



UNIVERSIDAD CENTRAL DE VENEZUELA

FACULTAD DE CIENCIAS

ESCUELA COMPUTACIÓN

CENTRO DE INFORMACIÓN DE SISTEMAS DE INFORMACIÓN

**DESARROLLO DE UNA SOLUCIÓN DE INTELIGENCIA DE  
NEGOCIOS PARA LA OBTENCIÓN DE INDICADORES DE  
GESTIÓN QUE APOYEN A LA TOMA DE DECISIONES EN EL  
ÁREA DE SEGURIDAD, HIGIENE Y AMBIENTE**

Trabajo especial de grado presentado ante la ilustre

Universidad Central de Venezuela por el

Br. Juan Carlos Gómez Pereira

C.I. 20.756.917

Tutor: Prof. Lic. Franky Uzcategui.

Caracas, octubre 2015

UNIVERSIDAD CENTRAL DE VENEZUELA

FACULTAD DE CIENCIAS

ESCUELA DE COMPUTACIÓN

ACTA

Quienes suscriben, miembros del jurado designado por el Consejo de la Escuela de Computación, para examinar el Trabajo Especial de Grado titulado “Desarrollo de una solución de inteligencia de negocios para la obtención de indicadores de gestión que apoyen a la toma de decisiones en el área de seguridad higiene y ambiente.” y presentado por el bachiller: Bachiller Juan Carlos Gómez Pereira, a los fines de optar al título de **Licenciado en Computación**, dejamos constancia de lo siguiente:

Leído como fue dicho trabajo, por cada uno de los miembros del jurado, se fijó el día \_\_\_ de \_\_\_\_\_ de \_\_\_\_\_, a las \_\_\_\_\_ horas, para que el autor lo defendiera en forma pública, lo que estos hicieron en la Sala \_\_\_ de la Escuela de Computación, mediante una presentación oral de su contenido, luego de lo cual respondieron a las preguntas formuladas. Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió aprobar con la nota de \_\_\_ puntos.

En fe de lo cual se levanta la presente Acta, en Caracas el día \_\_\_ de \_\_\_\_\_ de \_\_\_\_\_.

---

Prof. Franky Uzcategui

(Tutor)

---

Prof(a) Mercy Ospina

(Jurado)

---

Prof Wilfredo Rangel

(Jurado)

## **AGRADECIMIENTOS**

Son muchas las personas especiales que han formado parte de mi vida durante, especialmente durante el desarrollo de este trabajo de investigación. Sin ellos la realización de este trabajo hubiese sido más difícil, por no decir imposible.

A mi familia, padres y hermano, les agradezco por esforzarse y hacer posible el que pudiese estudiar y tener los principios y valores requeridos para poder lograr mis objetivos con integridad. No suelo decirles gracias, pero su existencia es mi valor máspreciado.

A mi tutor Franky Uzcategui, quien logro guiarme durante el trabajo, brindándome los conocimientos y soportando mi carácter. Es un grato honor haber trabajado con una profesional como usted.

Para mis múltiples compañeros de infancia, que fueron mi distracción y mi voz de aliento en momentos turbios, en especial a Jean Freites y Alejandro Bonilla, vuestras amistades son invaluable.

En pocos años he logrado conocer gente muy especial en la universidad, gente que ha estado atenta y que me han acompañado con palabras de apoyo cuando la situación lo ameritaba. Quiero hacer mención especial a mis colegas Ysidro Alba, Israel Rodríguez y Andrés Álvarez.

Por ultimo hacer mención a mis compañeros de trabajo en Detaltech Sistemas, siempre atento y dispuesto a brindar ayuda, siempre estimulando mis conocimientos y mi desarrollo como profesional.

Universidad Central de Venezuela.

Facultad de Ciencias

Escuela de Computación

CISI

**Desarrollo de una solución de inteligencia de negocios para la obtención de indicadores de gestión que apoyen a la toma de decisiones en el área de seguridad higiene y ambiente.**

**Autor:** Juan Gómez

**Tutor:** Franky Uzcategui.

**Fecha:** Caracas, octubre 2015.

**RESUMEN**

El procesamiento de grandes volúmenes de datos por parte de las organizaciones ha generado la necesidad de realizar acciones de análisis y seguimiento de las operaciones. Las organizaciones comúnmente utilizan herramientas básicas para el análisis de la información que consisten en la generación de reportes estáticos y predefinidos, lo que dificulta la flexibilidad en el análisis de los datos lo que imposibilita a los usuarios explorar las distintas perspectivas posibles a partir de la información existente en la organización.

En consecuencia, el objetivo de este trabajo especial de grado, consiste en el desarrollo de una solución de inteligencia de negocios que permita, a partir del sistema transaccional para el área de seguridad, higiene y ambiente de una organización explotadora de gas, generar indicadores de gestión claves que permitan ser visualizados por los usuarios a través de una plataforma que brinde flexibilidad a la hora de generar reportes e indagar por tanto, sobre los datos para tomar entonces decisiones acertadas en función de la información.

**Palabras Claves:** Inteligencia de Negocios, Almacén de Datos, Modelo dimensional, indicadores de gestión, toma de decisiones

## Índice de contenido

CAPÍTULO 1 .....	1
4.1.- Situación actual.....	1
4.2.- Planteamiento del problema.....	2
4.3.- Solución. ....	4
4.4.-Objetivos.....	8
4.4.1- General.....	8
4.4.2- Específicos. ....	8
4.5.- Alcance. ....	8
CAPITULO 2 .....	9
2.1 – Sistema de información.....	9
2.1.2.- Tipos de sistemas en base a las necesidades de información.....	9
2.2.- Base de datos. ....	11
2.2.1.- Definición. ....	11
2.2.2.-Modelo relacional. ....	11
2.3.- Almacén de datos (Datawarehouse).....	12
2.3.1.- Definición. ....	12
2.3.2.- Antecedentes. ....	12
2.3.3.- Características de un almacén de datos.....	14
2.3.4.- Mercado de datos (Data mart).....	16
2.4.- Modelado dimensional.....	16
2.4.1.- Definición. ....	16
2.4.2.-Hecho. ....	17
2.4.3.-Dimension. ....	17
2.4.4.- Jerarquía.....	18
2.4.5.- Fases del modelado dimensional.....	18
2.4.6.- Ambiente del modelo multidimensional. ....	21
2.4.7.- Cubo OLAP. ....	21
2.4.8.- Modelo dimensional en base de datos relacional.....	22
2.4.9.-Tablas.....	22
2.4.10.- Tipos de tablas de hechos.....	24

2.4.11.-Esquemas. ....	26
2.4.12.- Transición del modelo relacional al modelo dimensional.....	29
2.5.- Inteligencia de negocios.....	30
2.5.1.- Definición. ....	30
2.5.2.- Características. ....	31
2.5.3.- Arquitectura de una solución de inteligencia de negocios. ....	33
2.5.4.- Fuente de datos. ....	33
2.5.5.- Área intermedia.....	34
2.5.6.- Almacenamiento de datos. ....	36
2.5.7.- Componente de administración y control. ....	36
2.5.8.- Componente de distribución de la información. ....	36
2.5.9.- ¿Por qué se necesita inteligencia de negocios?. ....	38
2.5.10.- Beneficios de una solución de inteligencia de negocios. ....	39
2.5.11.- ¿Qué se requiere para una solución de inteligencia de negocios?.....	40
2.5.12.- Operadores de refinamiento OLAP. ....	40
2.6.- Herramientas de inteligencia de negocio. ....	44
2.6.1- Herramientas a utilizar.....	46
2.7.- Indicadores de gestión.....	49
2.7.1.-Definición. ....	49
2.7.2.- Características de los indicadores de gestión. ....	49
2.7.3.- Elementos de los indicadores de gestión.....	50
2.7.4.- Beneficios de los indicadores de gestión. ....	51
2.8.- Proceso de seguridad, higiene y ambiente (SHA).....	51
2.8.1.- Definición. ....	51
2.8.2.- Funciones involucradas en los procesos de SHA.....	52
2.8.3.- Disposiciones legales asociadas al departamento SHA en Venezuela.....	53
2.8.4.- Organigrama del departamento SHA. ....	54
2.8.5.- Sistema de información SHA.....	55
2.8.6.- Importancia del proceso SHA. ....	59
CAPÍTULO 3 .....	60
3.1.- Metodología Bill Inmon (Top-down). ....	60

3.2.-Metodología Ralph Kimball (Bottom-up).....	63
3.3.- Metodología mejores prácticas. ....	65
3.3.1.- Análisis del negocio. ....	66
3.3.2.- Desarrollo del almacén de datos. ....	66
3.3.3.- Desarrollo de requerimientos de información.....	66
3.3.4.- Implementación de la solución. ....	67
3.4.- ¿Cuál metodología escoger?. ....	67
CAPITULO 4 .....	70
4.1.-Análisis del negocio.....	70
4.1.1.- Identificar la forma de obtención de los requerimientos.....	70
4.1.2.- Identificar fuentes de datos. ....	71
4.2.- Desarrollo del almacén de datos. ....	79
4.2.1.- Crear área intermedia. ....	79
4.2.2.- Detallar proceso ETC para área intermedia. ....	82
4.2.3.- Crear almacén de datos. ....	88
4.2.4.- Detallar proceso ETC para almacén de datos. ....	96
4.2.5.-Verificar calidad de los datos.....	97
4.2.6.- Verificación de la calidad de los datos de los componentes del modelo implementado. .....	102
4.3.- Desarrollar requerimientos de información. ....	107
4.3.1.- Configuración del administrador. ....	108
4.3.2.- Generación de indicadores de gestión.....	109
4.3.3.- Desarrollo y distribución de consultas en el cuadro de mando.....	110
4.3.4.Verificar la calidad de los datos en la herramienta de manipulación y visualización de datos. ....	121
4.4.- Implementación de la solución. ....	123
4.4.1.- Ajuste de las consultas. ....	123
4.4.2.- Establecer esquemas de seguridad. ....	125
4.4.3.-Crear flujos de procesos automáticos.....	126
CONCLUSIONES Y RECOMENDACIONES .....	128
BIBLIOGRAFÍA.....	130

## Índice de Figuras

Figura 1. Proceso actual para satisfacer un requerimiento. ....	2
Figura 2. Actividades para satisfacer un requerimiento dentro de la solución planteada. ....	5
Figura 3. Arquitectura de solución. ....	6
Figura 4. Arquitectura de solución con las herramientas a utilizar. ....	7
Figura 5 Tipos de sistemas de información. ....	9
Figura 6 Jerarquía. ....	18
Figura 7 Ejemplo de granularidad en un modelo dimensional. ....	20
Figura 8 Ejemplo modelo dimensional. ....	21
Figura 9 Cubo OLAP. ....	22
Figura 10 Tabla de hecho. ....	23
Figura 11 Tabla dimensión. ....	24
Figura 12 Esquema estrella. ....	27
Figura 13 Modelo constelación. ....	28
Figura 14 Modelo copo de nieve. ....	29
Figura 15 Componentes del almacén de datos. ....	33
Figura 16 ejemplo de cuadro de datos por Pentaho BI Server. ....	38
Figura 17 operadores ROLL-UP y DRILL-DOWN. ....	42
Figura 18 Operadores ROLL-DRILL ACROSS. ....	42
Figura 19 Operadores SLICE y DICE. ....	43
Figura 20 Operador PIVOT. ....	44
Figura 21 Cuadrante mágico de Gartner. ....	44
Figura 22 Pentaho Data Integrator. ....	47
Figura 23 Ejemplo cuadro de mando Qlikview. ....	48
Figura 24 Organigrama típico del departamento SHA. ....	55
Figura 25 Funcionalidades del sistema transaccional del sistema SHA. ....	56
Figura 26 Flujo de actividades para crear un evento. ....	56
Figura 27 Flujo de actividades para crear el detalle de un evento. ....	57
Figura 28 Flujo de actividades para crear una acción correctiva. ....	58
Figura 29. Flujo de actividades para crear un elemento de configuración general. ....	58
Figura 30 Proceso de desarrollo según Bill Inmon. ....	61

Figura 31 Proceso de desarrollo según Ralph Kimball. ....	63
Figura 32 Ciclo de vida de la metodología Kimball. ....	64
Figura 33 Fases de metodología mejores prácticas. ....	65
Figura 34 Módulos del sistema relacional.....	71
Figura 35 modelo relacional del módulo de acciones correctivas.....	72
Figura 36 Modelo relacional del módulo Opersafe.....	73
Figura 37 Modelo relacional del módulo Registro técnico. ....	74
Figura 38 Modelo relacional del módulo Incidentes y Accidentes. ....	74
Figura 39 Representación de las dimensiones y sus respectivas jerarquías. ....	76
Figura 40 Modelo dimensional del proceso SHA. ....	77
Figura 41 Arquitectura de la solución. ....	78
Figura 42 Tabla eventos. ....	80
Figura 43 Tabla eventos generada en Oracle y visualizada en SQL Developer.....	82
Figura 44 Interfaz de inicio de Pentaho Data Integrator. ....	83
Figura 45 Creación de una conexión a una base de datos. ....	83
Figura 46 Representación de las tablas de la base de datos en Kettle. ....	84
Figura 47 La extracción de elementos de una tabla se hace a partir de una consulta SQL. ....	84
Figura 48 Flujo de trabajo para generar el proceso ETC de un tipo de eventos.....	85
Figura 49 Forma de concatenar distintas columnas en Kettle.....	85
Figura 50 En Kettle se requiere establecer como sera llenada cada columna de la taba destino. ....	86
Figura 51 Proceso ETC para la tabla eventos del área intermedia. ....	86
Figura 52 Pantalla para la creación de trabajo de Kettle. ....	87
Figura 53 Trabajo que permite cargar los datos en el área intermedia.....	87
Figura 54 Modelo dimensional preparado para su implementación en la base de datos.....	88
Figura 55 Base de datos creada para soportar el modelo dimensional.....	88
Figura 56 Proceso ETC para la tabla de hechos.....	96
Figura 57 Trabajo para implementar el modelo dimensional.....	97
Figura 58 Verificación de datos de la dimensión prioridades. ....	98
Figura 59 Verificación de datos de la dimensión estatus. ....	98
Figura 60 Verificación de datos de la dimensión responsables.....	99
Figura 61 Verificación de la cantidad de responsables. ....	99
Figura 62 Verificación de datos de la dimensión departamentos.....	100

Figura 63 Verificación de datos de la dimensión sublocalidades.....	100
Figura 64 Verificación de datos de la dimensión eventos.....	101
Figura 65 Verificación de datos de la dimensión tiempo.....	101
Figura 66 Verificación de la cantidad de acciones correctivas.....	102
Figura 67 Verificación de la tabla de hechos a partir de una consulta.....	102
Figura 68 Comprobación de datos del modelo con la dimensión tiempo. tiempo.....	103
Figura 69. Comprobación de datos del modelo con la dimensión tiempo.....	104
Figura 70. Comprobación de datos del modelo con la dimensión prioridades.....	104
Figura 71. Comprobación de datos del modelo con la dimensión sublocalidades.....	105
Figura 72. Comprobación de datos del modelo con la dimensión departamentos.....	105
Figura 73. Comprobación de datos del modelo con la dimensión eventos.....	106
Figura 74. Comprobación de datos del modelo con la dimensión responsables.....	106
Figura 75 Creación de conexión con la fuente de datos desde Qlikview.....	108
Figura 76 El script dentro de Qlikview define que datos se manejaran en el cuadro de mando.....	109
Figura 77 Ejecución del script.....	109
Figura 78 Cuadro de mando inicio.....	110
Figura 79 La dimensión tiempo y sus distintos roles en el modelo.....	111
Figura 80 Cuadro de mando general.....	111
Figura 81 Los elementos del cuadro de mando se adaptan a la selección en un elemento.....	112
Figura 82 Dimensión Gerencias donde puede navegarse a partir de la jerarquía establecida... ..	112
Figura 83 Forma de manipular los valores posibles de la dimensión estatus.....	113
Figura 84 Lista de responsables, puede buscar por un apellido específico.....	113
Figura 85 Representación de las prioridades de las acciones correctivas.....	113
Figura 86 Formas de mostrar los valores y jerarquía de la dimensión sublocalidades.....	114
Figura 87 Se puede fácilmente exportar a Excel un elemento dentro del cuadro de mando.....	114
Figura 88 Cuadro de mando enfocado a fechas.....	115
Figura 89 Distintas formas de representar un calendario.....	115
Figura 90 Calendario para gestionar la fecha de creación de las acciones correctivas.....	116
Figura 91 Qlikview evita crear una selección inválida dentro del calendario.....	116
Figura 92 Calendario para la fecha de clausura de acciones correctivas.....	117
Figura 93 Otra forma de manipular las fechas en Qlikview.....	117
Figura 94 Cantidad de acciones correctivas por prioridad mostrado en un gráfico de torta.....	118

Figura 95 Acciones correctivas vencidas y por vencer. ....	118
Figura 96 Cantidad de acciones correctivas retrasadas. ....	118
Figura 97 Cuadro de mando enfocado a los eventos. ....	119
Figura 98 Forma de navegar por la dimensión eventos a partir de la jerarquía. ....	119
Figura 99 Acciones correctivas por responsable y tipo. ....	120
Figura 100 Acciones correctivas por responsable y tipo de evento aplicando operación pivot. ....	120
Figura 101 Cantidad de acciones correctivas por estatus y prioridad. ....	120
Figura 102 Verificación de datos a partir del cálculo de acciones correctivas por prioridad. ....	121
Figura 103 Verificación de los datos realizando una selección en la dimensión sublocalidades. ....	122
Figura 104 Verificación de concordancia en el retraso de acciones correctivas. ....	122
Figura 105 Verificación de cantidad de acciones correctivas para el evento Opersafe. ....	122
Figura 106 Verificación de datos seleccionando un mes y visualizando sus responsables. ....	123
Figura 107 Proceso ETC periódico de la dimensión sublocalides. ....	126
Figura 108 Script que verifica y elimina datos actualizados. ....	127
Figura 109 Inserción de elementos nuevos. ....	127
Figura 110 Job de actualización periódica de la implementación del modelo dimensional. ....	127

## Índice de tablas

Tabla 1 Comparación entre las soluciones de inteligencia de negocios.....	46
Tabla 2 Comparación metodologías.....	68
Tabla 3 Descripción de las dimensiones. ....	76
Tabla 4 Tipo de eventos. ....	80
Tabla 5 Características de los atributos de la tabla eventos. ....	81
Tabla 6 Detalle de la dimensión estatus. ....	89
Tabla 7 Detalle de la dimensión prioridades. ....	89
Tabla 8 Detalle de la dimensión responsables.....	90
Tabla 9 Detalle de la dimensión departamentos.....	91
Tabla 10 Detalle de la dimensión sublocalidad.....	92
Tabla 11 Detalle de la dimensión eventos.....	93
Tabla 12 Detalle de la dimensión tiempo.....	94
Tabla 13 Detalle de la tabla de hecho de acciones correctivas.....	95
Tabla 14 Resumen de evaluación de primer prototipo.....	124
Tabla 15 Resumen de evaluación de primer prototipo. ....	125

## INTRODUCCIÓN

La sociedad vive la era de la información. Computadores, dispositivos móviles, servidores, todas las acciones que realizan los individuos generan datos y estos datos representan las actividades y valores de cualquier organización. Si se desea comprender cualquier empresa solo hace falta ver sus procesos de negocios y sus modelos de datos.

Sin embargo, complacer a esta sociedad consumista de información y a organizaciones dependientes de los datos que generan no ha sido un proceso fácil. Los requerimientos de aplicaciones sofisticadas han generado una evolución vertiginosa en el hardware y software de equipos que han hecho posible el procesamiento y almacenamiento de millones de elementos en fracciones de segundos. Este procesamiento y capacidad de almacenamiento barato y eficaz ha logrado que cualquier individuo u organización consuma algún recurso tecnológico.

La llegada de múltiples aplicaciones que generan datos sobre distintos elementos dentro de la organización ha generado una necesidad. Esta necesidad viene de la mano con un mercado competitivo, globalizado y donde el tomar una decisión se vuelve más complicado. El campo tecnológico de una organización entonces ha tenido que evolucionar para no solo acumular datos, sino para ayudar de forma eficiente y efectiva a los directivos de una organización a utilizar estos datos para generar decisiones acertadas.

Dichas necesidades analíticas vienen de la mano con requerimientos propios del usuario, de los procesos que ejecuta la organización o de la estructura o esquema de datos que maneja la empresa. En cualquiera de los casos, una solución de inteligencia de negocios se presenta como un sistema capaz de integrar múltiples fuentes de datos de una manera lógica para la toma de decisiones y a su vez presenta la forma en como consumir estos datos integrados para que la información se muestre de una manera que permita a la gerencia de la organización tomar decisiones de forma sencilla. Estas decisiones permiten ajustar ciertos elementos a partir de un estado actual de la organización o de un estado esperable de la organización a partir de análisis predictivo.

