área específica, que son de interés para un cierto grupo de usuarios. Discoverer también facilita el manejo de la información desde datos sumariados hasta datos muy detallados.

#### 3.7.1 Arquitectura del Oracle Discoverer

Discoverer mantiene soporte de bases de datos heterogéneas y está compuesto arquitectónicamente por los siguientes módulos: [Terán, 2006]

- **Discoverer Administrator:** Este producto sirve para crear, mantener y administrar datos en la capa de usuario final (EUL) y para definir la forma en que los usuarios interactúan con los datos.
- **Discoverer Plus:** Cuenta con una interfaz Web que puede ser consultada usando un navegador web. Cada usuario final, puede usar este componente para crear, visualizar y editar los libros de trabajo sobre internet o intranet.
- **Discoverer Viewer:** este componente es útil para visualizar los libros de trabajo y hojas de trabajo usando un navegador web. Los libros de trabajo creados en *Discoverer Plus* pueden ser abiertos, pero no pueden ser editados usando Viewer. Discoverer Viewer hace la información disponible en Internet o intranet y permitiendo que la información sea incluida en páginas Web o accesadas desde portales corporativos.
- **End User Layer:** este componente es una capa de metadatos que oculta la complejidad de la base de datos, permitiendo que los usuarios puedan interactuar con la base de datos sin tener que reconocer comandos de la misma. Esto permite a los usuarios utilizar el lenguaje de negocio con el que está más familiarizado.



**Figura 3.5 Arquitectura de Oracle Discoverer. [Terán, 2006]**

### 3.7.2 Publicación de Reportes

Oracle Discoverer permite compartir y publicar los informes en una gran variedad de formatos, incluyendo *HTML*, *ASCII*, *Excel* y correo electrónico para distribuir a un gran número de usuarios.

La exportación a *Excel* incluye la creación de cabeceras y el formateo de los datos, así como las sumas totales. Además los resultados de una matriz son exportados como tablas pivote facilitando el manejo de los resultados para distribuirlos a un grupo de personas.

Los usuarios pueden, sobre las consultas que realicen, elegir funciones analíticas establecidas por el administrador o definir las propias. Esta herramienta proporciona la posibilidad de crear funciones analíticas anidadas y soporta funcionalidades de *Oracle Database 10g* tales como creación de ranking, medias móviles, comparativas y tendencias lineales.

### 3.7.3 Rendimiento

Ofrece un elevado rendimiento por su empleo de *memoria caché*, paginación de documentos y monitorización de la actividad. Predice los tiempos de respuesta de cada consulta, informa al usuario y controla en todo momento el mejor uso de los recursos. Optimiza el tiempo de respuesta redirigiendo las consultas a tablas previamente sumarizadas. Los libros de trabajo de *Discoverer* se pueden incluir dentro de páginas web existentes, especificando las *URL's*. [Terán, 2006]

### 3.7.4 Integración con la Base de Datos

Discoverer tiene un alto nivel de integración con la base de datos, lo que simplifica la seguridad, la escalabilidad y el acceso a los datos. Así mismo, también está integrado con *OracleAS Reports Services*.

#### Conclusiones del Capítulo.

Se puede concluir que:

- Oracle implementa servicios OLAP para almacenar estructuras multidimensionales que facilitan la ejecución de consultas complejas, que se requieren sistemáticamente en temas de seguridad ciudadana.
- Oracle ofrece una solución completa a los procesos del Data Warehouse: ETL, almacenamiento y visualización.
- Para definir los procesos de ETL se usarán herramientas que automaticen la limpieza de los datos a partir de las fuentes y el refrescado a partir de los cambios.
- Las herramientas de la opción Oracle Spatial permiten almacenar información georeferenciada, que puede ser utilizada en la construcción de mapas temáticos, que son absolutamente necesarios a la hora de manejar estadísticas de temas de seguridad.
- Oracle Discoverer proporciona una interfaz amigable para la explotación de estructuras multidimensionales (cubos OLAP).
- La arquitectura de software propuesta se ajusta a las necesidades del Data Warehouse del CTAISC.