El objetivo de este trabajo especial de grado es el estudio de conceptos, metodologías y herramientas para el desarrollo de una solución de inteligencia de negocios para el proceso de seguridad higiene y ambiente (SHA) de una organización extractora de gas.

El siguiente trabajo de investigación está estructurado en los siguientes capítulos:

Capítulo 1: se plantea la propuesta especial de grado a partir de una situación actual que presenta una problemática y donde se plantea una solución, objetivos a partir de un alcance establecido.

Capítulo 2: en este capítulo se presentan todos los conceptos que deben manejarse para poder comprender la estructura, tipo, objetivos de una solución de negocios. También serán estudiadas algunas herramientas para generar soluciones de inteligencia de negocio existentes en el mercado.

Por último se comprenderá los indicadores de gestión y su valor así como los objetivos y funciones de un departamento de seguridad, higiene y ambiente con el objetivo de entender como los indicadores de gestión pueden ser utilizados en dicho departamento para mejorar su funcionamiento.

Capítulo 3: se describen 3 metodologías para el desarrollo de una solución de inteligencia de negocio, las de Ralph Kimball, Bill Immon y una basada en mejores prácticas a partir de la metodología de Kimball e Immon. A partir de dichas metodologías se establecerán sus ventajas y diferencias para la creación de una solución de inteligencia de negocio.

Capítulo 4: Se realiza la solución planteada en el capítulo 1. En este sentido se trabaja a partir de los objetivos específicos la realización integral de una solución de inteligencia de negocios aplicada al área de seguridad, higiene y ambiente, trabajando específicamente sobre las acciones correctivas que se generan en dicha área.

Por último son realizadas el conjunto de conclusiones y recomendaciones a partir del desarrollo del trabajo especial de grado.

## **CAPÍTULO 1**

### **PROBLEMA DE INVESTIGACIÓN**

#### **4.1.- Situación actual.**

Hoy en día las organizaciones están conformadas por departamentos que realizan múltiples actividades en conjunto para lograr el máximo desempeño de la organización. El proceso de seguridad, higiene y ambiente (SHA) es un proceso que abarca un área fundamental en aquellas organizaciones que realizan operaciones delicadas que requieren un monitoreo, evaluación de incidentes para realizar las correcciones oportunas.

Por lo general las actividades correspondientes a los distintos procesos dentro de la organización, incluido el proceso SHA, son manejados con sistemas transaccionales que registran todos los eventos que ocurren.

La organización en la cual se refiere esta propuesta de trabajo especial de grado es una organización que se encarga de la extracción de gas. Por las características de una organización que realiza procesos de extracción sobre un recurso inflamable, las tareas de seguridad e higiene son fundamentales para preservar la integridad de los bienes (personal) y el ambiente.

El departamento de SHA está conformado por un conjunto de sub-divisiones que manejan los distintos eventos que ocurren en la organización en materia de seguridad, higiene y ambiente. Estos sub-departamentos o divisiones tienen un sistema transaccional que les permiten registrar los datos de un evento. Dichos datos varían según el tipo de evento. A partir del registro de un evento a su vez se pueden plantear un conjunto de acciones correctivas, que son las acciones que se deben realizar a partir de la ocurrencia de un evento.

Actualmente, el proceso para resolver un requerimiento analítico se observa en la figura 1. si algún miembro del sector estratégico-analítico de la organización requieren algún requerimiento analítico, deben comunicarse con el departamento o gerencia encargada del departamento SHA, para que el personal del área observe si el requerimiento puede satisfacerse con algún reporte que se posea. Si esto no ocurre se requiere que el departamento de informática genere el reporte que cumpla con el requerimiento. Como este proceso puede

llevar tiempo (entre 2 y 4 semanas), los encargados del área a partir de los reportes existentes y utilizando hojas de cálculos realizan de forma manual el reporte para satisfacer de forma provisional el requerimiento existente, aunque esto trae la posibilidad de que exista un error manual por parte del personal y la utilización de un recurso de la empresa de forma completa en la resolución de dicho requerimiento analítico.

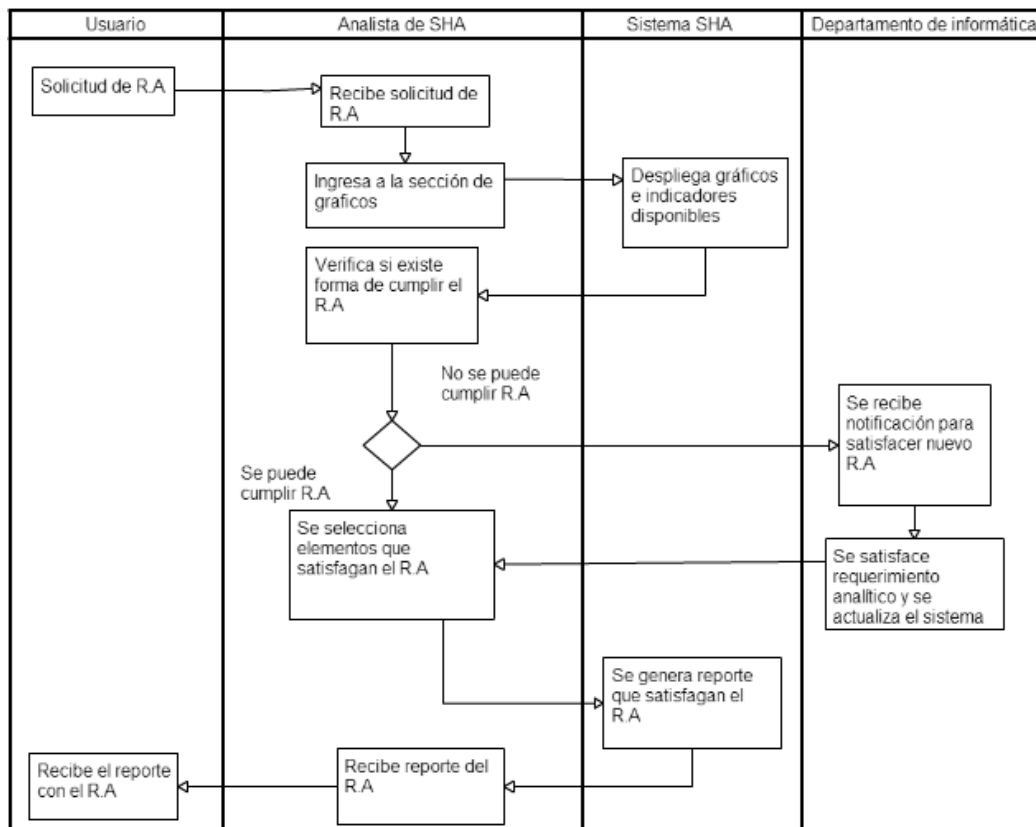


Figura 1. Proceso actual para satisfacer un requerimiento.

#### 4.2.- Planteamiento del problema.

Las organizaciones que no poseen un sistema analítico dedicado, es decir, que solo manejan sistemas transaccionales (OLTP) para el manejo de operaciones, tienen eventualmente problemas para realizar análisis, monitoreo y tomar decisiones en función de los datos recolectada. El proceso de SHA es un área que registra múltiples operaciones dentro de la organización, sin embargo una vez almacenado los datos, si no se posee una herramienta para su procesamiento y su visualización a los usuarios se les dificulta la tarea de obtener información provechosa que pueda ayudar al desempeño de la organización. Pese a que el

área de SHA de la organización posee un sistema transaccional que permite registrar todos los datos referentes a este proceso. Las capacidades analíticas de esta área se ven mermadas por la ineficacia de dicho sistema de poder generar información para ser utilizada por la gerencia de la organización. Una solución inicial que busco la empresa fue la generación de indicadores y reportes estáticos que abarcaran los indicadores de gestión más importantes.

Sin embargo esta solución parcial posee serios inconvenientes que son descritos a continuación:

- **Poca flexibilidad:** los indicadores de gestión que se observan y la forma como se visualizan no puede ser modificado. Esto implica que la gerencia del área SHA o la organización no tiene otra forma de obtener información para tomar decisiones. En este sentido, esa poca flexibilidad dificulta las capacidades analíticas las cuales requieren manipular los datos, modificar los indicadores, la forma como estos indicadores se observan con el fin de obtener una vista de los datos que genere información útil en un momento dado.
- **Trabajo manual:** el poder satisfacer una necesidad analítica de un usuario implica que el mismo contacte a los encargados del área de SHA para que puedan establecer si la necesidad puede resolverse. Si la necesidad puede resolverse se envía el usuario. Esta comunicación manual donde el usuario realmente no tiene acceso a herramientas analíticas hace ineficiente el proceso de visualización de indicadores para la toma de decisiones.
- **Problemas en generación de nuevos informes o gráficas:** la gerencia suele requerir nuevas formas de visualizar los datos o modificar los reportes. Producto del punto anterior, estas modificaciones requiere que sea contactado el equipo de informática de la organización o de terceros (consultoría) para que realicen dichos cambios. Esto representa costos adicionales, la agregación de nuevos elementos a la aplicación lo que puede afectar su desempeño y un tiempo de espera mientras se están realizando las modificaciones. Este tiempo de espera, que puede ser de entre 2 y 4 semanas, implica que una necesidad analítica no pueda resolverse en el momento que esta se genera, lo que muestra que el sistema no ayude a la toma de decisiones en el momento oportuno.

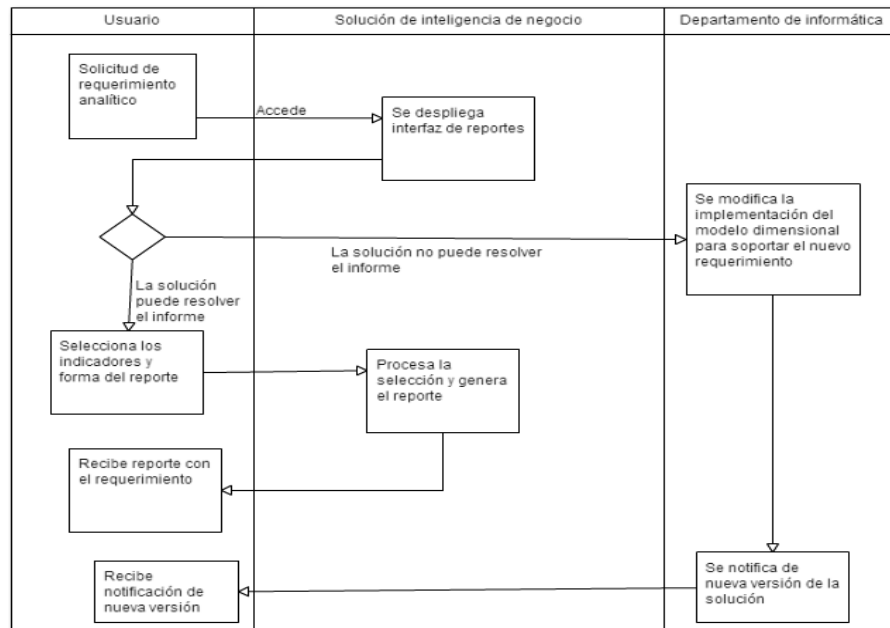
- **Procesamiento innecesario:** la herramienta genera todo el conjunto de gráficos para un evento particular aunque en ese momento, un determinado elemento de visualización no sea requerido. Esto implica que la aplicación genera un procesamiento innecesario que afecta el rendimiento. Además como no existen operadores OLAP, es decir operadores que permitan filtrar los datos utilizada en los informes, todos los datos son utilizados y procesados para generar los datos. Esto implica que a medida que el sistema transaccional acumule datos de las transacciones realizadas, el sistema procesara más datos y disminuirá su capacidad de respuesta.

#### **4.3.- Solución.**

Dada la situación actual, con sus respectivos problemas anteriormente señalados, se plantea como solución la creación de un modelo dimensional que permite obtener indicadores de gestión relevantes para la organización, para posteriormente permitir a los usuarios obtener de forma rápida y flexible información relevante y oportuna para la toma de decisiones.

El levantamiento de requerimientos de la solución vendrá dada por el modelo transaccional que utiliza la organización para manejar el proceso de SHA. Por lo que a partir del modelo transaccional, se generara un modelo dimensional con la capacidad de satisfacer las necesidades analíticas de esta área dentro de la organización.

La solución planteada permitirá, como se observa en la figura 2, poder satisfacer un requerimiento de cualquier componente de la organización a partir del acceso a la solución de inteligencia de negocios. La solución brindara la capacidad de que el usuario manipule las variables del indicador de forma que pueda obtener el requerimiento solicitado. Puede ocurrir que la solución no posea la capacidad para satisfacer el requerimiento. Como se genera el modelo dimensional a partir del esquema transaccional, esta imposibilidad de cubrir un requerimiento puede ocurrir porque el esquema transaccional no posee los datos para generar el requerimiento o el proceso de SHA ha sido modificado y por tanto el esquema transaccional ha recibido modificaciones. En ambos casos el departamento de informática tiene que realizar las modificaciones en la implementación del modelo dimensional para que el usuario pueda obtener su requerimiento.



**Figura 2. Actividades para satisfacer un requerimiento dentro de la solución planteada.**

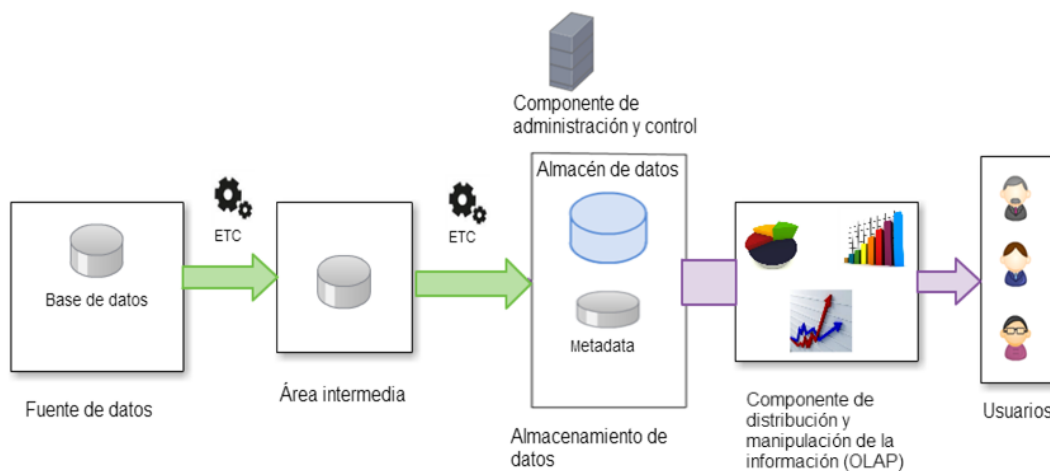
Aunque no se conocen los requerimientos a priori porque los requerimientos se obtendrán del esquema de datos y su implementación busca poder cumplir con cualquier requerimiento de información. De antemano se sabe que la solución abarcará los siguientes requerimientos:

- Cantidad de acciones correctivas por departamento.
- Estado de las acciones correctivas en un instante de tiempo.
- Evaluar la evolución de las acciones correctivas en el tiempo en función de su estado, departamento o gerencia.
- Identificar responsables de acciones correctivas vencidas o por vencer según el tipo de evento.

La solución planteada se establece a partir de una base de datos relacional que soporta los distintos módulos del proceso de SHA, así como las acciones correctivas correspondientes a los distintos eventos. La solución está conformada como puede verse por la figura 3, por los siguientes componentes:

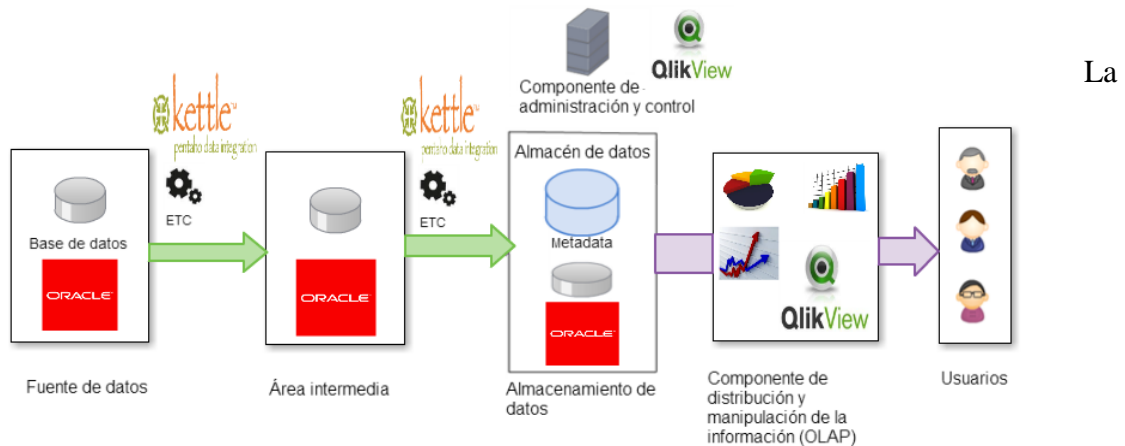
- Un área intermedia que permitirá estandarizar, consolidar y transformar los datos para el análisis.

- Un conjunto de modelos dimensionales que serán generados a partir del particionamiento lógico de los datos contenida en el área intermedia y el sistema transaccional. Esto genera como resultado un esquema estrella como forma de implementación del modelo dimensional.
- Dos procesos de ETC: el primer proceso ETC permitirá extraer solo los datos relevantes del sistema transaccional del área de SHA. El segundo proceso ETC permitirá dividir los datos en distintos esquemas estrellas.
- Herramienta de inteligencia de negocios que permita una vez generado los modelos dimensionales generar un conjunto de elementos analíticos que permita a los usuarios de la organización obtener información útil.



**Figura 3. Arquitectura de solución.**

Producto de que la organización utiliza el sistema manejador de base de datos Oracle para el soporte de los datos del departamento SHA y considerando las capacidades de dicho manejador, el mismo será utilizado para el área intermedia y el almacén de datos. Para el desarrollo de los procesos ETC será utilizada la herramienta Pentaho Data Integrator (Kettle) y como herramienta de inteligencia de negocios Qlikview. Por tanto el esquema de la solución (incluido las herramientas) puede observarse en la figura 4.



**Figura 4. Arquitectura de solución con las herramientas a utilizar.**

selección de Oracle como base de datos en los distintos niveles de la solución se debe al hecho de que la organización utiliza Oracle como base de datos, adicionalmente, la base de datos de Oracle es una base de datos confiable, eficiente y efectiva para los requerimientos de la solución, por lo que se prefirió añadir otro motor de base de datos a la solución.

Como herramienta de distribución y manipulación de la información se utiliza Qlikview porque la organización, la cual es a su vez una filial de otra organización, utilizan Qlikview como herramienta de inteligencia de negocios. Adicionalmente el estudio de la herramienta demuestra ser de fácil implementación, enfocada a medianas empresas y con la suficiente visión de negocio para poder abarcar la implementación del modelo dimensional y extraer los datos del almacén de datos para generar los indicadores correctivos.

En el caso de la herramienta de ETL, se estudiaron múltiples opciones y se decidió por la herramienta de pentaho, Kettle. Kettle es una herramienta intuitiva, de fácil instalación y generación de procesos de ETL, además la versión comunitaria posee todas las herramientas requeridas para la solución.

Por último, el componente de administración y control será utilizado Qlikview, debido a que Qlikview permite actualizar el modelo de datos y los datos a utilizar en el cuadro de mandos, generar privilegios según los distintos usuarios de la organización y modificar parámetros de configuración como forma de imprimir cuadros de mandos, opciones para enviar reportes por correo, forma de acceder al servidor en caso de que se acceda al cuadro de mando de forma remota, etc.

#### **4.4.-Objetivos.**

##### **4.4.1- General.**

Desarrollar una solución de inteligencia de negocios para la obtención de indicadores de gestión que apoyen a la toma de decisiones en el área de seguridad, higiene y ambiente.

##### **4.4.2- Específicos.**

Se describen los objetivos específicos planteados para la propuesta establecida:

- Definir el modelo dimensional en base al sistema transaccional.
- Desarrollar el almacén de datos.
- Desarrollar los requerimientos de información (indicadores de gestión).
- Implementar la solución.

##### **4.5.- Alcance.**

El desarrollo de la solución de inteligencia de negocios se enfocara al diseño y despliegue de los indicadores asociados al proceso de seguridad higiene y ambiente, específicamente al módulo de acciones correctivas utilizando los datos del sistema transaccional como fuente para la carga del almacén de datos. Se seguirá la metodología de mejores prácticas utilizando como plataforma tecnológica Oracle como repositorio de datos en los distintos niveles de almacenamiento de datos, Pentaho para la generación de los procesos ETC y Qlik para la manipulación de los datos. Adicionalmente se construirán un conjunto de indicadores de gestión piloto para poder observar la aplicación de la herramienta.

## CAPITULO 2

### MARCO CONCEPTUAL

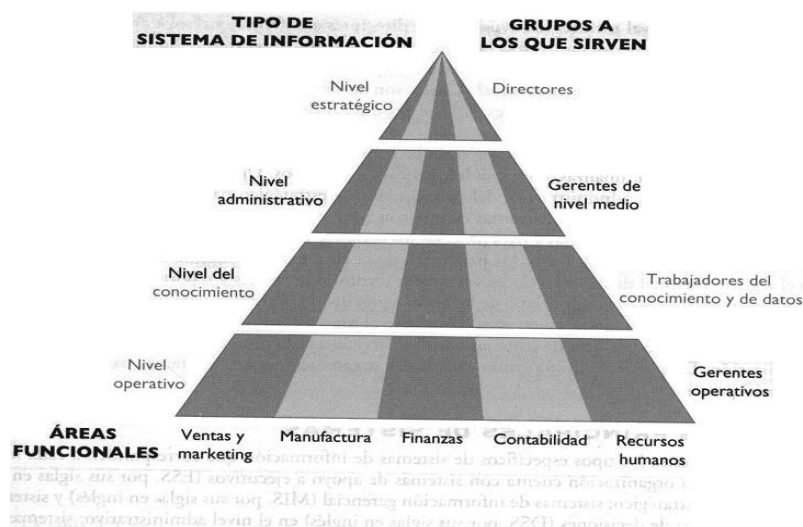
#### 2.1 – Sistema de información.

##### 2.1.1.- Definición.

Son un conjunto de procesos que interactúan entre sí para crear, almacenar, recopilar y procesar datos para la generación de información a la organización que permita apoyar las actividades realizadas en la organización, tales como monitoreo y control de operaciones y soporte a la toma de decisiones.

##### 2.1.2.- Tipos de sistemas en base a las necesidades de información.

En función a los distintos roles que desempeñan los usuarios o empleados dentro de una organización, las necesidades y el tipo de información a requerir variara en cuanto al nivel de agregación y forma de visualización. En este sentido, como se puede ver en la figura 5, a partir del nivel que se encuentre un usuario dentro de la pirámide, podemos obtener diferentes tipos de sistemas de información.



**Figura 5 Tipos de sistemas de información.**

**Fuente: Laudon, K. y Laudon J. (2008). *Sistemas de información gerencial: Administración de la empresa digital* (10ª ed.). México: PEARSON Education.**

- Nivel operativo (sistemas de procesamiento de transacciones).

Son sistemas que permiten a las áreas de operaciones de una empresa, registrar todas las transacciones o eventos dentro de la organización de una forma automatizada, que permite a los gerentes de operaciones realizar consultas básicas, por ejemplo, la cantidad de un producto en inventario. Como también, permite el seguimiento de las operaciones que se ejecutan en un momento determinado.

Suelen ser sistemas diseñados para soportar la carga de muchas transacciones por lo general sencillas y habituales.

- Nivel conocimiento (Sistemas de información de trabajo de conocimiento).

Promueven la creación y administración de conocimiento en el área operativa. Por lo general este nivel está conformado por personal profesional (ingenieros, científicos, programadores) que generan o diseminan información creando documentación, productos y sistemas.

- Nivel administrativo (sistema de información gerencial y sistemas de apoyo a la toma de decisiones).

Conformado por la gerencia intermedia, los sistemas de información para el área administrativa permiten supervisar, coordinar, administrar y ayudar en la toma de decisiones.

Los sistemas de información gerencial se encargan de mostrar y resumir la información proveniente de las transacciones (nivel operativo) para ser visualizada por la gerencia. Por lo general, se generan informes periódicos y se aplica agregación en los datos para mostrar de forma resumida el desempeño de la organización.

Los sistemas de apoyo a la toma de decisiones ayudan a la gerencia intermedia a definir decisiones. Suelen enfocarse en problemas particulares, con cambios rápidos y que no se pueda establecer un patrón de resolución predefinido. Por lo que estos sistemas buscan utilizar datos internos y externos a la organización para generar modelos, predicciones, y un conjunto de indicadores que permitan ayudar a la gerencia intermedia a definir tácticas que mejoren el desempeño de la organización.

- Nivel Estratégico (Sistema de apoyo a ejecutivos).

Los sistemas de apoyo a ejecutivos son sistemas enfocados a la alta gerencia de una organización. Suelen ayudar a la toma de decisiones no rutinarias (creación de una nueva fábrica, extensión del mercado a un nuevo país, etc) que suelen tener alto impacto en la organización durante años. Estos sistemas presentan una información aún más resumida que los sistemas de información gerencial y la presentación de dicha información suele darse a partir de elementos gráficos (histogramas, gráfico de torta, mapas) que ayuden a la comprensión del estado de la organización.

## **2.2.- Base de datos.**

### **2.2.1.- Definición.**

Según Molina (2008) una base de datos es una colección de datos relacionados lógicamente que existen por un periodo de tiempo determinado y que es manejada por un sistema manejador de base de datos. Las bases de datos permiten servir como repositorio de datos para el almacenamiento, procesamiento y extracción de datos.

En una base de datos, aparte de los datos que son almacenados representado las operaciones de las organización, también hay una serie de datos que describen estos datos, como el tipo de dato, ubicación física, relación con otros datos, etc. A estos datos se le conoce como metadato.

Los datos que pueden ser almacenados en una base de datos dependerá del sistema manejador de base de datos, más a nivel conceptual pueden soportar cualquier de datos (audio, texto, imágenes, etc.). Dentro de las bases de datos, destaca el concepto de base de dato relacional, la cual es un tipo de base de datos que se genera a partir de un modelo dimensional.

### **2.2.2.-Modelo relacional.**

Es un modelo ampliamente utilizado e implementad en el desarrollo de sistemas transaccionales. El modelo relacional se sustenta en el concepto matemático de relación, donde los datos son representados en forma de “tablas”, donde cada instancia de la tabla se interpreta como un conjunto de valores ordenados. A su vez cada instancia de una tabla puede estar interconectada a otra instancia en cualquier otra tabla.

La normalización, específicamente la tercera forma normal definida por E.F Codd (1971) busca generar atomicidad en los datos y por tanto eliminar la redundancia en el esquema. Esta forma normal se obtiene al evitar dependencias entre los datos de una misma tabla, teniendo un solo atributo clave en la tabla y eliminando dependencias transitivas entre atributos.

Cuando se implementa un esquema relacional para sustentar las operaciones organizacionales, la 3 forma normal suele ser implementada para buscar la integridad en los datos y por tanto asegurar la veracidad de los mismos.

### **2.3.- Almacén de datos (Datawarehouse).**

#### **2.3.1.- Definición.**

Según Bill Inmon (1992) “Un datawarehouse es una colección de información creada para soportar la toma de decisiones”. La existencia del almacén de datos va a permitir la construcción de sistemas para la toma de decisiones, esto porque el almacén de datos es el puente entre los datos operacionales y unos datos diseñados para herramientas de análisis. Esto lo logra el almacén de datos integrando todas las fuentes de datos de la organización, en un único modelo.

La forma como se estructura los datos dentro de un almacén de datos para su almacenamiento se basa por tanto, en una estructura que permita la toma de decisiones. Esta estructura usualmente se va a basar en un modelo dimensional, porque según Kimball (2002) “una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis” y el modelo dimensional está diseñado para realizar consultas y análisis por parte de los usuarios finales.

#### **2.3.2.- Antecedentes.**

Desde la creación de los computadores, estos se han visto como un elemento importante para las organizaciones. Su evolución en costo, prestaciones y dimensiones han permitido que no solo grandes organizaciones puedan poseerlos sino que cualquier organización pueda poseer algún tipo de sistema. En los años 1970 empiezan a parecer las hojas de cálculo, siendo la

primera herramienta de software orientada a la manipulación de datos para usuarios de las organizaciones. En los años 70 a partir de los trabajos de E.F.Codd sobre sistemas relaciones, se crearon las primeras tecnologías capaces de soportar base de datos relacionales, lo cual fue un importante paso para el almacenamiento, distribución y utilización de los datos de las organizaciones.

La implantación de las bases de datos relacionales permitió a las compañías mejorar su capacidad de manejar información, lo que ayudo a la mejora en sus procesos y a un crecimiento de sus capacidades operativas. Sin embargo, estos primeros sistemas generados en los años 70 y 80 requerían del pedido constante de consultas al departamento de TI de la organización que generaban la consulta y el posterior reporte. Lo que se terminó denominando centros de información.

Los sistemas de soporte a la toma de decisiones surgieron con la creación de software de una complejidad mayor buscando brindar información estratégica de forma más sencilla. Sin embargo en dichos sistemas ocurría que existía una espiral de requerimientos hacia el departamento de TI, que requería estar modificando constantemente los sistemas para generar nuevos reportes.

“Buscando los orígenes de los almacenes de datos es difícil conseguir el mejor punto de partida dentro de cada uno de los diferentes libros y artículos que fueron publicado. Lo que es claro es que muchas personas tenían la idea y estuvieron trabajando con el concepto antes de que fuera publicado en un trabajo con el título de ‘almacén de datos’ ” (Smi, 1999).

Durante los años 80 y 90 empezaron a surgir ideas que fueron generando lo que hoy es llamado almacén de dato. Por los años 1980 E.F Codd creo los sistemas analíticos de procesamiento en línea (OLAP) los cuales están estrechamente ligados a los almacenes de datos. A partir de los sistemas OLAP, que definían un modelo diseñado explícitamente para el análisis de datos adicional a los sistemas operacionales existentes en la organización, en el año 1991, William Inmon público un libro por el cual se le considera el padre de los almacenes de datos (aunque ya se estaba trabajando con la idea desde antes) “construyendo el almacén de datos”. Este libro definía y explicaba las características de un almacén de datos,

Ralph Kimball en 1994 contribuyó con su libro “caja de herramientas de almacenes de datos” proporcionando una guía para diseñar un almacén de datos.

Con estos dos libros como bases, las mejoras de las tecnologías y sistemas que se diseñaron para el análisis de las organizaciones, los almacenes de datos pasaron a ser la piedra angular para los sistemas de toma de decisiones, por su capacidad de generar una base para la construcción de herramientas complejas que consulten datos que han sido estructurado para análisis complejos.

### **2.3.3.- Características de un almacén de datos.**

Según Ponniah(2001) los almacenes de datos cuentan con 4 características fundamentales:

- Data orientada a temas.

En los sistemas operacionales, el repositorio de datos (base de datos) se estructura en función de las aplicaciones existentes en la organización, para darle soportes a las mismas. Por tanto, se puede tener los datos de un cliente en una tienda en línea porque la aplicación requiere de este dato para ejecutar alguna funcionalidad, como tener dirección del cliente para enviar algún objeto comprado.

En los almacenes de datos, la estructuración de los datos no va a depender de aplicaciones, sino de los temas dentro de un negocio. Nos referimos a un tema, como un área del negocio englobada a partir de sus características que representan algún interés para la organización. Un ejemplo de esto pueden ser las áreas de ventas, compras y recursos humanos.

Ya que el almacén de datos se va a orientar de un tema, este buscará a partir de los indicadores, generarle a los analistas de la organización la capacidad de tomar decisiones sobre el área, mejorando el rendimiento total de la organización.

- Data variante en el tiempo.

Para el funcionamiento de los sistemas operacionales, los datos almacenados son aquellos que sirven para completar las distintas funcionalidades existentes. Estos son los valores actuales que se manejen en la organización. Por ejemplo, en un sistema de cuentas bancarias,

aunque obviamente se almacena los datos históricos para su consulta, el saldo actual de una cuenta es el importante para realizar las operaciones sobre la misma, como retirar dinero de un cajero. En dicho retiro al cajero, cuánto dinero tenía esa cuenta el día de ayer es irrelevante. Por lo general los sistemas operacionales guardan datos históricos para el control y auditoría de operaciones, mas no propiamente para su funcionamiento.

En el caso de los almacenes de datos, su objetivo es el análisis y soporte a las decisiones. Para realizar un análisis sobre los clientes y su manejo con las cuentas, tener solo el saldo actual de la cuenta es insuficiente. En este sentido se requiere que tener el histórico de la cuenta de los saldos para poder hacer un análisis. Por tanto en el almacén de datos se tendrá por un elemento, los distintos valores que tuvo el mismo, para evaluar su variación en el tiempo. Esto se hace guardando la información periódicamente al almacén de datos.

El almacén de datos debe permitir analizar información pasada, relacionar información presente y predecir hechos futuros.

- Data no volátil.

Siendo el soporte de los datos de las distintas aplicaciones dentro de la organización. Los datos en un sistema operacional están en constante cambio. Como son operaciones en tiempo real, las inserciones, eliminaciones y modificaciones pueden ocurrir en cualquier instante de tiempo. Por tanto hablamos de una data volátil. Su existencia puede cambiar en cualquier instante.

Esto no ocurre en los almacenes de datos, esencialmente porque, en primera instancia, los almacenes de datos no están soportando aplicaciones. Por lo que la modificación de los datos operacionales a partir de una transacción no debe implicar una modificación inmediata en los datos del almacén. En este sentido en función de la naturaleza de la organización y el área de negocio, se establecen momentos de carga, donde se envía los datos al almacén de datos (esto puede ser una vez al día, una vez a la semana, etc.). Y por otro lado, los datos en un almacén de datos como guarda datos históricos, estos datos debería no modificarse o eliminarse, las operaciones sobre un almacén solo son de carga de datos en un instante de tiempo de tiempo definido y de lectura para realizar análisis.

- Data integrada.

Los sistemas operacionales suelen trabajar con un solo modelo de datos, ejecutándose sobre un proceso específico. En los almacenes de datos, como existe una búsqueda de analizar los datos a partir de la recolección de los datos de los sistemas operacionales así como fuentes externas a la organización que sirvan para realizar análisis. Se tendrán distintas fuentes de datos, lo que implica distintos formatos en los datos, en cómo se guardan los mismos eventos con nomenclaturas distintas, etc.

Por esto, el almacén de dato debe tener la capacidad de integrar todas estas fuentes de dato para generar datos consolidados (que todos los datos se almacene bajo los mismos parámetros, estructura, etc.) que permita realizar análisis.

#### **2.3.4.- Mercado de datos (Data mart).**

Un mercado de datos es esencialmente un almacén de dato más pequeño pero con las mismas características (punto anterior). La mayor diferencia es su alcance, lo que incide directamente en el tamaño del mismo. Según Inmon (2005,p.16) “el nivel departamental o de mercado de datos está formada por requerimientos del usuario final en una forma específica situados dentro de las necesidades de un departamento.”. Este alcance en función a un departamento o un proceso de negocio específico hace que la cantidad de datos manejados por mercado de datos sea inferior al de un almacén de datos, que tiene un alcance global.

Por lo general estos mercados de datos atacan un único problema en el área de ventas, proveedores, clientes, por lo que el modelo estrella que los define es simple y su implementación es rápida en comparación a la de un almacén de datos. Debido a su limitado campo de acción, suelen unirse muchos mercados de datos a partir de dimensiones o atributos en común, para obtener una visión global dentro de la organización.

#### **2.4.- Modelado dimensional.**

##### **2.4.1.- Definición.**

“el modelado dimensional es una técnica de larga trayectoria para hacer las bases de datos simples. En un caso tras otros, por más de cinco décadas. Organizaciones TI, consultoras y

usuarios de negocios han gravitado naturalmente a una simple estructura dimensional para captar la necesidad humana fundamental por la simplicidad. La simplicidad es crítica porque asegura que los usuarios pueden fácilmente entender los datos, así como permite software que naveguen y encuentren resultados rápida y eficientemente” (Kimball,2013,p.7).

El modelo dimensional, que es el producto generado a partir del proceso de modelado dimensional, es clave para la generación de sistemas que ayuden a la toma de decisiones, porque permiten facilitar el entendimiento de los datos y mejorar las prestaciones del sistema. Esencialmente, los modelos relacionales que son los utilizados en los sistemas transaccionales, si bien permiten el registro y carga de una gran cantidad de transacciones, carecen de la capacidad de realizar consultas o una navegación compleja sobre los datos. Esto ocurre porque el modelo dimensional se estructura en función de valores o requerimientos claves para la organización y partir de esta necesidad generar un modelo que permita el análisis para la toma de decisiones.

Para realizar un análisis de un requerimiento o área de la organización, el modelo dimensional consta esencialmente de dos elementos. Por un lado valores o medidas que sirven como indicadores (hecho) y el conjunto de parámetros que definen la existencia de ese hecho en particular (dimensión).

#### **2.4.2.-Hecho.**

Es una representación de una medida dentro de la organización. Los hechos dentro del modelo dimensional permiten evaluar el desempeño de la organización en un proceso específico. Por lo general los hechos suelen ser numéricos y se puede aplicar sobre ellos alguna operación de agregación (suma, promedio, etc.). Kimball (2013).

#### **2.4.3.-Dimension.**

Según Kimball (2013,p.13) “las dimensiones provén el punto de entrada a la data”. Una dimensión es una característica que permite analizar a los hechos. En este sentido, la dimensión le va a agregar valor al hecho, situándolo en un contexto que lo define, por lo que, un hecho, suele estar relacionado con una o varias dimensiones.

Ya que los datos dentro de la dimensión busca describir un hecho. Estos son tipo cualitativo. Siendo una característica resaltante el hecho de que dichos atributos pueden organizarse para dar distintos niveles de detalle a la hora de realizar un análisis particular. A esta organización se le llama jerarquía.

#### 2.4.4.- Jerarquía.

Una jerarquía define la forma en como un conjunto de atributos de una dimensión pueden ser agrupados para realizar análisis. Por ejemplo, como se observa en la figura 6, dentro de la dimensión tiempo, una jerarquía puede darse con los atributos mes, trimestre, año.

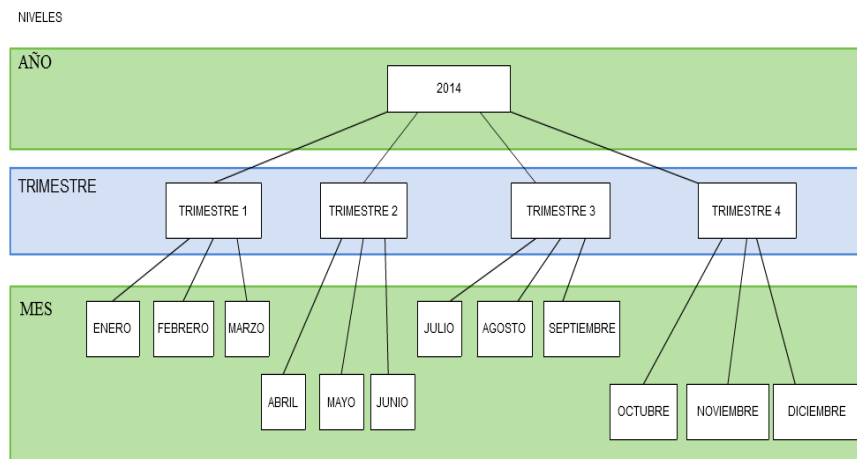


Figura 6 Jerarquía.

#### 2.4.5.- Fases del modelado dimensional.

Para desarrollar un modelo dimensional, Kimball (2013.p.38) sugiere cuatro fases para su construcción. Dichas fases son presentadas a continuación.

- **Seleccionar el proceso de negocio.**

Dentro de una organización, existen múltiples tareas o procesos que se interrelacionan entre sí. El primer paso en el modelado dimensional consiste en identificar cual proceso de negocio se quiere modelar. El proceso define la granularidad, hechos y dimensiones.

Un aspecto importante en la selección del proceso de negocio, es establecer los requerimientos analíticos. En este sentido el levantamiento de requerimientos y por

consecuencia la creación del modelo dimensional, pueden darse a partir de distintos escenarios, desde los cuales es necesario la construcción de un modelo dimensional como base para la creación de herramientas que permitan realizar análisis y ayudar en la toma de decisiones. Existen tres escenarios posibles desde los cuales se establecen los requerimientos de un modelo dimensional:

- **Listado de requerimientos:** se parte la construcción del modelo dimensional a partir de una lista de requerimientos de información suministradas por el dueño del proceso. En este sentido, el modelo de datos va a estar ajustado a extraer, calcular y cargar los datos requeridos para cumplir con los requerimientos preestablecidos.
- **Análisis del proceso:** en este escenario el modelo dimensional busca a partir del entendimiento del proceso de negocio establecer y resolver las necesidades analíticas. Esto requiere que el diseñador comprenda el negocio, los factores que influyen en el mismo, etc. Por lo que el mismo requiere comprender los sistemas transaccionales que operan en el negocio así como utilizar cualquier otro material que le suministra información del proceso.
- **Modelado de datos transaccional:** en función de los modelos transaccionales físicos que operan en la organización se realiza el modelo dimensional. No existen requerimientos a priori, por lo que se busca es crear un modelo dimensional capaz de satisfacer cualquier requerimiento posible del negocio.