## **Conclusiones.**

Como conclusiones del presente Trabajo de Diploma se plantean las siguientes:

1. Se caracterizó el estado del arte en el tratamiento de información sobre seguridad ciudadana en el mundo, donde se muestran ejemplos de sistemas informáticos utilizados donde se trabajan estos temas directa o indirectamente.
2. Se logró conceptualizar las necesidades de información para la versión inicial de la población del almacén de datos así como se logró referirlas a fuentes de datos conocidas de los Órganos de Seguridad en Venezuela.
3. Se propuso un modelo y las herramientas para su implementación informática, que resuelve el problema científico, en particular el problema de la seguridad ciudadana en la Republica Bolivariana de Venezuela.

## **Recomendaciones.**

Para el trabajo futuro, se tienen las siguientes recomendaciones:

1. Implementar el modelo propuesto para el Data Warehouse del CTAISC, considerando las extensiones que surjan durante las especificaciones detalladas de los requisitos.
2. Concebir un conjunto de Data Marts para cada área de negocio relacionada con el tratamiento de la información sobre la seguridad ciudadana.
3. Diseñar los reportes sobre las herramientas de visualización adquiridas.

## Referencias Bibliográficas.

[Action Ware S.L, 2004] Action Ware S.L. (2004). CUBOS MULTIDIMENSIONALES. Disponible en: <http://www.actionware.es/cubo.htm> . Consultada el [13 de junio del 2007].

[ALBET, 2006] Proyecto Técnico “CREACIÓN DEL CENTRO DE TRATAMIENTO Y ANÁLISIS DE INFORMACIÓN DE SEGURIDAD CIUDADANA (CTAISC) Y METODOLOGÍA PARA LA IMPLANTACIÓN DE CENTROS DE EMERGENCIAS 171 EN LA REPÚBLICA BOLIVARIANA DE VENEZUELA”. ALBET (2006).

[Aizen, 1997] Aizen, Marina (1997). “Las enseñanzas de mister Giuliani” Disponible en: <http://www.ser2000.org.ar/protect/Autores/comstat.htm>

[Arquilla & Ronfeldt, 2003] Arquilla, J. y Ronfeldt, D. (2003): Redes y guerra en red. El futuro del terrorismo, el crimen organizado y el activismo político, Madrid, Alianza.

[Asamblea Nacional Constituyente, 2000] Asamblea Nacional Constituyente (2000). Constitución de la República Bolivariana de Venezuela. Caracas, Venezuela.

[Clements, 1990] Clements, Kevin (1990), “Toward a Sociology of Security”, Conflict Research Consortium, University of Colorado. Working Paper 90-4, July, 22 pp.

[Corey, 1998] Corey, M. J. (1998). Oracle Data Warehousing. John Wiley & Sons, Inc.

[Enriquez,Martínez,...,2007] Enríquez, Juan Á., Martínez, Tomás,... (2007). “Funcionalidad 4: Almacenes de Datos y Bases de Datos XML”. Universidad de Castilla- La Mancha. Escuela Superior de Informática.

[García, 2006] Lourdes García. (2006). Bases de Datos y Data Warehousing, Universidad de La Habana.

[García, 2007] Ana M. García Pérez (2007). Procesos Elementales de Negocio. Centro de Tratamiento y Análisis de la Información de Seguridad Ciudadana Venezuela.

[Giménez, 2003] Matilde Celma Giménez (2003). Almacenes de Datos (Data Warehouse). Departamento de Sistemas Informáticos y Computación. Universidad Politécnica de Valencia. España.

[Hobbs, Hillson & Lawande, 2005] Liliam Hobbs, Susan Hillson, Shilpa Lawande, Pete Smith (2005). Oracle Database 10g Data Warehousing. John Wiley Publishing, Inc. USA

[Hraba,1998] Hraba, Joseph et al. (1998), "Perceived Risk of Crime in the Czech Republic", Journal of Research in Crime and Delinquency, vol. 35, nº 2, pp. 225-242.

[Imhoff, 2003] Imhoff, C., N. Galemmo, et al. (2003). Mastering Data Warehouse Design. Indiana, John Wiley Publishing, Inc.

[Inmon, 2002] W. H. Inmon (2002). Building Data Warehouse, Third Edition. USA. John Wiley & Sons, Inc.

[Inmon, 2005] W. H. Inmon (2005). Building Data Warehouse, Fourth Edition. USA. John Wiley & Sons, Inc.

[Instituto Nacional de Estadística, 2006] Instituto Nacional de Estadística. Encuesta de Victimización y Percepción Policial. Año 2006. Venezuela.