En cualquiera de los tres escenarios, los modelos transaccionales e incluso el proceso de negocio pueden ser modificados para poder satisfacer algún requerimiento.

- **Definir la granularidad.**

Inmon (2005,p.41) define la granularidad como “el nivel de detalle o sumario de las unidades de data [...] Entre más detalle existe, menor es el nivel de granularidad”. Los sistemas operacionales al registrar todas las transacciones de una organización, van a tener el nivel más bajo de granularidad, por ejemplo, una venta, el registro de un cliente, etc.

En el modelado dimensional, es importante el término de granularidad, porque esto va a definir los indicadores o métricas dentro de la tabla de hechos. Un bajo nivel de granularidad

permite mayor flexibilidad en la creación de reportes y manejo de los datos para generar nuevas agregaciones y análisis. Por ejemplo, si se tiene la inscripción de un estudiante a la universidad como elemento a medir, se puede describir la granularidad como el elemento a medir según el contexto. Por tanto, según la figura 7 medir la cantidad de estudiantes inscritos a una escuela provenientes de un municipio en un semestre determinado define la granularidad.

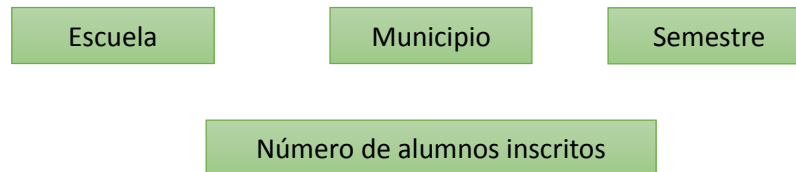


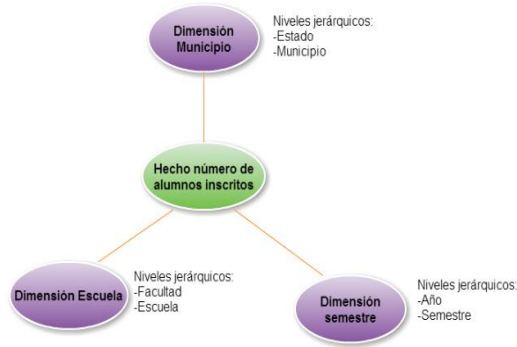
Figura 7 Ejemplo de granularidad en un modelo dimensional

- **Identificar dimensiones.**

Como se describió antes (2.5.3), las dimensiones caracterizan los hechos dentro del modelo dimensional. Por lo que identificar los elementos que le dan sentido a la ocurrencia o existencia de un dato que sirve para la toma de decisiones es fundamental para que nuestro modelo dimensional tenga la capacidad de manipular los hechos en función de distintas variables, para la granularidad planteada en la figura 7, las dimensiones que caracterizan la inscripción de un estudiante son el semestre, la comunidad y la escuela.

- **Identificar hechos.**

Como último paso, a partir de la granularidad y las dimensiones, se establecen los hechos, que no son más que las medidas con las cuales los analistas se apoyaran para tomar decisiones. Para la granularidad planteada como ejemplo, el hecho sería la existencia de una inscripción como tal, porque se busca realizar análisis en función de la suma de inscripciones. A partir de las dimensiones y el hecho, podemos generar el modelo dimensional, presentado en la figura 8.



**Figura 8 Ejemplo modelo dimensional.**

#### **2.4.6.- Ambiente del modelo multidimensional.**

Según Ponniah (2001), un modelo dimensional puede ser implementado en una base datos relacional o multidimensional. Si esta implementado en una base datos relacional, nos referimos a un modelo estrella o copo de nieve (del termino snowflake en inglés), si se ha implementado en una base de datos multidimensional hablamos de un cubo OLAP.

En ambos ambientes se tiene un mismo diseño lógico, en el sentido de que existen un conjunto de indicadores que son descritos por una serie de dimensiones. Sin embargo su diseño e implementación a nivel físico va a variar por la forma en que las bases de datos trabajan, diferencia que puede verse en los tipos de sistemas OLAP (ver punto 2.2).

#### **2.4.7.- Cubo OLAP.**

Es una representación de un modelo dimensional que se base en tener un conjunto de dimensiones que se interceptan mostrando una serie de indicadores a medir a partir de la implementación en una base de datos multidimensional de nuestro modelo de negocio. El termino cubo refiere a la existencia de 3 dimensiones asociadas a un hecho, como se puede ver en la figura 9. Sin embargo, el concepto de cubo OLAP puede aplicarse para n dimensiones que trabajen sobre k indicadores. Cuando es necesario tener más de 3 dimensiones se coloca el término de hipercubos. Cano (2007).

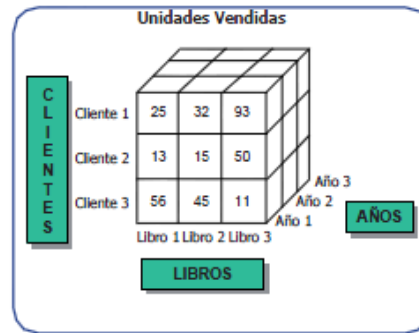


Figura 9 Cubo OLAP.

Fuente: Cano, J. L. (2007). *Business Intelligence: Competir con información*. España:ESADE.

#### 2.4.8.- Modelo dimensional en base de datos relacional.

Molina(2008) refiere a que la utilización de base de datos relacionales para la implementación de modelos dimensionales es común, en parte por la flexibilidad, escalabilidad y popularidad de dichas bases de datos. Esto implica, que se tenga que adaptar la construcción de dichas bases de datos para poder soportar el modelo dimensional y por tanto las necesidades de los usuarios. Una base de datos relacional, que se basa en un modelo relacional, parte de la existencia de un conjunto de tablas que guardan los datos y que se relacionan entre sí en función del modelo del negocio.

Para un modelo dimensional, tendremos dos tipos de tablas. Estas tablas y su organización a su vez conforman el esquema que define el modelo dimensional.

#### 2.4.9.-Tablas.

Son un elemento típico en ambientes de bases de datos relacionales. Las mismas permiten a partir de un conjunto de atributos, almacenar datos estructurados, es decir, registros conformados por un conjunto de atributos establecidos. Dentro de los almacenes de datos existen dos tipos de tablas que corresponden a los dos tipos de elementos existentes en el almacén, la tabla de hechos y la tabla dimensión. (kimball, 2013, p.12).

- **Tabla de hechos.**

“la tabla de hechos en un modelo dimensional guarda las medidas de desempeño resultantes de los eventos de un proceso de negocio de la organización. Uno debe esforzarse en guardar el más bajo nivel de data de las medidas resultantes” (kimball, 2013, p.12).

Dentro del modelo dimensional, en una tabla de hechos como podemos ver en la figura 10, podemos conseguir 3 tipos de atributos. El conjunto de indicadores que miden un evento particular. Una clave primaria que identifica a ese hecho y un conjunto de claves primarias que relacionan a ese hecho con un conjunto de dimensiones.

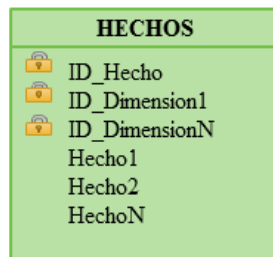


Figura 10 Tabla de hecho.

**Fuente:** Kimball, R. (1998). *The datawarehouse lifecycle toolkit* (1ª ed.). Estados unidos de América:Wiley.

- **Tabla de hechos sin hechos (Factless table).**

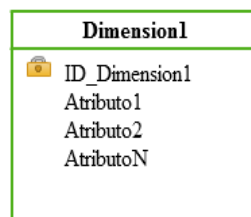
Ocurre cuando el elemento a medir es la ocurrencia de un evento más que una medida en específico, por lo que la tabla de hechos no tendrá hechos. Por ejemplo, en la inscripción de un estudiante no existe ninguna medida, lo que registra la tabla de hecho es la existencia del evento, es decir la inscripción de un estudiante. Por lo general las tablas de hechos tienen medidas porque los procesos de negocios a las que se les asocia tienen indicadores cuantitativos que pueden ser utilizados para medir la productividad de un área de negocio. (Kimball,2013,p.44).

- **Tabla dimensión.**

“Las tablas dimensionales son las integras compañeras de las tablas de hecho. Las tablas de dimensión contienen el contexto textual asociado con el evento medido en el proceso de negocio. Ellas describen el “quien, que, donde, cuando, como, y porque” asociado al evento” (Kimball,2013,p.13).

La cantidad de atributos que puede tener una tabla dimensión es variable, llegando a poseer decenas a solo unos pocos atributos, entre los cuales pueden existir atributos con grandes textos como contenido. A su vez dichos atributos pueden estar asociados lógicamente en función de una jerarquía.

Para cada elemento o fila de la tabla, se tiene una clave primaria que lo identifica unívocamente, lo cual es importante para realizar la posterior unión o relación con la tabla de hechos. En la figura 11 se puede ver la estructura básica de una tabla dimensión, la cual está conformada por un identificador y n campos.



**Figura 11** Tabla dimensión.

**Fuente:** Kimball, R. (1998). *The datawarehouse lifecycle toolkit* (1ª ed.). Estados Unidos de América:Wiley.

#### 2.4.10.- Tipos de tablas de hechos.

A partir de la forma en que se define la granularidad y la forma en cómo se obtendrán las medidas. Kimball(2013) establece tres tipos de tablas de hecho, descritas a continuación:

- **Tabla de hechos transaccional.**

Es este tipo de tabla de hecho, se mide la ocurrencia de un evento, por tanto, cada instancia de la tabla de hechos representa un evento. Debido a esto, la inserción de un registro suele contener todas las medidas y respectivas asociaciones con las dimensiones, por lo que cada instancia es atómica y completa, por lo que hecho representado posee desde un primer momento, la máxima capacidad de manipulación a partir de la granularidad establecida.

Este tipo de tabla de hecho es la más común. En el ejemplo de la inscripción de un alumno a la universidad, observamos como al insertar en la tabla de hechos un ingreso específico a partir de la ocurrencia y registro del mismo, poseemos una tabla de hechos transaccional, porque cada instancia de nuestra tabla de hechos representa un evento con todas sus

mediciones y dimensiones asociadas. Cuando la forma de levantamiento de requerimientos se hace a partir del modelo relacional, la tabla de hechos resultante suele ser una tabla de hecho de este tipo.

- **Tabla de hechos acumulativa.**

Esta tipo de tabla se crea cuando se requiere medir un hecho a través de un proceso. Debido a esta forma de medición, una instancia de la tabla no representa la realización un hecho “completo”, como por ejemplo una venta. La creación de una instancia por el contrario, representa el inicio de un proceso y las dimensiones existentes en el modelo, se irán completando y asociando a la tabla de hecho a medida que el proceso se realice.

Para esto generar una tabla de hecho acumulativa, es necesario que la medida se pueda describir a partir de un proceso finito, y que se tenga claramente definido el inicio y el fin del proceso. Si en el ejemplo de inscripción de estudiante, adicionalmente se desea medir si el estudiante se retira o se gradúa y si se gradúa en cuanto tiempo lo hace, una tabla de hechos de este tipo puede resultarnos adecuada, en el sentido de que la inserción de una instancia en la tabla de hecho representa la inscripción de un alumno (inicio del proceso) y la graduación o retiro de la universidad por parte del mismo representa el fin del proceso. Al crear una instancia, no se puede determinar el valor de todas las dimensiones que caracterizan un hecho ya que no se puede saber en el momento de inscripción del estudiante, el momento en el que él se graduara.

- **Tabla de hechos foto periódica.**

A diferencia de la tabla de hechos transaccional que mide un evento y la tabla de hechos que mide un proceso, este tipo de tabla de hechos busca medir en función de un periodo de tiempo determinado. En este sentido, se representa una suma de actividades en la instancia de una tabla de hechos. La diferencia entre este tipo y el tipo acumulativo, es que el acumulativo el tiempo para la agregación de una nueva fase dentro del proceso es indeterminado, mientras que en la foto periódica se define un tiempo en el cual se volverá a evaluar el hecho a medir.

Si se desease medir para una facultad, a partir del municipio que provienen los estudiantes, la cantidad de estudiantes existentes por semana, este tipo de tabla de hechos es la adecuada.

Porque permite tomar fotos periódicas de la cantidad de estudiantes existentes. Como se observa la medición no viene dada por la ocurrencia de un hecho o el seguimiento de un proceso, sino por un periodo de tiempo establecido en el que se busca medir un hecho.

#### **2.4.11.-Esquemas.**

Es el grupo de tablas y organización de las mismas que en conjunto conforman un modelo dimensional o un modelo relacional.

Dentro del área de almacenes de datos, existen un conjunto de esquemas de importancia que suelen ser utilizados para la implementación del modelo dimensional. Dichos esquemas son los siguientes.

- **Esquema estrella.**

Es una forma de disponer las tablas de un modelo dimensional, dentro de una base de datos relacional. El nombre de modelo estrella viene dado a que el modelo está formado por una tabla formada por una tabla de hecho que se relaciona con todas las tablas de dimensiones, lo que genera una figura similar a una estrella, esto porque el basamento del modelo estrella es la desnormalización de los datos generando dimensiones que individualmente se conectan con un evento particular.

Como se ve en la figura 12, se tiene una de hecho rendimiento, que tiene asociado un indicador, que es la nota. En función de este hecho, a partir de claves foráneas se asocian un conjunto de tablas de dimensión que describen esa nota y nos permite a su vez generar reportes. Por ejemplo, tener todas las notas del año 2014 a partir de la dimensión tiempo de la materia base de datos a partir de la dimensión materia.

La cantidad de dimensiones puede variar en función de las necesidades del problema, sin embargo, se recomienda no tener más de 15 dimensiones por un modelo estrella.

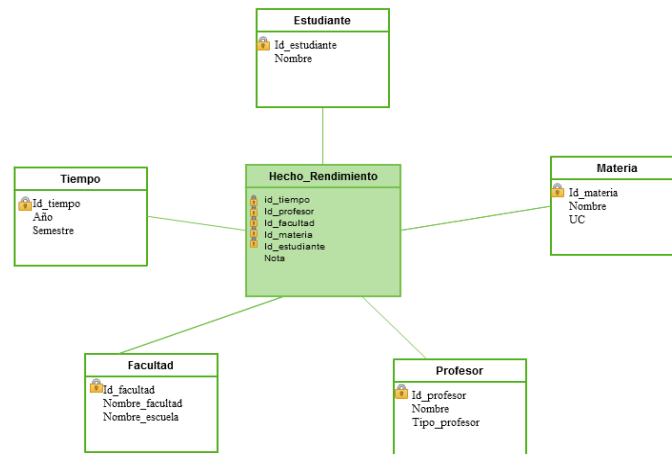


Figura 12 Esquema estrella.

- **Esquema constelación.**

El esquema constelación es la conformación de muchos esquemas estrellas unidos lógicamente. Su nombre viene dado por el término astronómico de constelación que hace referencia a un cúmulo de estrellas.

Un esquema constelación permite unir distintos procesos de negocios o áreas analíticas de una organización, a partir de dimensiones o elementos de la tabla de hechos (esto último poco frecuente) en común. Esto amplía la capacidad de generar reportes que reflejen una situación más general de la organización así como una mayor posibilidad de utilizar distintos conjuntos de datos para un análisis particular.

Como se ve en la figura 13, a partir de un conjunto de dimensiones (estudiante, tiempo, facultad) permite unir dos posibles modelos estrellas, uno que mide el rendimiento estudiantil con otra que mantiene el registro de admisión de estudiantes en la universidad para análisis de todo lo que son las propiedades de estos nuevos estudiante, es decir, cuando ingresaron, como ingresaron, de que liceo vienen y hacia que facultad van. Esta unión de modelos en un esquema mayor permite una mayor capacidad de análisis en el sentido que por ejemplo,

podría medir como es el rendimiento estudiantil de una materia dada en función de un liceo o de la forma en que ingreso el estudiante a la universidad.

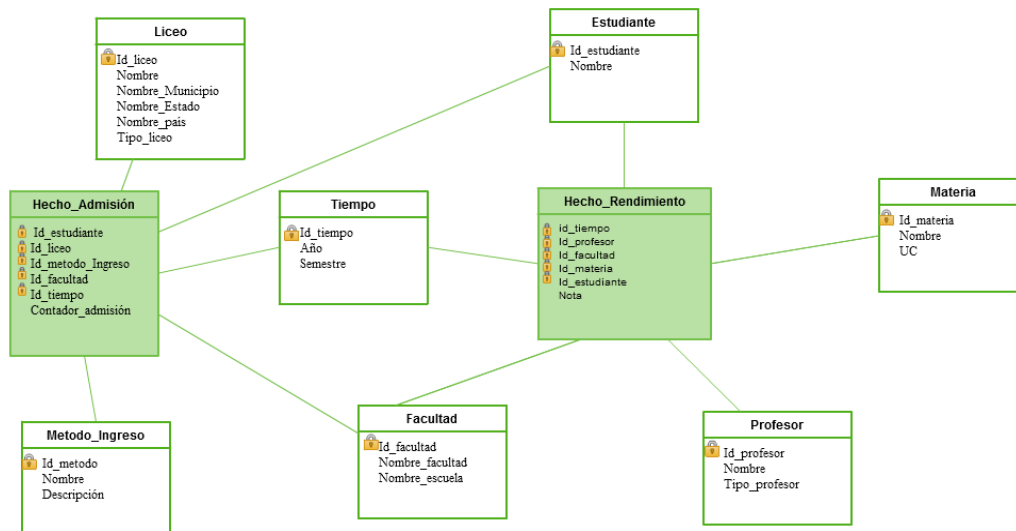
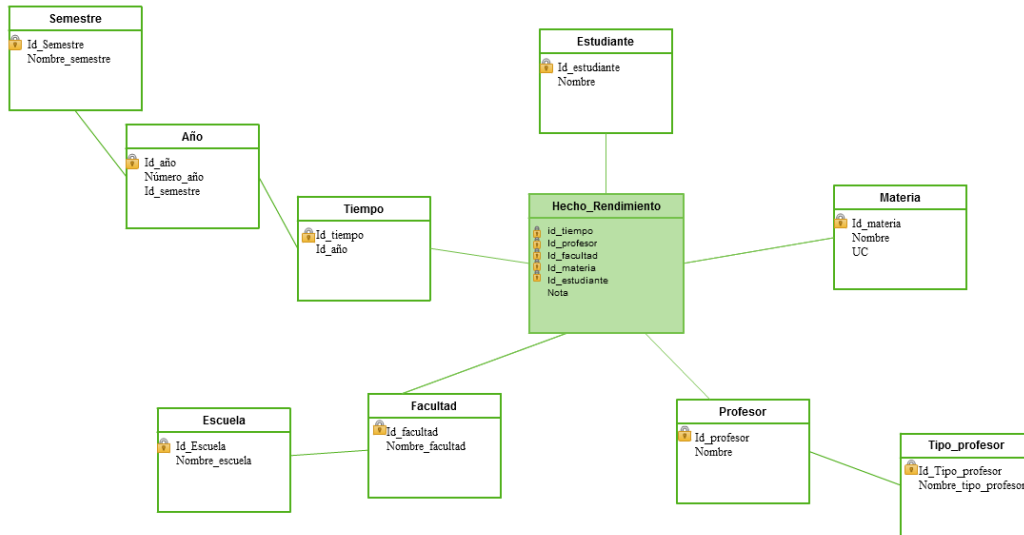


Figura 13 Modelo constelación.

- **Esquema copo de nieve.**

Es un esquema similar al esquema estrella con la diferencia de que a partir de una dimensión o “punta de la estrella” esta se extiende generando más dimensiones. Un hecho importante es que estas dimensiones no se van a unir con la tabla de hecho, por lo que tendremos relaciones entre dimensiones, algo que no ocurre en el modelo estrella.

El nombre de copo de nieve viene dado a la forma de un copo de nieve que se extiende en función de un punto central, en este caso la tabla de hechos. Por lo general estas dimensiones que se extienden vienen a representar la jerarquía o niveles de una dimensión específica, la cual, a diferencia del modelo estrella que se desnormaliza totalmente para formar una sola dimensión, esta jerarquía se extiende por varias tablas. Por lo que el esquema copo de nieve presenta un mayor grado de normalización que el modelo estrella.



**Figura 14 Modelo copo de nieve.**

Como se observa en la figura 14, el modelo previamente generado para el modelo estrella (figura 12) se le ha aplicado algunos procesos de normalización, lo que ha generado dimensiones adicionales como escuela o semestre.

#### **2.4.12.- Transición del modelo relacional al modelo dimensional.**

Kimball(2013) establece que la utilización de bases de sistemas manejadores de base de datos relacionales que soportan las operaciones transaccionales a partir de esquemas relacionales, implica que la implementación de un esquema dimensional es necesario una transformación del esquema relacional al esquema dimensional.

Un esquema relacional, por lo general suele implementar la tercera forma normal (3FN) para garantizar integridad y reducir la redundancia (mejorar la eficiencia en memoria), sin embargo los esquemas dimensionales buscan minimizar la cantidad de relaciones para simplificar el modelo así como aumentar el desempeño. La razón por la que el esquema relacional es deficiente en el procesamiento de consultas de lectura, es que para evitar la dependencia de datos en una misma tabla, se requiere construir muchas tablas que se relacionen entre sí. Realizar una consulta implica acceder a muchas tablas a partir procesos de unión (Join en inglés) para obtener el resultado, este proceso es costoso.

El esquema dimensional no necesita la 3FN porque parte de un esquema relacional que si esta normalizado, por lo que la integridad ya está comprobada. En el caso de la redundancia, los sistemas analíticos priorizan la velocidad de ejecución aunado al hecho de que los costos de memoria secundaria se ha visto reducido a través de los años.

Por tanto la transformación de un esquema dimensional a un esquema relacional implica el proceso inverso a la normalización, desnormalización. La desnormalización implica la creación de redundancia controlada en el esquema para aumentar su desempeño. Por ejemplo, en el caso de tener en 3FN una tabla para los datos de la ciudad y una tabla para los datos del país, donde existe una relación entre estas tablas, la desnormalización implica unir estas dos tablas. Esta unión hace que datos de la ciudad que dependen del país estén en la misma tabla (se pierde la 3FN), pero que a su vez, el procesamiento de una consulta que dependa de los datos de la ciudad y el país no requieran consultar sobre dos tablas.

Por lo general el proceso de desnormalización ubica a la base de datos en un estado de posibles inconsistencias, por lo que su utilización es solo recomendada para sistemas analíticos donde a través de un proceso de ETC se obtengan datos de una base de datos en donde efectivamente se garantice la consistencia en los datos.

## **2.5.- Inteligencia de negocios.**

### **2.5.1.- Definición.**

La búsqueda de utilizar la datos organizacionales para soportar decisiones complejas es una necesidad que se ha presentado desde mediados del siglo XX. La globalización, la producción en masa, mercados competitivos y la creación de computadores cada vez más sofisticados han posibilitado atacar esta necesidad.

Ya a finales de los años 50 Luhn (1959, p. 314) definía la inteligencia y los negocios como “Los negocios son una colección de actividades cuales sean para un propósito, sean ciencia, tecnología, comercio, industria, leyes [...]. La noción de inteligencia también es definida en un caso general como la habilidad de encontrar interrelaciones entre hechos presentes en una manera que guie las acciones a tomar para un objetivo específico.” Aunque Luhn en los años 50 no podía implantar una solución de inteligencia de negocios porque no existía la

arquitectura que soportara dicha implementación, la noción de negocio, como conjunto de tareas y de inteligencia, como la capacidad de generar análisis para una respuesta específica es un buen inicio para comprender la esencia de que es la inteligencia de negocios.

Un concepto más adecuado a la actualidad de inteligencia de negocio puede ser el dado por IBM en su página web como “la inteligencia de negocios significa usar la data para hacer mejores decisiones. Se trata del acceso, análisis, y descubrimiento de nuevas oportunidades”. La inteligencia de negocios va un paso más allá de los almacenes de datos, que como tal proveen una plataforma para tener datos de una forma adecuada para al análisis. Pero que la realización de estos análisis en dicho almacén de dato directamente resulta ineficiente.

Durante varias décadas, la inclusión de tecnologías más baratas a las organizaciones implicó la utilización de herramientas que fueron sofisticándose para tener la capacidad de analizar datos. Las organizaciones empezaron a acumular grandes cúmulos de datos y la complejidad de los modelos que soportaban esos datos hizo necesaria la utilización de sistemas adecuados para estas operaciones analíticas. Esto originó la creación de sistemas de procesamiento analíticos en línea (OLAP) y de almacenes de datos.

La inteligencia de negocios no busca reemplazar a los almacenes de datos. Por el contrario, es una evolución natural de los mismos, porque la generación de un almacén de datos se basaba en su construcción más que en su posterior acceso, lo cual hacía que muchas organizaciones no vieran la capacidad de realizar análisis con los mismos. La inteligencia de negocio complementa esto focalizándose en proveer información a partir de un conjunto de herramientas a los usuarios finales, en busca de hacer fácil el acceso a dicha información lo que puede ayudar entonces, con la ayuda de toma de decisiones de una forma asertiva.

### **2.5.2.- Características.**

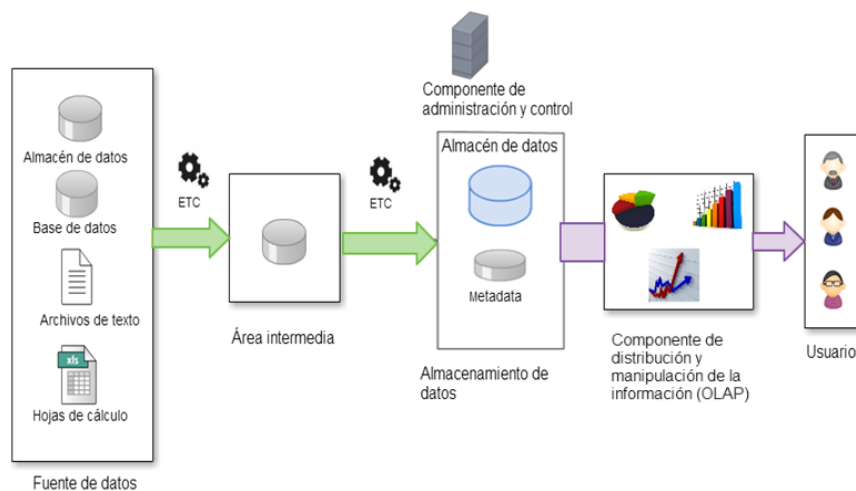
Cano(2007) establece que una Solución de inteligencia de negocios tiene que tener la capacidad para adaptarse a los distintos escenarios que pueden presentarse dentro de una organización, sin embargo, existen características en común que presentan cualquier solución de inteligencia de negocios. Entre estas podemos señalar:

- Responde preguntas básicas del negocio: Un sistema de inteligencia de negocio tiene que ser capaz de dar respuesta a preguntas simples dentro de la organización de una forma fácil. Dichas preguntas pueden ser por ejemplo, cantidad de ventas por mes, franquicia con más ingresos el trimestre pasado, cantidad de gastos de mantenimientos, etc.
- Responde preguntas analíticas/probabilísticas: Son preguntas más complejas que pueden surgir en una organización y que, el sistema de inteligencia de negocios debe ayudar a facilitar su respuesta. Ejemplo de estas preguntas son determinar las ventas del próximo trimestre, cuales son los productos que se compran juntos con más frecuencia, cuáles son mis productos claves, etc. La capacidad de brindar información por parte del sistema de inteligencia de negocio en estas interrogantes puede darle la capacidad a la organización de tomar ventajas frente a sus competidores y obtener un mejor desempeño.
- Provee información a cualquier usuario final: pese a que está orientado a la media y a la alta gerencia de una organización, los sistemas de inteligencia de negocios tienen que tener la capacidad de proveer información a los que otorgan información como a los que consumen dicha información, es decir, un gerente puede utilizar el sistema de inteligencia de negocios para visualizar datos, pero, el encargado de área que es el que está en el lugar donde se están generando los datos, también tiene que ser capaz de obtener información del sistema. Sin embargo, la cantidad de información que vera cada individuo dentro de la organización dependerá de su cargo y el acceso que tenga a la misma.
- Comunicar resultados: el sistema de inteligencia de negocios tiene que proveer la capacidad de informar a partir de situaciones establecidas, cualquier ocurrencia o anomalía a los encargados correspondientes. Esto no significa que un sistema de inteligencia de negocio no es un sistema de alerta, es decir, el sistema de inteligencia de negocio no está diseñado para informar si no hay producto en el almacén, sino está diseñado para informar si existe una desviación con respecto a la meta de ventas anuales para de esta forma, los encargados de ventas puedan evaluar la situación y tomar las medidas pertinentes.

- Manipulación de los datos: la capacidad que se tenga para comparar, sumarizar y analizar datos desde diferente ángulos aumentara el valor del sistema de negocio. Entendiendo que en muchos casos, la cantidad de datos que nos permitirá manipular el sistema puede generar análisis o comparaciones antes no contempladas, obteniendo patrones o comportamientos desconocidos. Por tanto, es importante la flexibilidad que nos provea el sistema de inteligencia de negocios con los datos que son utilizados. La existencia de un modelo robusto, poco flexible de visualización dificulta la capacidad de análisis.

### 2.5.3.- Arquitectura de una solución de inteligencia de negocios.

Basados en Cano (2007), la arquitectura para la implementación de una solución de inteligencia de negocios se basa en los siguientes componentes (figura 15).



**Figura 15 Componentes del almacén de datos.**

**Fuente:** Cano, J. L. (2007). *Business Intelligence: Competir con información*. España: ESADE.

### 2.5.4.- Fuente de datos.

Son aquellos repositorios que servirán para llenar de datos nuestro almacén de datos. Estas fuentes de datos pueden tener distintos formatos en las que se está guardando los datos requeridos, por lo que podemos tener archivos de textos, hojas de cálculo, datos guardados en base datos (relacional o no), etc. A su vez el tipo de dato a requerir para poblar nuestro almacén de datos puede variar según la naturaleza del problema.

Por lo general los datos de los sistemas operacionales y sus transacciones suelen ser la base para cualquier almacén de datos. Esto porque el almacén de datos suele resultar de la necesidad de operar análisis más complejos sobre los datos manejados en los sistemas transaccionales dentro de la organización.

También las organizaciones pueden poseer datos internos referida a los clientes, productos que no podría ser tomada en cuenta por los sistemas operacionales (porque son irrelevantes a la hora de realizar los procesos dentro de la organización) pero pueden ser importantes para realizar análisis. Por ejemplo, la información demográfica de un usuario, dentro de los sistemas operacionales puede ser tomada por protocolo a la hora de registrar un cliente, más dentro de estos sistemas, no representa una información útil. Pero para realizar un análisis de mercadeo de un producto según zonas geográficas, esta información es fundamental. Este tipo de dato complementa los datos operacionales y le agrega valor de análisis a la misma.

Otra forma de datos útil para usar como fuente de datos, puede ser todos los datos históricos que se guarde en la organización. Por lo general las organizaciones por ley o propia forma de trabajo, suelen tener datos almacenados por un período de varios meses o años. Dentro del problema o alcance del almacén de datos, hay que evaluar si estos datos pueden añadirle valor al almacén de datos.

Por último, la organización no es un ente aislado, sino que interactúa con otros sistemas, con otras organizaciones, con variaciones de precio, leyes, cambios climáticos, etc. Todos estos datos externos a la organización puede ser útil incorporarla al almacén de datos para realizar análisis a partir de los elementos externos que afectan a la organización.

#### **2.5.5.- Área intermedia.**

Antes de cargar los datos correspondientes a las distintas fuentes de datos en el almacén de datos. Estos datos tienen que ser cargados, por lo que debe realizarle procesos para integrarla para posteriormente ser cargada dentro del almacén de datos.

El almacén de datos está diseñado para soportar operaciones de aplicaciones que consulten datos para la realización de análisis. Estos procesos previos que tratan los datos deben por tanto ejecutarse en un lugar adicional, este lugar es el área intermedia. El área intermedia es

el puente entre un conjunto de fuentes de datos y los datos que realmente posee el almacén de datos.

En el área intermedia ocurre un proceso de extracción-transformación-carga (ETC). El proceso de ETC es crucial para el éxito del almacén de datos ya que permiten utilizar datos correctos para el análisis. Actualmente las casas de software tienen herramientas especializadas para facilitar el proceso de ETC entre las fuentes de datos y el almacén. En las fases del ETC se realiza lo siguiente:

- **Extracción.**

Consiste en la obtención de los datos desde las distintas fuentes de datos que soportan el almacén de datos. En esta fase es importante considerar la diversidad de las fuentes, en el sentido de la ubicación de las mismas, el formato en que se está almacenando los datos, que modelo soporta los datos que se están obteniendo, etc. Esta fase permite tener todo el conjunto de datos con los que se va a trabajar en un área común.

- **Transformación.**

Por lo general, las distintas fuentes de datos son generadas por distintos sistemas operacionales o aplicaciones. Esto implica que la forma en como fueron desarrollados dichos sistemas varíen, por lo que, siendo los datos la base en la cual estos modelos se desarrollan. El tipo, formato, características de los datos puede variar. Por ejemplo, dos sistemas pueden manejar el sexo de un cliente, sin embargo en un sistema se almacena en el formato “F” o “M” y en el otro sistema “masculino” o “femenino”. La inclusión de estas dos fuentes de datos para generar nuestro almacén de datos puede generar un inconveniente porque podría dar la impresión de que existen 4 tipos de sexo.

Todos estos problemas entre la compatibilidad de los datos entre distintas fuentes de datos o incluso, de una fuente de dato con nuestro almacén de datos son tratados en esta fase. Esta fase requiere un alto entendimiento de las fuentes de datos y sus características, así como de las características de nuestro almacén de datos para generar modificaciones en los datos que permitan refinar el dato a un punto que pueda ser utilizada para realizar análisis correctos dentro del almacén de datos.

- **Carga.**

Una vez generada las modificaciones pertinentes en los datos, se procede a mover estos datos al almacén de datos. Cuando son cargados los datos, el sistema manejador donde se encuentra nuestras bases de datos verificara la validez de los datos en referencia a si existe una correspondencia con los datos cargados y el modelo definido para el almacén de datos. Luego de este punto, los datos se encuentran en nuestro almacén de datos en una forma que es provechosa para la organización.

#### **2.5.6.- Almacenamiento de datos.**

En esta área se ubican los datos que están siendo manejados por el almacén de datos, estos esencialmente se dividen en dos. El almacén de datos, que refiere a los datos de la organización almacenada según el modelo dimensional definido. Y la metada, la cual refiere a un conjunto de datos que van a definir los datos en el almacén de datos.

En este sentido la metadata Definen las características del modelo (tablas y sus relaciones), los tipos de datos, nombre de los atributos, estadísticas sobre esos datos y demás elementos que van a darle significado a los datos almacenados dentro del almacén de datos.

#### **2.5.7.- Componente de administración y control.**

Es el sistema que se encarga de orquestar y gestionar todas las operaciones dentro del almacén de datos. Por tanto será el encargado de ejecutar los procesos ETC dentro del sistema, de monitorear el rendimiento, controlar el acceso sobre el almacén de datos, etc. A su vez puede utilizar información de la metadata para optimizar el rendimiento y las funciones de control sobre el almacén de datos (basándose en que la metadata contiene información del almacén de datos).

#### **2.5.8.- Componente de distribución de la información.**

Uno de los elementos más importantes en una solución es la capacidad de mostrar la información de una forma que permita agilizar el análisis por parte de los usuarios. Este componente está conformado por todas las aplicaciones o herramientas que sirven para desplegar la información del almacén de datos, a partir de una conexión entre esa aplicación

y el almacén de datos, entendiendo que dicha aplicación no necesariamente está ejecutándose en el mismo servidor donde se encuentra el almacén de datos.

La forma más básica de distribuir la información es a través de un reporte prediseñado, sin embargo las necesidades de análisis más complejos, han hecho que se creen herramientas que permitan soporte de operaciones estadísticas, de minería de datos, realización de reportes personalizados.

Según Cano (2007) las principales herramientas de la inteligencia de negocios para la presentación de datos son:

- Generadores de informe: herramientas para programadores profesionales para crear modelo de informes que serán utilizados por un grupo o área dentro de la organización.
- Herramientas de usuario final de consultas e informes: diseñadas para que los usuarios finales puedan generar o modificar informes predefinidos. No requiere que el usuario final tenga habilidades de programación.
- Herramientas OLAP: permite a los usuarios finales observar los datos finales desde una vista multidimensional (se suele buscar la manipulación de 3 dimensiones para generar un cubo). Facilita el análisis de datos desde distintas perspectivas.
- Herramientas de planificación, modelización y consolidación: diseñados para darle la capacidad a los usuarios finales de generar a partir de los datos elaborar planificaciones, levantar presupuestos o previsiones. Suelen necesitar de un conjunto de objetivos o métricas, por ejemplo el tiempo a calcular el presupuesto o la previsión de ganancia a partir de una región específica. Estos valores de entrada, son utilizados por la herramienta de inteligencia de negocios para generar las salidas correspondientes.
- Herramientas de minería de datos: la minería de datos es el proceso de manipulación de datos para la obtención de información útil para una organización. Por tanto esta herramienta permite la creación de modelos estadísticos a partir de los datos obtenida. Esto puede ser especialmente útil para conseguir patrones de consumo dentro de los usuarios, detectar fraudes financieros, etc.

- Herramienta de cuadro de mando: Suele presentarse como una interfaz con múltiples indicadores que muestran de forma rápida valores críticos dentro de la organización (figura 16). Tienen a su vez, la capacidad de, dado un indicador o gráfico, poder presentar más información de un elemento en específico si esto es requerido.

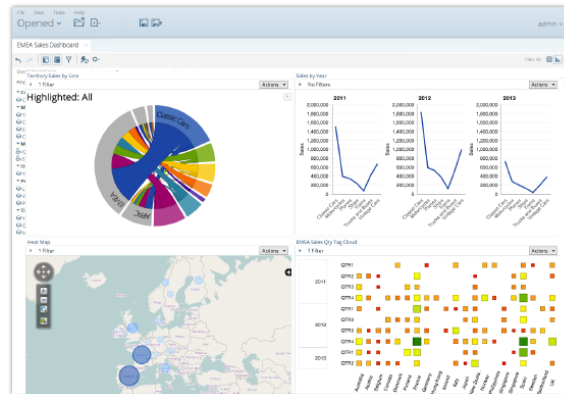


Figura 16 ejemplo de cuadro de datos por Pentaho BI Server.

Pentaho (s.f). *Pentaho*. Recuperado el 27 de enero de 2015 de: <http://www.pentaho.com/>.

### 2.5.9.- ¿Por qué se necesita inteligencia de negocios?.

Una solución de inteligencia de negocios permite aportar beneficios a distintos departamentos ayudando las diferentes operaciones que se realizan Cano (2007). Por tanto una solución de inteligencia de negocios es necesaria si:

- Se tienen datos que no son posibles analizarlos por la complejidad del modelo que los define, lo que genera que el tiempo preparando los datos para su análisis sea mayor que el tiempo que realmente se utiliza para analizarla.
- Es difícil establecer si los objetivos dentro de la organización son alcanzados. Sobre todo objetivos departamentales, regionales, etc. A los cuales son difíciles hacerles seguimiento.
- La realización de informes en hojas de cálculos es complicado por el cruzamiento que se necesita hacer sobre los datos, es decir, los datos tienen muchas dependencias y cuesta representarlos. Lo que hace que los analistas se enfrenten a una tarea complicada por la poca flexibilidad que poseen.

- Trabajo extra por parte de los empleados que requieren realizar informes o documentos de forma manual.
- Se pierden oportunidades de negocios. La organización tiene una lenta respuesta a los cambios y no logra mantenerse al nivel de las expectativas del mercado.

#### **2.5.10.- Beneficios de una solución de inteligencia de negocios.**

Los beneficios dentro de una solución de negocios pueden dividirse en tres grandes grupos Canno(2007) estos son:

- **Beneficios tangibles:** son aquellos beneficios que pueden ser medidos y que representan una mejora dentro de la organización de forma directa. Ejemplo de estos beneficios se dan en reducción de costos, al evaluar fallos en el manejo de inventario, exceso de personal, procedimientos innecesarios, proveedores ineficientes o costosos. Mejorando los ingresos al ayudar a establecer nuevos mercados, productos, promociones.
- **Beneficios intangibles:** el acceso a la información de forma más fácil y sencilla a usuarios finales dentro de la organización, pueden ayudar a la competitividad de la organización. Estos beneficios ayudan a generar beneficios tangibles dentro de la organización, es decir, aumentan la rentabilidad de la organización porque se mejoran ciertos procesos dentro de la misma. Estos beneficios se presentan en mejora del trato con los clientes, aumento de la satisfacción de los clientes, mayor facilidad para realizar informes y procesos analíticos, mayor integración de la información, conseguir ventajas competitivas, etc.
- **Beneficios estratégicos:** Aquellos que permiten facilitar las decisiones estratégicas dentro de la organización, estas son, las decisiones que definen las operaciones a seguir dentro de la organización para obtener beneficios tangibles. Esto significa, mejorar la forma en que se toman las decisiones con respecto a nuevos mercados, precios de productos, políticas de ventas, etc. Haciendo que dichas decisiones se tomen de forma oportuna, en el momento exacto, así como teniendo la capacidad de monitorear las decisiones tomadas y su impacto dentro de la organización para tener la capacidad de ajustar cualquier elemento-variable si hace falta.