[Landáez, 2002] Nelly Landáez Arcaya (2002). Inseguridad ciudadana. Versión Resumida del Trabajo de Grado con el que se optó al Título Magíster en Ciencia Política Mención Gerencia Pública. Universidad Central de Venezuela

[Martínez, 2002] Valle, Marco A. (2002)- Coordinador Nacional Proyecto de Seguridad Ciudadana, Programa de las Naciones Unidas para el Desarrollo (PNUD), Nicaragua. Disponible en: <http://www.iigov.org/ss/article.drt>

[Montenegro, 2003] Montenegro, Germán (2003). Informe de Presentación –“Sistema de Recolección, Procesamiento y Análisis de la información Delictiva”. Subsecretaría de Planificación y Logística de Seguridad, del Ministerio de Seguridad de la Provincia de Buenos Aires. Argentina.

[Nader, 2003] Nader, Javier (2003). Tesis de Magíster en Ingeniería de Software. Sistema de Apoyo Gerencial Universitario. Argentina.

[Nariño, 2007] Nariño, John (2007). Presentación de Oracle para el Ministerio del Poder Popular para Relaciones Interiores y Justicia. Venezuela.

[Palomar & Trujillo, 2001-2002] Palomar, Manuel; Trujillo, Juan C. (2001-2002). Uso y Diseño de Bases de Datos Multidimensionales y Almacenes de Datos. Universidad de Alicante. España.

[Pence & Creeth, 2002] Pence, N, Creeth, R. "An Introduction to OLAP". Disponible en: <http://www.OLAPreport.com/>. [Consulta: Diciembre de 2006].

[Poe et al., 1998] Poe, Vidette; Klauer, Patricia; Brobst, Stephen 1998. "Building a Data Warehouse for Decision Support". Prentice Hall. ISBN: 0-13-769639-9.

[Ríos, 2006] Ríos, Wilfredo. Trabajo de Diploma presentado para optar por el título de Ingeniero en Ciencias Informáticas. Año 2006. Universidad de Ciencias Informáticas. pp:1-3

[Sansó-Rubert, 2006] Sansó-Rubert, Daniel. El papel de la información en la lucha contra la delincuencia organizada transnacional. UNISCI Discussion Papers, N° 12 (Octubre / October 2006).

[Segundo & Rozas, 2006] Segundo, A. and W. Rozas. (2006). "Business Intelligence: Mito o Realidad." Consultada: 17 de noviembre del 2006] Disponible en: <http://www.directivoscede.com/opiniones/detail.php?id=1167>

[Sperley, 1999] Sperley, Eric (1999). The Enterprise Data Warehouse: Planning, Building, and Implementation. O'Reilly & Associates, Inc.

[Tarre, 2006] Tarre, Marcos (2006) Notas de Seguridad. Especial para la Cámara de Comercio, Industria y Servicios de Caracas.

[Terán, 2006] Terán, Gustavo (2006). Maestría de Informática. Procuraduría General del Estado. Implementación de Data Warehouse con Oracle Discoverer.

[Thomé, 2004] Thomé, H.I. (2004). Victimización y cultura de la seguridad ciudadana en Europa. Tesis Doctoral para optar al título de Doctor en Sociología. Departamento de Sociología y Análisis de las Organizaciones. Universidad de Barcelona. 545 pp. España.

[Thomsen, 2002] Thomsen, Erick (2002) OLAP Solutions, Building Multidimensional Information Systems, Second Edition. Jhon Wiley & Sons, Inc. USA. .

[Valdivia, 2002] Claudio Valdivia R. (2002). Seguridad Ciudadana, Una tarea de todos – un manual para la acción. Instituto de Defensa Legal. Chile.

[Windom, 1995] Widom, J. (1995) Research Problems in Data Warehousing. Proc. of 4th Int'l Conference on Information and Knowledge Management (CIKM), Technical Report. Dept. of Computer Science.

[Zechinelli, 2001] Zechinelli, J. L. (2001).Data Warehouses. Fundación Universidad de las Américas, Puebla. Disponible en: <http://www.udlap.mx/~zechinell/>

## **Bibliografía.**

Ramos, S. (2006). Conferencia: Data Transformation Services y Analysis Services. Microsoft SQL Server 2000.

García, L. (2007). Conferencia: Bases de Datos y Data Warehousing. Departamento de Ciencia de la Computación. Facultad de Matemática y Computación. Universidad de La Habana.