### **2.5.11.- ¿Qué se requiere para una solución de inteligencia de negocios?.**

Para aplicar una solución de inteligencia de negocios, según Almeida, Ishikawy, Reinschmidt (1999) dentro de la organización es necesario:

- Soporte de una aplicación preestablecida que sea una solución de inteligencia de negocios. Si bien una solución de inteligencia de negocios puede ser creada desde 0 por una organización, lo recomendable es usar una de las soluciones existentes en el mercado puesto que las mismas suelen presentar solidez, facilidad y rapidez para su implementación y buena capacidad adaptativa a las distintas situaciones que pueden presentarse dentro de la organización.
- Una solución que represente un buen balance costo-beneficio para la organización, que permita su viabilidad y a su vez aumentar la capacidad competitiva de la organización.
- Un acceso rápido y sencillo a la información operacional de la organización, involucrando a la mayor cantidad de usuarios finales (departamentos).
- Tener soporte para el manejo de las nuevas tecnologías de la información, incluido técnicas de análisis de datos y sistemas de procesamiento analítico en línea (OLAP).
- Un ambiente abierto con capacidades de escalar. Esto implica poder hacer crecer nuestra solución agregando nuevos módulos, ya sea porque se necesitan nuevos tipos de informes o análisis o porque se requiere la incorporación de nuevas unidades de negocios.

### **2.5.12.- Operadores de refinamiento OLAP.**

Cano(2007) habla de los operadores de refinamiento o manipuladores de consultas dentro de un sistema OLAP como operaciones que brindan una gran herramienta para realizar análisis sobre un conjunto de datos. Estos operadores esencialmente permitirán ajustar nuestros informes según requerimientos específicos permitiendo así, que el sistema tenga alta capacidad de respuesta a los distintos análisis que puedan plantearse dentro de la organización.

- **ROLL-UP.**

Es una operación de agregación sobre los datos o dimensiones en las que se está realizando un cubo. En función de las jerarquías establecidas en las dimensiones, se puede subir el nivel de granularidad, moviéndonos a través de estos niveles de una jerarquía, para generar un informe con un menor nivel de detalle pero que englobe más datos.

Una operación de agregación implica que el hecho que está siendo reflejado en el cubo, tiene que realizársele un nuevo cálculo de agregación, esto en función de solo las dimensiones que hayan sido reducidas en la operación roll-up, esto si el hecho puede ser sumariado con otros hechos, como monto de ventas, unidades vendidas, etc. Por ejemplo, si observamos en la figura 17, se tiene un cubo OLAP que muestra los ingresos de estudiantes por facultad, a partir del municipio (ubicación del liceo donde estudió el estudiante) en función del momento donde se realizó el ingreso. La operación de roll-up que realiza una agregación sobre la dimensión tiempo, implica un recalcule del hecho sumando el valor de cada semestre para generar un nuevo valor, No se requiere tener que recalcular nuevamente en función de las facultades, porque esta dimensión no ha variado.

- **DRILL-DOWN.**

Es la operación opuesta a Roll-up, es decir, es una operación de desagregación sobre las dimensiones a partir de las jerarquías que existen en las mismas. Esto disminuye el nivel de granularidad en el cubo, lo que permite una mayor capacidad de análisis detallados sobre un requerimiento específico puesto que se ven los hechos en función de un menor nivel de agregación.

Al igual que en el roll-up, una operación de desagregación requiere un recalcule de los hechos, sin embargo, este recalcule si implica una nueva evaluación del hecho en función de todas las dimensiones, es decir, si observamos la figura 17, realizar una operación de drill-down, pasando año a semestre, implica que el sistema tiene que calcular los valores para cada semestre en función de la facultad y municipio, puesto que si el sistema no tiene guardado en memoria cache, es imposible para él saber cómo ocurre la subdivisión de las cantidades.

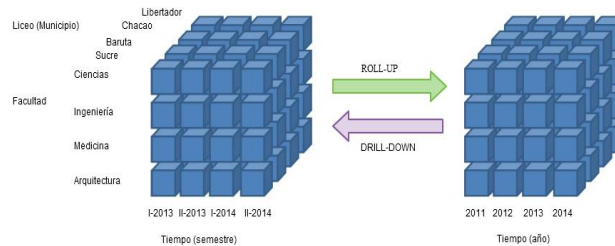


Figura 17 operadores ROLL-UP y DRILL-DOWN.

- **DRILL-ACROSS.**

Implica la agregación de una nueva dimensión al informe o cubo que se esté realizando. En este sentido, el informe se va a disgregar o generar un mayor nivel de detalle, no por movernos dentro de una jerarquía, sino porque la inclusión de la nueva dimensión va a generar subgrupos, como se puede ver en la figura 18, la agregación de la dimensión facultad, crea un mayor detalle (disgregación de los datos) puesto que, por cada municipio y semestre, ahora existirán subconjuntos conformados por cada facultad. Implica un recalcule total de los hechos.

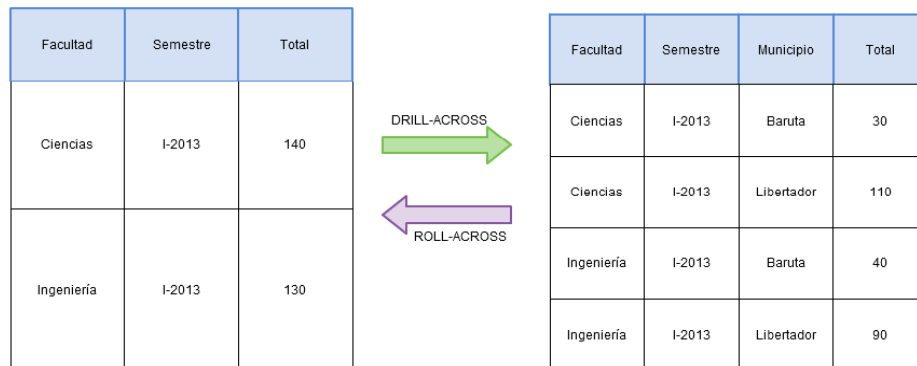


Figura 18 Operadores ROLL-DRILL ACROSS.

- **ROLL-ACROSS.**

Es la operación contraria al DRILL-ACROSS, implica la eliminación de una o más dimensiones dentro de un cubo. Esta eliminación elimina subconjuntos de datos (figura 18), lo que genera una agregación de los mismos. Al igual que con Roll-up, si el hecho es un valor que se puede sumarizar, la aplicación de un roll-across no implica un recalcule total en función de las dimensiones restantes.

- **SLICE.**

Consiste en la reducción de la dimensionalidad del cubo generado a partir de una selección dentro de una de las dimensiones del mismo. Esto se realiza manteniendo fijo el valor de una dimensión, mientras las otras dimensiones varían en función de esta. En la figura 19, se ha tomado de la dimensión tiempo (semestre), el semestre (I-2014), por lo que la misma solo tendrá este valor fijo, mientras las otras dimensiones variaran sus valores mostrando todas las métricas respectivas.

- **DICE.**

Consiste en la generación de un sub-cubo a partir de la selección en una o más dimensiones. Si se realiza sobre una sola dimensión fijando un solo valor para la misma, es un slice. Debido a que se está realizando una selección (figura 19), se genera un cubo de menor tamaño pues se están reduciendo el número de miembros para las dimensiones que se les ha aplicado la selección. La operación dice, permite a partir de un cubo, ubicar el análisis en una parte específica del mismo, lo que ayuda al proceso de abstracción de los usuarios.

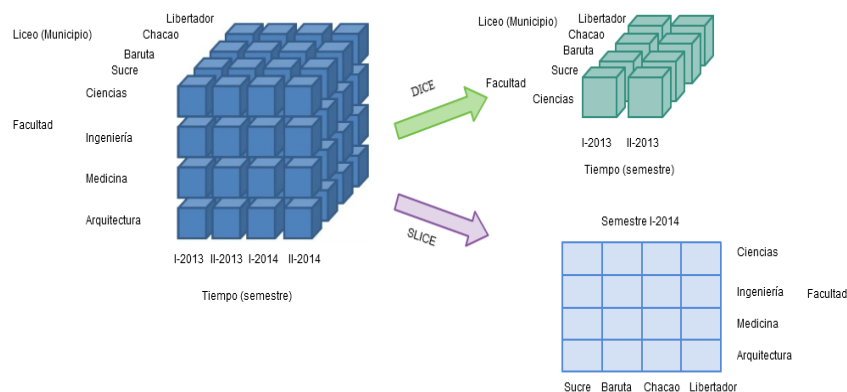


Figura 19 Operadores SLICE y DICE.

- **PIVOT.**

Permite modificar la ubicación de las dimensiones dentro de nuestro cubo, generando un nuevo cubo. Este reajuste de posiciones permite cambiar la forma en como los datos son mostrados, lo que puede generar una mayor comprensión de las métricas existentes dentro del cubo, facilitando la toma de decisiones. Como observamos en la figura 20, aunque los datos del cubo son los mismos, la reubicación de las dimensiones hace que la visualización de los datos cambie.

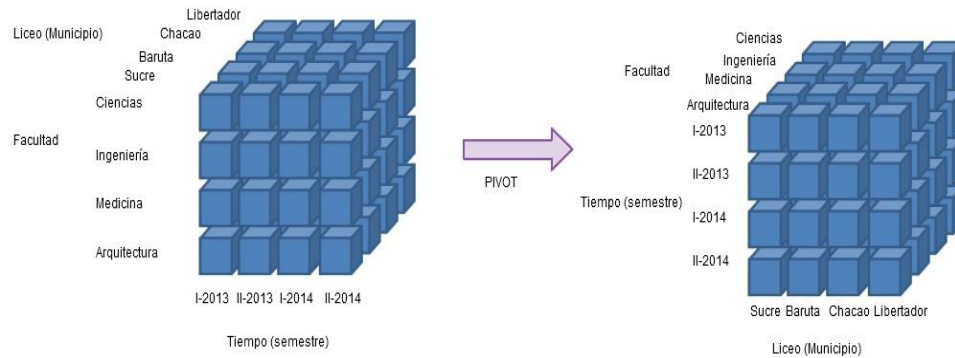


Figura 20 Operador PIVOT.

## 2.6.- Herramientas de inteligencia de negocio.

La utilización de una herramienta de inteligencia de negocio de una casa de software puede ayudar al éxito de un proyecto de inteligencia de negocio. Gartner (2014). Gartner es una consultora encargada de estudiar soluciones de software, en la figura 21, se observa la comparación de las distintas herramientas tomando como medidas la habilidad de ejecución y la complejidad del negocio.



Figura 21 Cuadrante mágico de Gartner.

Fuente: Gartner (2014). Gartner Magic Quadrant .Recuperado el 15 de enero de 2015 de:  
[http://www.gartner.com/technology/research/methodologies/research\\_mq.jsp](http://www.gartner.com/technology/research/methodologies/research_mq.jsp).

Una solución de inteligencia de negocio son un conjunto de aplicaciones empaquetadas, las cuales están orientadas a adaptarse a ciertas características o necesidades organizacionales. Por esto, tendremos una variedad de herramientas de inteligencia de negocio en el mercado, entre las cuales se resaltan:

- **Tableau.**

Es una plataforma de inteligencia de negocios conformada esencialmente por cinco módulos principales que proveen todas las funcionalidades necesarias básicas para una solución de inteligencia de negocios. La empresa que fabrica dicha solución lleva su mismo nombre y fue creada en el 2003 en California, Estados Unidos. Es una de las soluciones de negocio con mayor crecimiento en el área. Su crecimiento en los últimos años ha sido notorio, de forma que en el informe de Gartner de soluciones de inteligencia de negocios el cual es uno de los informes más utilizados para medir el impacto de una solución o compañía en un área específica dentro de las tecnologías de la información, en el 2014, fue la solución líder dentro del área de inteligencia de negocio (figura 21). Parte de su éxito se debe a la fácil implementación de su solución, costos reducidos de licencia, velocidades de ejecución, así como una rápida realización y puesta en producción de métodos de acceso a la información, lo que permite obtener rápidamente un retorno a la inversión.

- **Qlik.**

Es una empresa fundada en 1993 en Pennsylvania Estados Unidos que se dedica a construir soluciones de inteligencia de negocios y visualización de datos. La base de la construcción de sus soluciones es la de generar herramientas que puedan ser puestas en producción rápidamente, abaratando así costos en elaboración. Al igual que Tableau, según Gartner en el cuadrante de soluciones de inteligencia de negocios (figura 21), es una herramienta con facilidad para ser ejecutada.

- **Pentaho Open Source Business Intelligence.**

Son un conjunto de herramientas de inteligencia de negocios bajo filosofía de software libre desarrolladas por la corporación pentaho, compañía fundada en 2004 en Orlando, Estados Unidos. Es una solución que posee un conjunto de componentes que permiten la elaboración de cuadros de mando, análisis de minería de datos, presentación de informes flexibles, etc.

Toda la solución de Pentaho está desarrollada en Java, por lo cual es altamente portable, es decir, tiene flexibilidad de ejecutarse en distintos ambientes.

- **Oracle Business Intelligence.**

Es una solución de inteligencia de negocios creada por la corporación Oracle. Oracle es una de las compañías de desarrollo de software gerencial más grande del mundo. Creada en 1977 en California Estados Unidos, inicialmente Oracle se basaba en la construcción de sistemas manejadores de bases de datos, sin embargo a través de los años fueron generando (o adquiriendo a través de la compra de compañías) distintas soluciones de software para abarcar todo los sistemas requeridos en una organización.

La solución de Oracle para inteligencia de negocios en un conjunto de herramientas agrupadas bajo la plataforma Oracle Business Intelligence Standart Edition One.

En la tabla 1 puede verse una comparación entre las herramientas de solución descritas previamente.

Solución	Licencia	ETC	Orientación	Herramientas de manipulación de datos	Costos	Habilidad de ejecución	Complejidad de la visión del negocio
<b>Qlik</b>	Comunitaria y privada.	No.	Pequeña y medianas empresas.	Si	Moderados.	Alta.	Alta.
<b>Tableau</b>	Comunitaria y privada.	No.	Pequeña y medianas empresas.	Si	Moderados.	Alta.	Alta.
<b>Pentaho</b>	Comunitaria y privada.	Si.	Pequeña y medianas empresas.	Si	Bajos.	Limitada.	Moderada.
<b>Oracle</b>	Privada.	Si.	Grandes corporaciones.	Si	Altos.	Limitada.	Alta.

**Tabla 1 Comparación entre las soluciones de inteligencia de negocios.**

**Fuente: Gartner (2014). Gartner Magic Quadrant .Recuperado el 15 de enero de 2015 de:**  
[http://www.gartner.com/technology/research/methodologies/research\\_mq.jsp](http://www.gartner.com/technology/research/methodologies/research_mq.jsp).

### **2.6.1- Herramientas a utilizar.**

Las herramientas a utilizar vienen dadas de 3 organizaciones distintas (Oracle, Pentaho, Qlik). En donde Oracle es utilizado como motor de base de datos, Pentaho para la realización

de los procesos de ETC y Qlik como herramienta de manipulación y distribución de la información.

A continuación serán descritas herramientas que serán utilizadas para generar la solución de inteligencia de negocios:

- Oracle Database

Es un manejador de bases de datos relacionales de amplio uso en el mercado. Se caracteriza por ser de fácil instalación, ser completa, tener capacidad de generar respaldos, administración automatizada de memoria, soporte para distintas fuentes de datos (xml, archivos de texto, imágenes) y de poder ser instalada en distintos sistemas operativos. Dentro de la solución de Oracle bussines intelligence, Oracle Database es la base de datos encargada de registrar todos los datos de los procesos operacionales que ocurren dentro de la organización.

- Pentaho data integration (Kettle).

Herramienta que permite la realización de procesos de extracción-transformación-carga (ETC). Permite a través de una interfaz gráfica establecer procedimientos ETC sobre el conjunto de datos, esta herramienta puede facilitar el desarrollo de almacenes de datos, lo que implícitamente ayuda a la construcción de soluciones de datos. En la figura 22 se puede observar la interfaz gráfica que nos provee Pentaho data Integrator para la manipulación de las distintas fuentes de datos y definición de procesos de ETC.

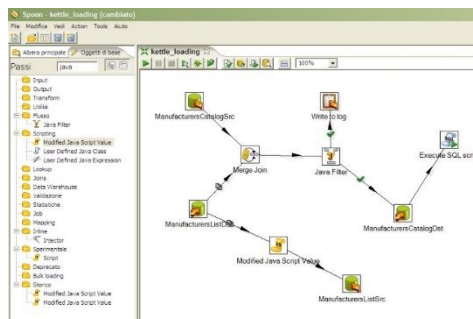


Figura 22 Pentaho Data Integrator.

Fuente: Pentaho (s.f). *Pentaho*. Recuperado el 27 de enero de 2015 de: <http://www.pentaho.com/>.

- Qlikview.

Es la herramienta principal de Qlik para el área de inteligencia de negocios. Permite un fácil manejo de los datos a partir de la conexión a múltiples fuentes de datos. A parte de su interfaz sencilla y usable posee dos características significativas. La primera es que carga todos los datos así como la configuración del cuadro de mando en un archivo tipo .qw, este archivo posee un alto factor de compresión (hasta 10 veces el tamaño original) y permite que la solución pueda ser ejecutada fácilmente en otro ordenador puesto que se elimina la dependencia a la fuente de datos.

La segunda característica es que se basa en el almacenamiento en memoria principal para una rápida ejecución de las consultas. Esencialmente carga todos los datos y sus asociaciones lógicas en memoria principal (ayudándose de su factor de compresión), basan este concepto en que la velocidad de respuesta es fundamental y los bajos precios de memoria permiten una fácil escalabilidad. Desde la versión 11, Qlikview puede hacer paginación de sus datos, lo que elimino el problema de manejar datos que fueran mas grandes que la capacidad de memoria disponible. En la figura 23, se ve un ejemplo de un cuadro de mando creado por Qlikview.

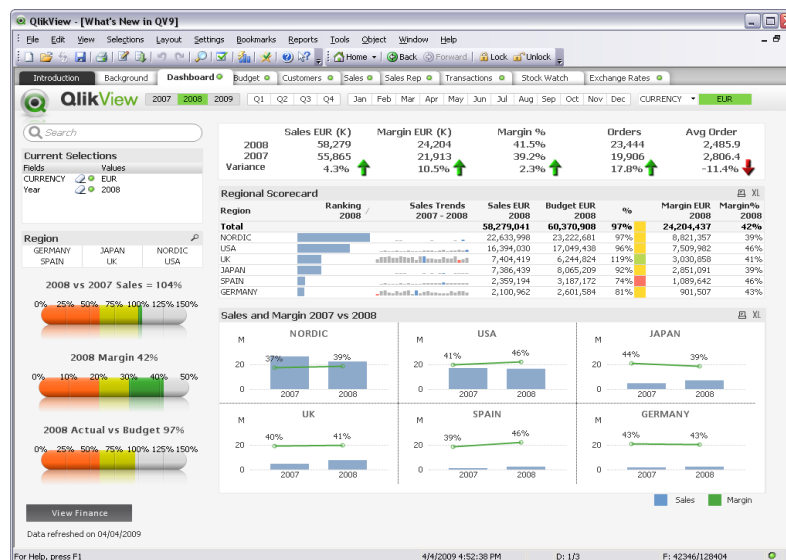


Figura 23 Ejemplo cuadro de mando Qlikview.

Fuente: Qlik. (s.f). Qlik Products. Recuperado el 27 de enero de 2015 de:  
<http://www.qlik.com/us/explore/products/>.

## 2.7.- Indicadores de gestión.

### 2.7.1.-Definición.

“Un indicador de gestión es la expresión cuantitativa del comportamiento y desempeño de un proceso, cuya magnitud, al ser comparada con algún nivel de referencia, puede estar señalando una desviación sobre la cual se toman acciones correctivas o preventivas según el caso.” (Jaramillo, 2013, p1). El concepto de indicador de gestión tiene mucha importancia en los proyectos de inteligencia de negocio, porque ellos implican la esencia de los mismos, esto es, la utilización de métricas para evaluar desempeño en función de mejorar este desempeño a través de la cuantificación de los resultados. Basados en el modelo estrella, los indicadores de gestión vienen reflejados por los hechos que existen en la tabla de hechos, siendo las dimensiones las que van a caracterizar estos indicadores de gestión.

Jaramillo (2013, p1) también establece que “Empleándolos en forma oportuna y actualizada, los indicadores permiten tener control adecuado sobre una situación dada; la principal razón de su importancia radica en que es posible predecir y actuar con base en las tendencias positivas o negativas observadas en su desempeño global.”. En este sentido varios puntos se pueden destacar de los indicadores de gestión. Primeramente permiten a la toma de decisiones, y una de las bases de los sistemas para la toma de decisiones son los proyectos de inteligencia de negocio. Permiten realizar predicciones porque se pueden establecer predicciones del valor que tomara el indicador en momento de tiempo. Su valor depende de la forma en que estos se muestren, siendo actualizados y de fácil comprensión para los analistas de negocios.

### 2.7.2.- Características de los indicadores de gestión.

Dentro de las características de los indicadores de gestión Rozo (2013, p95) establece las siguientes como básicas para que un indicador de gestión sea efectivo:

- **Simplicidad:** el indicador debe ser sencillo, de fácil comprensión y que permita medir de forma clara y oportuna un evento.
- **Adecuación:** el indicador debe tener la capacidad de adaptarse a la arquitectura organizacional para poder representar un evento de forma correcta

- **Validez en el tiempo:** un indicador de gestión debe ser permanente en el tiempo o en su defecto, establecer el período de tiempo durante el cual el mismo es válido.
- **Participación de los usuarios:** Los usuarios deben participar activamente en la creación del indicador de gestión, comprendiendo que los usuarios de negocios comprenden las necesidades analíticas de la organización y son ellos los que requieren de estos indicadores para tomar decisiones.
- **Oportunidad:** capacidad de que los datos que son representados por los indicadores sean recolectados de forma oportuna, es decir, la capacidad del indicador de estar disponible para su uso en el momento adecuado.

### 2.7.3.- Elementos de los indicadores de gestión.

Jaramillo (2013, p4) establece varios componentes que debe poseer un indicador de gestión. A continuación se describen dichos componentes:

- **Definición:** la descripción del indicador de gestión
- **Objetivo:** cuál es la razón de ser del indicador, lo que implica que se tiene que establecer qué valor va a tener la generación de ese indicador para la organización, por ejemplo a partir de un indicador que mida los gastos en productos de limpieza por sucursal dentro de una cadena de comida, optimizar el uso de productos de limpieza para reducir costos.
- **Valores de referencia:** debido a que un indicador de gestión es un valor cuantificable, el mismo debe de tener un punto con el que comparar para evaluarlo. El valor de referencia nos permite establecer la situación de la organización a partir del valor de los indicadores en un momento dado. Para obtener un valor de referencia se puede:
  - utilizar valores históricos de ese indicador.
  - valores estándares o teóricos, es decir cuál es el valor que el sistema debería tener en función de las características del sistema.
  - Valores determinado por los usuarios.
  - Valores por competencia, lo cual implica utilizar como referencia el desempeño de un competidor que posea una estructura similar dentro de su organización.

- Responsabilidad: es la forma en como el indicador será utilizado.
- Puntos de medición: la forma como el indicador será medido, esto significa establecer la fuente de los datos a utilizar y como los datos serán utilizados a partir de operaciones de agregación para generar el indicador.
- Periodicidad: cada cuanto tiempo el indicador será recalculado.
- El sistema de procesamiento y toma de decisiones: implica establecer cuáles serán los sistemas que servirán para generar datos a los indicadores y luego como estos indicadores serán mostrados a los usuarios.

#### **2.7.4.- Beneficios de los indicadores de gestión.**

Según Jaramillo (2013) se pueden destacar muchos beneficios si se aplica un sistema donde efectivamente los indicadores de gestión son utilizados por la organización para la toma de decisiones. Uno de los beneficios más importantes de los indicadores de gestión es que permiten medir el rendimiento de la organización, a partir de la medición se pueden generar cambios, evaluar esos cambios y establecer la efectividad de los mismos. Además los indicadores de gestión permiten la toma de decisiones basadas en hechos y no en elementos subjetivos, es decir, la toma de una decisión se fundamenta en un rendimiento, una tendencia medida dentro de la organización.

Por tanto los indicadores de gestión son la base para el monitoreo de procesos de negocios dentro de la organización, la evaluación de su desempeño con la competencia para posteriormente hacer una eficiente gerencia de cambios dentro de la organización que permitan optimizar los procesos dentro de la misma.

#### **2.8.- Proceso de seguridad, higiene y ambiente (SHA).**

##### **2.8.1.- Definición.**

Según Chiavenato (2000) Los procesos involucrados al área de SHA existen en aquellas organizaciones donde las características de sus procesos de negocio requieren que la misma posea una unidad encargada de toda el área de seguridad, higiene y ambiente. Por lo general dichas organizaciones realizan manipulación de sustancias tóxicas, contaminantes o

inflamables que pueden perjudicar al personal o los bienes de la empresa así como el entorno donde se desarrollan estas actividades.

Debido a que los procesos de SHA no existe en todas las organizaciones y que depende mucho del área de negocio donde la organización realice sus operaciones, las características y formas de trabajo de dicho departamento va a tener una gran variabilidad, es decir, los mecanismos de un departamento en SHA para el área de petróleo serán totalmente distintos a un departamento SHA en el área de realización de productos farmacológicos. Esta variación no es tan acentuada en otros departamentos, como el de cobranza o de manejo de proveedores, donde, aunque existen elementos propios dentro de la organización o área de negocio, los procesos de negocios dentro de todas las organizaciones tienen una estrecha similitud y pueden generalizarse.

### **2.8.2.- Funciones involucradas en los procesos de SHA.**

Las funciones en los procesos de SHA van a variar mucho entre organizaciones, sin embargo existen operaciones en común que realizan estos procesos Pérez(2011). Dichas funciones son expuestas a continuación:

- Revisar y aprobar las políticas de seguridad: se encarga de evaluar todos los mecanismos o normativas que posee la organización.
- Realizar inspecciones periódicas de seguridad: dentro de estas inspecciones, el área de SHA se encarga de establecer la periodicidad, tipo y forma de evaluar los distintos componentes dentro de la organización, esto puede incluir partes mecánicas o evaluación del personal.
- Establecer normas adecuadas de seguridad, deben concordar con las disposiciones legales: dentro del proceso de SHA se tienen que alinear las políticas de seguridad con estándares establecidos en la industria o en el país donde se estén desarrollando las actividades.
- Poner en funcionamiento y mejorar el programa de seguridad: a partir de un plan base de seguridad, se realizan evaluaciones para su mejora.

- Asesorarse sobre problema de seguridad: el área SHA es la encargada de buscar cualquier material, instrucción o elemento que permita mejorar la seguridad dentro de la organización.
- Ocuparse del control de las enfermedades ocupacionales: este control debe darse en áreas de negocio donde el trabajo del personal pueda suponer la generación de cuadros clínicos, por ejemplo, enfermedades cutáneas por el manejo de materiales irritantes, intoxicación por gases tóxicos entre otros.
- Asesorarse sobre problemas del medio ambiente: las operaciones de una organización pueden dañar el medio ambiente, la erosión de la tierra, la generación de desechos tóxicos o la contaminación del agua pueden ser algunos de los problemas asociados al desarrollo de las operaciones de una organización. El área de SHA debe buscar toda la asesoría necesaria así como realizar todos los mecanismos posibles para evitar que las operaciones de la organización perturben el ambiente.
- Ejecutar plan de primeros auxilios: implica la realización de un plan de primeros auxilios y si posible ejecución en caso de que surja una eventualidad en la organización.

### **2.8.3.- Disposiciones legales asociadas al departamento SHA en Venezuela.**

El departamento de SHA suele estar ligado a las normativas que imperan en el país donde se realizan las operaciones, por tanto, aunque se trate de una empresa transnacional, dicha organizaciones tiene que adaptar sus normativas a las establecidas en el país.

Para Venezuela se tienen distintas leyes para el área de seguridad, higiene y ambiente, las cuales se describen a continuación:

- Ley Orgánica de Prevención, Condiciones y Medio Ambiente de Trabajo (LOPCYMAT).

Es la ley principal a lo referido con seguridad e higiene laboral, es una ley que se encarga de desarrollar los derechos de los trabajadores, como poseer adecuadas condiciones de seguridad y medio ambiente de trabajo.

- Ley orgánica del ambiente.

Conjunto de disposiciones legales que enmarcan delitos y sus respectivas sanciones, que pueden ser cometidos por una persona natural o jurídica a los espacios naturales del país.

- Código penal venezolano y ley orgánica del trabajo.

Hace referencia a las condiciones básicas de seguridad e higiene en donde el trabajador puede desempeñar sus labores.

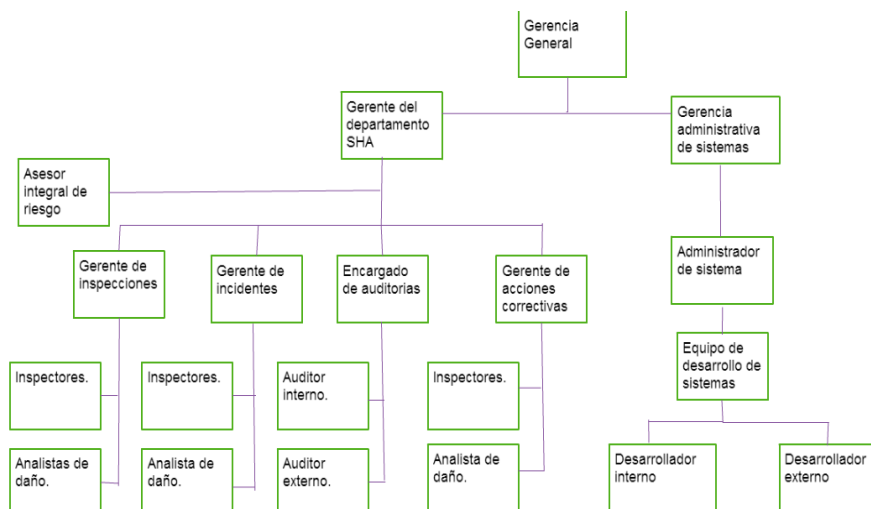
En materia del medio ambiente también existen una serie de tratados y convenios internacionales sobre el tratado de residuos específicos, como el Convenio sobre la Prevención de la Contaminación del Mar por Vertimiento de Desechos y otras Materias, el convenio para la Protección de la Capa de Ozono, etc.

En el sentido de seguridad industrial, La Organización Internacional del Trabajo (OTI) promueve convenios y tratados que busquen dar condiciones dignas y seguras de tratado.

#### **2.8.4.- Organigrama del departamento SHA.**

El organigrama de departamento o área de seguridad, higiene y ambiente va a depender del tipo de organización donde exista y por tanto sus distintos elementos se adaptarán a satisfacer las necesidades específicas de la empresa donde se ubique.

Sin embargo, los departamentos de SHA suelen tener una estructura similar, descrita en la figura 24. A partir del gerente del departamento, existen un conjunto de gerencias particulares para cubrir las distintas funciones existentes, estas funciones varían y abarcan esencialmente gerencias asociadas a las inspecciones de equipo, simulacros, registro de incidentes, acciones correctivas, etc. Cada gerencia posee inspectores que son los que realizan el registro de equipos u operaciones asociadas al departamento SHA y analistas de daños que son los encargados de establecer el impacto de un evento en la organización. Dentro de todas estas gerencias, dos de las mismas tienen una estructura particular. La gerencia de auditoria y la gerencia de sistemas.



**Figura 24 Organigrama típico del departamento SHA.**

**Fuente: Pérez Jaramillo, Carlos Mario (2013). Los indicadores de gestión. Recuperado el 7 de marzo de 2015 de: <http://www.escuelagobierno.org/inputs/los%20indicadores%20de%20gestion.pdf>.**

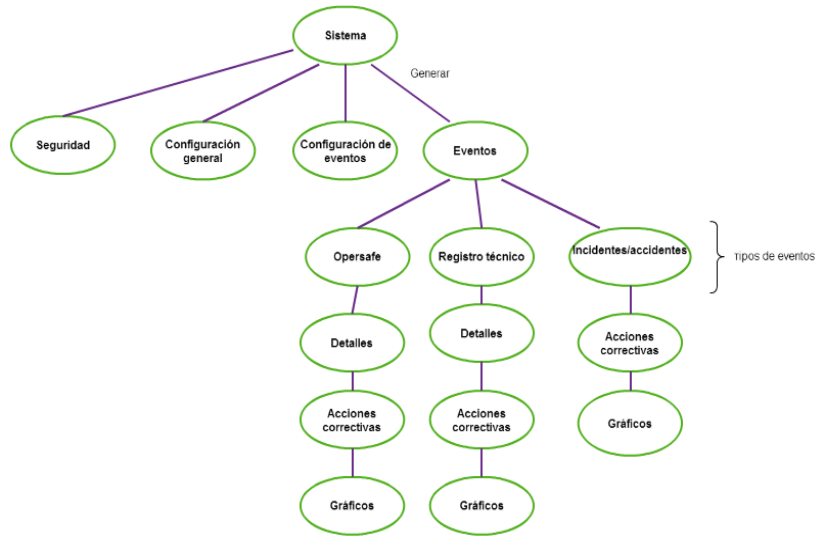
Para las auditorías, se subdivide en auditores internos, propios de la organización y auditores externos que son contratados para realizar auditorías en un momento específico.

Para la gerencia administrativa de sistemas, existe un gerente de sistema, un administrador de sistema y un equipo de desarrollo de sistemas, el cual puede estar conformado por desarrolladores propios de la nómina de la organización o desarrolladores externos contratados por terceros.

Adicionalmente existe un conjunto de asesores o analistas integral de riesgos, que buscan integrar los datos de las distintas gerencias para ofrecer una visión general del departamento.

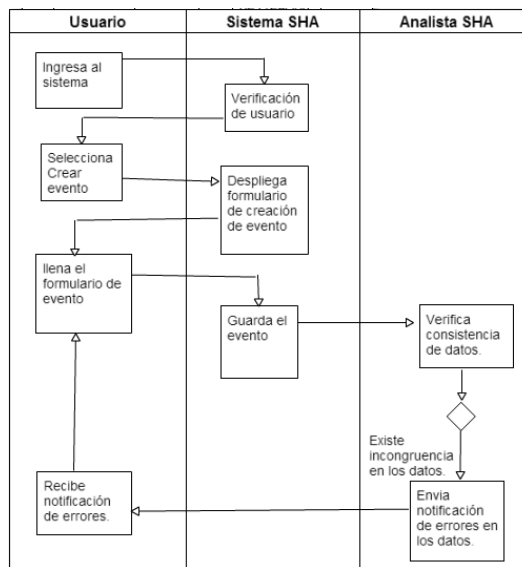
### **2.8.5.- Sistema de información SHA.**

La figura 25, muestra el conjunto de funcionalidades actual del sistema transaccional de la organización. El sistema tiene un módulo de seguridad donde se maneja al acceso a los distintos elementos del sistema. Un módulo de configuración general donde se gestionan los elementos comunes a los eventos y acciones correctivas, como el departamento o localidad donde ocurre la situación. Configuración de eventos donde, en función del evento se adaptan los parámetros establecidos y el módulo de eventos donde se crean eventos, sus detalles y acciones correctivas así como se pueden obtener gráficos estadísticos de un evento.



**Figura 25 Funcionalidades del sistema transaccional del sistema SHA.**

La figura 26 se muestra el flujo de la actividad básica del sistema que es la creación de un evento. Para esto, el usuario ingresa al sistema a partir de una identificación y crea un evento en función de un tipo de evento. De forma periódica, analistas o empleados que conforman parte de la gerencia media del departamento de seguridad, higiene y ambiente realizan validación de los datos para confirmar su exactitud con los hechos ocurridos en la organización.

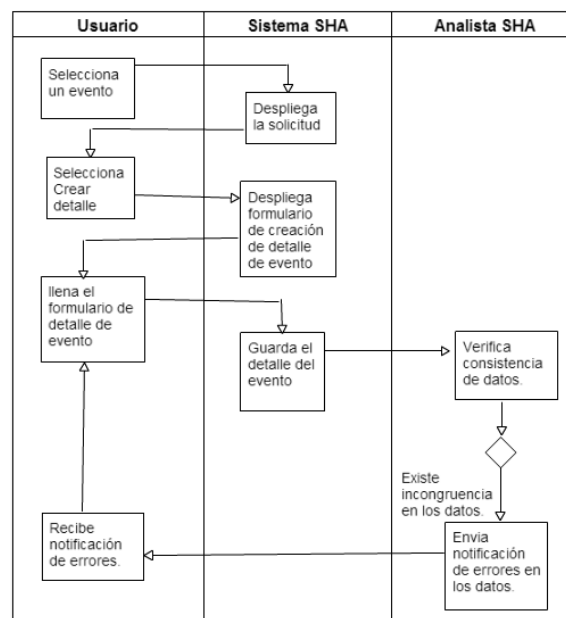


**Figura 26 Flujo de actividades para crear un evento.**

Para un evento, se pueden tener distintos tipos de eventos. Estos son

- Opersafe: registro de hallazgos encontrados en la organización importantes para el área de SHA.
- Registro técnico: inspecciones a los distintos equipos.
- Incidentes/accidentes: registro de alguna irregularidad o accidente laboral.

Los eventos de opersafe y registro técnico, tienen además detalles. Cada detalle corresponde con la incidencia o registro de un suceso. Por ejemplo, aunque el evento registro técnico representa todas las inspecciones de un año en específico, es el detalle el que registra una inspección específica. La creación de un detalle corresponde al flujo de actividades de la figura 27. Adicionalmente analistas del departamento de seguridad, higiene y ambiente verifican la existencia del detalle y la consistencia de los datos, este proceso de verificación adicional, al igual que con los eventos, ocurre de forma periódica



**Figura 27 Flujo de actividades para crear el detalle de un evento.**

Para los eventos (en algunos casos de forma transitiva a partir de un detalle), se tiene el módulo de acciones correctivas (flujo de actividades de la figura 28). Este módulo permite establecer el conjunto de medidas a realizar a partir del registro de un suceso.

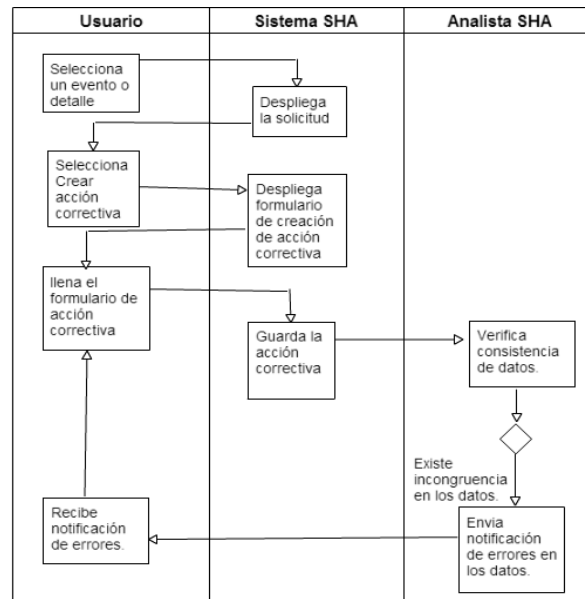


Figura 28 Flujo de actividades para crear una acción correctiva.

Para la creación de eventos, detalles de eventos, etc. Se requiere elementos generales que son utilizados a la hora de crear dichos objetos. Algunos de estos elementos son generales, como una localidad, otros son específicos de un evento, como el tipo de prueba para el evento registro técnico. La figura 29, muestra el flujo de actividades asociado a la creación de un elemento de configuración que será utilizado posteriormente en la creación de otros objetos dentro del sistema.

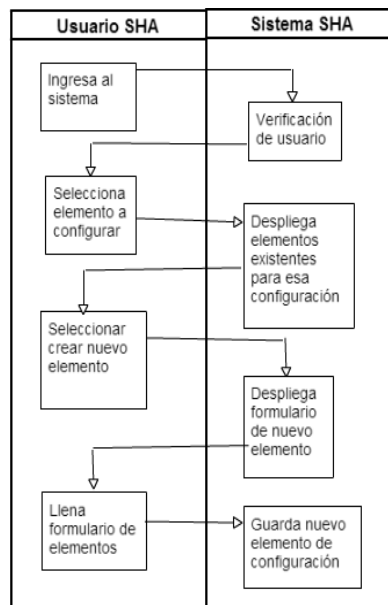


Figura 29. Flujo de actividades para crear un elemento de configuración general.

Adicionalmente cada evento tiene un conjunto de gráficos que buscan representar ciertos indicadores en función de las acciones correctivas, esto porque las acciones correctivas son

el núcleo del sistema puesto que muestra efectivamente cuales operaciones está realizando la organización.