Marotta, Adriana (1999). Conferencia: DWH Bedelia. InCo.

Humphries H. (2002). Data Warehousing Architecture and Implementation. O'Reilly & Associates, Inc. ISBN: 0-13-080902-0, 384 p.

Jones, M.E" (2002), Dimensional Modeling: Identifying, Classifying & Applying Patterns. Drexel University, College of Information Science & Technology, Philadelphia.

Serrano, Manuel (2001). Conferencia: Almacenamiento y Recuperación de la Información. UCLM.

Boza, Andrés (2004). Data Warehouse para la gestión por procesos en el sistema productivo. Second World Conference on POM and 15th Annual POM Conference, Cancun, Mexico.

## **Glosario de Términos.**

Conocimiento: Es el acervo de información utilizado en el proceso de la toma de decisiones.

Data Mart: Vista parcial de un almacén de datos.

Data Warehousing: El proceso de organizar la información en una forma que crea conocimiento basado en datos. Los productos de software que presentan este conocimiento a los usuarios se llaman Herramientas de Inteligencia de Negocios (Business Intelligence Tools).

Datos: Describen la realidad o percepción de la existencia humana, corporativa institucional, comunitaria, etc. Las computadoras almacenan y procesan datos. Al nivel más bajo los datos no tienen significado alguno.

Drill-down: Disgregar.

Escalabilidad: Es la propiedad deseable de un sistema, una red o un proceso, que indica su habilidad para manejar el crecimiento continuo de trabajo de manera fluida o para estar preparado para hacerse más grande sin perder calidad en los servicios ofrecidos. En general, también se podría definir como la capacidad del sistema informático de cambiar su tamaño o configuración para adaptarse a las circunstancias cambiantes.

Gestión de Conocimiento (Knowledge Management): Es la disciplina que busca enfocar el uso de las Tecnologías de Información en las personas, con el fin de que estas y sus organizaciones aprendan a utilizar los recursos y fuentes de información para el logro de objetivos estratégicos.

Información: Es lo que una persona es capaz de entender sobre la realidad. Los sistemas de información en la actualidad utilizan computadoras para procesar y presentar los datos en un formato comprensible para el ser humano.

MOLAP: multidimensional OLAP.

OLTP: On Line Transaction Processing.

OLAP: On Line Analytic Processing.

ROLAP: Relational OLAP.

Roll-up: Agregar.

Snowflake schema: Esquema en copo de nieve.

Star schema: Esquema en estrella.

## Anexo1. Licencias



## Proyecto CTAISC

No	Descripción	Cantidad Total
<b>Bases de Datos (Capa WareHouse y RAC) 2 Integrity rx7640 Intel Itanium 4 CPU dual core</b>		
15	Oracle Database Enterprise Edition (PERPETUAL)	8
16	Software Update License & Support for O.E. Edit	8
17	Oracle Real Application Clusters	8
18	Software Update License & Support for RAC	8
19	OLAP	8
20	Software Update License & Support for OLAP	8
21	Spatial	8
22	Software Update License & Support for Spatial	8
23	CD-PACK version 10G	1
24	Instalación y Configuración de Oracle Real Application Cluster (RAC), OLAP y Spatial, con las siguientes características: Dos (2) Servidores con cuatro (4) procesadores instalados en cada uno. Incluye pruebas, ajustes y garantía.	1
<b>Bases de Datos (Capa ETL) 1 DL580 Intel 2 CPU dual core</b>		
25	Oracle Database Enterprise Edition (PERPETUAL)	2
26	Software Update License & Support for O.E. Edit	2
27	Warehouse Builder Enterprise ETL	2
28	Software Update License & Support for ETL	2
29	Warehouse Builder Data Quality	2
30	Software Update License & Support for BuilderData Qty	2
31	Instalación y Configuración de Oracle Database Enterprise Edition, Warehouse Builder Enterprise ETL y Warehouse Builder Data Quality con las siguientes características: Un (1) servidor con dos (2) procesadores. Incluye pruebas, ajustes y garantía.	1
<b>Bases de Datos (Capa Visualización) 1 DL580 Intel 2 CPU dual core</b>		
32	Business Intelligence Standard Edition	2
33	Instalación y Configuración de Business Intelligence Standard Edition con las siguientes características: Un (1) servidor con dos (2) procesadores. Incluye pruebas, ajustes y garantía.	1