### **2.8.6.- Importancia del proceso SHA.**

Peréz (2011) remarca que el área de SHA dentro de organizaciones que realizan operaciones delicadas es vital, esto porque a partir de sus funciones genera múltiples beneficios dentro de la organización. Estos beneficios vienen dado en primera instancia en el establecimiento de políticas que buscan preservar los bienes de la empresa, por ejemplo, al controlar las inspecciones se pueden evitar daños en los componentes que puedan ocasionar daños mayores producto del fallo de un componente.

Un proceso de SHA eficiente además es fundamental para garantizar la integridad de los empleados. La seguridad de los empleados es algo importante para la organización por múltiples razones como los éticos donde no debería arriesgarse la vida de un trabajador en un proceso de negocio, los económicos evitando gastos médicos adicionales, demandas así como desde el punto de vista político, una organización que no cuida la seguridad de sus empleados se pone a merced del escarnio público lo que puede dañar la imagen de la organización y afectar los beneficios de la organización.

Por último, con el auge de los temas ecológicos, las organizaciones no se pueden dar el gusto de afectar el medio ambiente, no solo existe un daño ecológico muchas veces irreparable y que podría a la final dañar el ambiente donde se realizan las operaciones de la organización al punto de afectar la productividad de la misma, sino que pueden darse demandas e incluso clausura de operaciones por parte de los fiscales competentes en el área.

Por tanto el área SHA realizando sus múltiples operaciones según sea las características de la organización, permite que los procesos del negocio se realicen de forma limpia, es decir, de una forma tal donde los daños hacia los bienes o el ambiente sean mínimos.

## CAPÍTULO 3

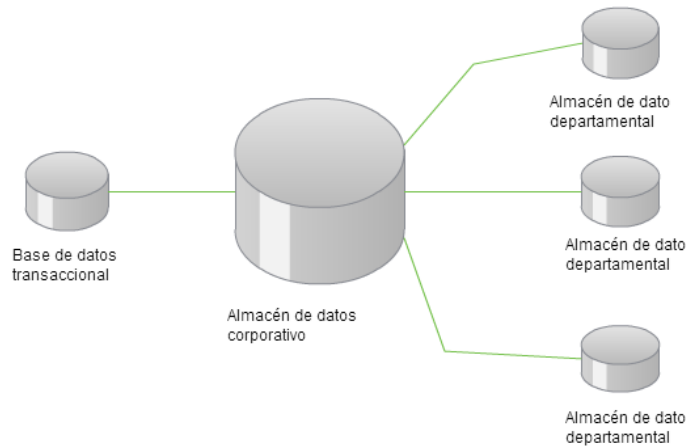
### MARCO METODOLÓGICO

El marco conceptual procedimental que conlleva la elaboración de una solución de inteligencia de negocio, viene estrechamente relacionada con el marco conceptual para la elaboración de un almacén de datos. Esto ocurre, porque dentro de la infraestructura una solución de inteligencia de negocios, el almacén de datos (y todo lo que tiene que ver con su elaboración) es una pieza fundamental, puesto que su desarrollo abarca desde la captación de datos desde las distintas fuentes, hasta la elaboración de un modelo que permita integrar los datos de una forma idónea para realizar procesos analíticos sobre datos. Basado en esto existen dos componentes esenciales dentro de una solución de negocios, los datos (por lo general almacenados en un almacén de datos) y las herramientas que el cliente tiene para visualizar y manipular los datos (agregación de datos, minería de datos, generación de cubos OLAP). Dichas herramientas son por lo general aplicaciones que agrupa una casa de software y que poseen características definidas. Tienen un proceso de instalación y configuración, donde se conectan al servidor que provee los datos y a partir de este punto, el usuario puede generar los elementos requeridos para la toma de decisiones.

Sobre todo el universo que tiene que ver el desarrollo de una estructura de datos propicia para la toma de decisiones (almacenes de datos), sobresalen dos teorías de desarrollo de dos figuras muy importantes dentro del área, Ralph Kimball y Bill Inmon. Aunque ambos autores tienen puntos en común, los mismos tienen filosofías distintas de trabajo para alcanzar los objetivos y el desarrollo del almacén de datos en general.

#### **3.1.- Metodología Bill Inmon (Top-down).**

Inmon establece en su libro “Building the datawarehouse” en 1992 una forma de desarrollo de un almacén de datos. Esta forma de desarrollo de un almacén de datos puede ser descrito, como se observa en la figura 30, en que la construcción de un almacén de datos corporativo parte de la fuente de datos operacional y luego se construye el conjunto de almacenes de datos departamentales.



**Figura 30** Proceso de desarrollo según Bill Inmon.

**Fuente:** Inmon, I. (1992). *Building the Operational Data Store* (1ª ed.). Estados Unidos de América: John Wiley and Sons.

Las respectivas consideraciones que señala Inmon son descritas a continuación:

- El inicio del desarrollo de un almacén de datos viene ligado al modelo de datos organizacional que posee la empresa. Este modelo de datos es el que sustenta los procesos operacionales de la organización y del cual se realiza una variación para generar un almacén de datos.
- Todos los requerimientos de un almacén de datos no pueden ser conocidos a priori. Esto viene dado porque una vez que se desarrolla y se implementa el almacén de datos es que dichos requerimientos pueden salir a la luz, por lo que el desarrollo de un almacén de datos es cíclico e incremental.

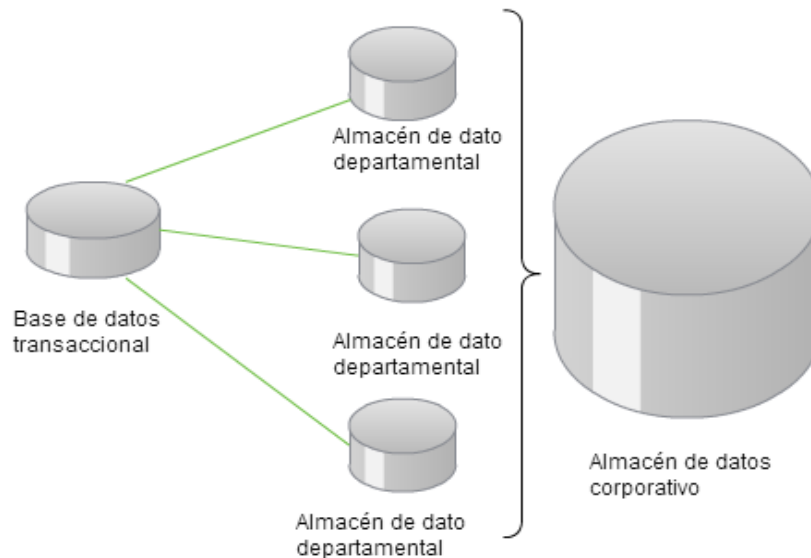
Este enfoque de desarrollo se basa en que el almacén de datos se basa en datos, no en procesos que buscan objetivos específicos (requerimientos funcionales y no funcionales claramente definidos). Por tanto todos los requerimientos sobre el almacén de datos no pueden definirse desde un principio, porque los usuarios no han tenido la capacidad de interactuar con el sistema de soporte de decisiones. Es después del uso de la información que genera el almacén de datos que ciertos requerimientos pueden ser conocidos.

- Los ambientes donde existen los sistemas y operacionales y el almacén de datos deben de ser distintos (a nivel de base de datos y equipo). Esto asegura que el procesamiento analítico no afecte el rendimiento de los sistemas ya existentes.
- La extracción inicial de datos es grande porque se requiere poblar con toda los datos histórica el almacén de datos, además hay que tener especial cuidado en la granularidad y las marcas de tiempo. Las marcas de tiempo denotan el hecho de que el almacén de datos guarda históricos y un solo registro puede tener múltiples imágenes o valores en función de determinados instantes de tiempo. Luego existen un conjunto de cargas periódicas preestablecidas.
- Establece como punto de partida la creación de un almacén de datos empresarial, es decir, empezar con una visión general de la organización. Esto requiere un amplio conocimiento de todas las estructuras y relaciones de la organización de forma tal de plasmar en el diseño del almacén de datos todos estos componentes.
- Posteriormente se generan mercados de datos que serán subconjuntos del almacén de datos que se crean a partir de agrupaciones lógicas existentes, como por ejemplo agrupar los elementos que tienen que ver con el área de ventas y generar un mercado de datos.
- Inmon refuta la utilización del esquema estrella como modelo para sustentar un almacén de datos, puesto que “El enfoque estrella multidimensional es un enfoque donde el diseño de la base de datos es basado en las ocurrencias de datos dentro de un área de negocio y como esta área es accedida desde este punto. El enfoque estrella aplica al mundo de los mercados de datos, no al de los almacenes de datos. Es un error construir un almacén de datos usando un modelo estrella porque el almacén de datos terminara por ser óptimo para un grupo de usuarios y costoso para todos los demás”. Inmon plantea la utilización de un modelo normalizado para realizar el almacén de datos corporativo puesto que el mismo siempre estará afectado por el punto de vista (departamento) que guie u oriente su desarrollo, por lo que la normalización garantiza un enfoque general que sirve como punto de partida a los distintos mercados de datos, que son los que si realizaran los procesos de desnormalización a partir de las necesidades específicas de las distintas áreas dentro de la organización.

- Deben desarrollarse modelos que describan la organización y sus modelos de datos y que a su vez sirvan de guía para el desarrollo del almacén de datos así como la posterior creación de los mercados de datos.

### 3.2.-Metodología Ralph Kimball (Bottom-up).

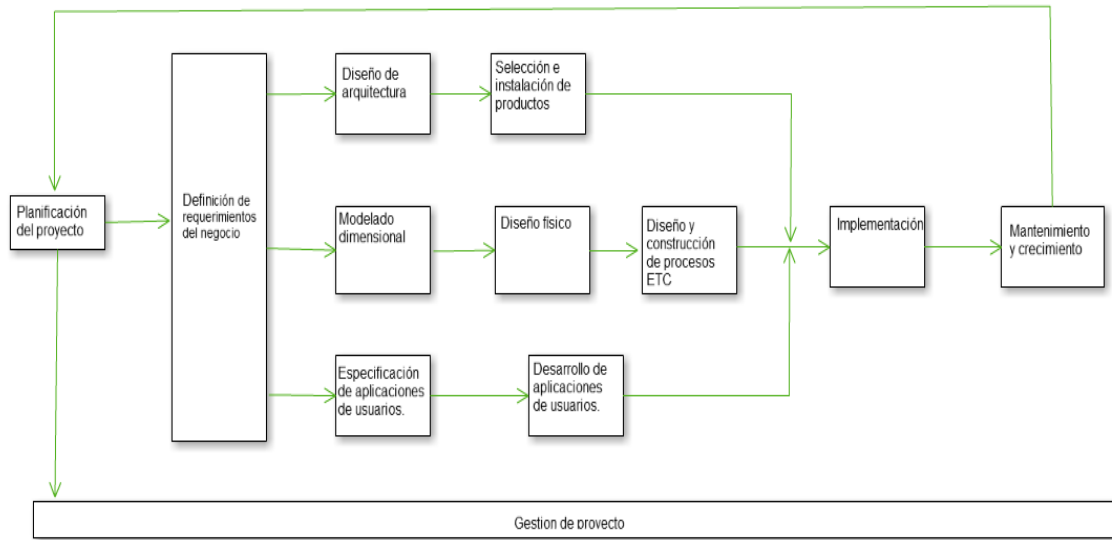
La metodología de Kimball es ampliamente utilizada en las corporaciones por su capacidad de ofrecer resultados medibles en un corto periodo de tiempo (en comparación a la metodología de Inmon). Para la realización de la solución, como se observa en la figura 31, se parte de la creación de almacenes de datos departamentales (mercados de datos) a partir de los datos operaciones. El conjunto de mercados de datos conforman al final el almacén de datos corporativo.



**Figura 31 Proceso de desarrollo según Ralph Kimball.**

**Fuente: Kimball, R. (1998). *The datawarehouse lifecycle toolkit* (1ª ed.). Estados Unidos de América:Wiley.**

La metodología de Kimball esta descrita en su libro publicado en 1998 “The Data Warehouse lyfecicle toolkit”, en este libro Kimbal establece una serie de pasos que deben seguirse para la realización de un almacén de datos. Este conjunto de pasos no son lineales, como se ve en la figura 32, en el ciclo de desarrollo ciertas tareas pueden ser desarrolladas de forma paralela.



**Figura 32** Ciclo de vida de la metodología Kimball.

**Fuente:** Kimball, R. (1998). *The datawarehouse lifecycle toolkit* (1ª ed.). Estados Unidos de América:Wiley.

El conjunto de pasos que plantea Kimball parte de la planificación del proyecto, que establece el área de negocio a trabajar, para a partir de la definición de los requerimientos analíticos, comenzar de forma paralela tres procesos.

- Arquitectura.

Se diseña, seleccionan e instalan el conjunto de elementos de hardware y software que soportara la solución.

- Construcción del almacén de datos.

Conlleva la generación del modelo dimensional, su implementación en una base de datos a partir de la generación de los procesos de extracción, transformación y carga pertinentes.

- Aplicaciones de usuarios.

Se diseñan y seleccionan las formas como los usuarios accederán a la solución de inteligencia de negocios.

Una vez realizadas estas tres fases, se implementa la solución en la organización. Debido a que hablamos de una solución incremental la implementación requiere su respectivo mantenimiento y crecimiento, adaptándose a las nuevas necesidades analíticas que puedan existir.

La gestión del proyecto nos permite a su vez administrar los recursos para que puedan darse las distintas fases del proyecto.

### 3.3.- Metodología mejores prácticas.

Es una metodología que busca obtener lo mejor de la metodología de Kimball a nivel de velocidad y trabajo ascendente, mas adaptando el desarrollo a las necesidades particulares mostradas en las organizaciones venezolanas.

Planteada en “manual Metodología de Desarrollo de Sistemas de Información de Inteligencia de Negocio” de Tian Consultores, la metodología de mejores prácticas se basa en 4 fases, descritas en la figura 33. Las cuales son descritas a continuación:

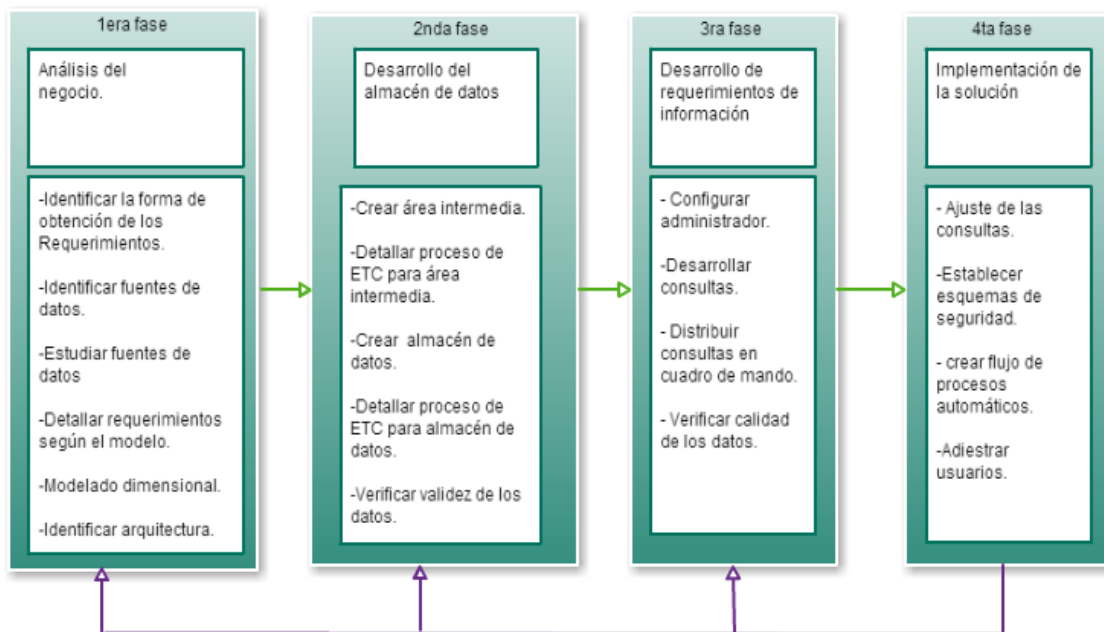


Figura 33 Fases de metodología mejores prácticas.

Fuente: TIAN CONSULTORES. (2010). “Manual Metodología de Desarrollo de Sistemas de Información de Inteligencia de Negocio”. Caracas,Venezuela.

### **3.3.1.- Análisis del negocio.**

En esta fase se describe cual es el alcance y objetivo de la solución de inteligencia de negocios. Para esto se estudia cual será el modelo de análisis del negocios, el cual determinara como serán planteados los requerimientos. Basado en esto, los requerimientos pueden definirse según 3 escenarios, descritos en la sección de modelado dimensional (marco conceptual punto 3.6.5).

A partir de cómo serán levantados los requerimientos se procede a estudiar las distintas fuentes de datos para desarrollar el modelo dimensional También se define la arquitectura a utilizar para el desarrollo del proyecto.

### **3.3.2.- Desarrollo del almacén de datos.**

Una vez establecido el modelo de datos y la infraestructura a utilizar, se desarrolla el área intermedia y el almacén de datos en función de las herramientas seleccionadas. A partir de esto se crean los procesos de extracción, transformación y carga (ETC) que se encargaran de pasar los datos de las distintas fuentes al área intermedia, para posteriormente ubicar los datos en el almacén de datos. En ambos procesos se tiene que realizar pruebas en la calidad de los datos para validar los mismos. Por último, una vez implementado el almacén se analiza y documentar los requerimientos establecidos en el modelo de negocios y que han sido considerados en la realización del modelo.

### **3.3.3.- Desarrollo de requerimientos de información.**

Implica la construcción de la metada de la solución así como la definición del proceso de conexión entre el almacén de datos y la herramienta de inteligencia de negocios. Una vez los datos son cargados en la herramienta se procede a realizar las consultas que satisfagan los requerimientos del negocio así como establecer la forma como se verán los indicadores dentro del cuadro de mando. Por último se verifica que los datos mostrados en el cuadro de mando correspondan a los datos existentes en la fuente de datos.

### 3.3.4.- Implementación de la solución.

Esta etapa consiste en la puesta en ejecución de la solución de inteligencia de negocios. Esto genera a su vez dos tareas básicas: el adiestramiento de los usuarios finales y el ajuste de los indicadores dentro del cuadro de mando a partir de requerimientos en la forma de presentación de los mismos.

Adicionalmente se establece todo lo relacionado a la seguridad de la inteligencia, estableciendo usuarios, roles y restricciones sobre las vistas.

### 3.4.- ¿Cuál metodología escoger?.

Desde la conceptualización de las distintas metodologías, puede establecerse una serie de comparaciones, descritas en la tabla 2.

Metodología Inmon.	Metodología Kimball.	Metodología mejores prácticas.
Los requerimientos no se conocen al principio.	Los requerimientos definen la implementación.	Los requerimientos definen la implementación.
Método descendente.	Método ascendente.	Método ascendente.
Se define modelo, arquitectura e implementación.	Se define de forma paralela arquitectura, modelo e implementación.	Se define modelo, arquitectura e implementación.
Implementación lenta.	Implementación rápida.	Implementación rápida.
Ideal para soluciones complejas.	Ideal para soluciones escalables.	Ideal para soluciones escalables.
Parte de que no todos los requerimientos están establecidos.	Parte de que los requerimientos están establecidos.	Parte de que los requerimientos pueden

		darse en distintos escenarios.
Adaptada a necesidades universales.	Adaptada a necesidades universales.	Adaptada a las necesidades de las organizaciones venezolanas.
No contempla de forma directa la visualización de los indicadores.	No contempla de forma directa la visualización de los indicadores.	Contempla la visualización de los indicadores para el análisis.
Es una metodología para el desarrollo de un almacén de datos.	Es una metodología para el desarrollo de un almacén de datos.	Es una metodología para el desarrollo de una solución de inteligencia de negocios.

**Tabla 2 Comparación metodologías.**

Bill Inmon plantea una solución general desde la cual se desglosen soluciones departamentales, o desarrollo “Top-down”. Inmon establece que se debe realizar primeramente un almacén de datos que abarque toda la estructura de la organización, pasando por un proceso de integración y purificación de los datos. Posteriormente se generan mercados de datos que serán subconjuntos del almacén de datos que se crean a partir de agrupaciones lógicas existentes, como por ejemplo agrupar los elementos que tienen que ver con el área de ventas y generar un mercado de datos.

Ralph Kimball ofrece una solución opuesta a la ofrecida por Inmon, puesto que se basa en la construcción de mercados de datos departamentales, para su posterior integración en un almacén de datos que abarque toda la organización, este tipo de diseño se denomina “Bottom-up”. Ralph establece que la complejidad de las organizaciones dificulta la realización de una solución inicial que abarque toda la organización. Por lo que la realización de mercado de datos que apoyen procesos organizacionales específicos es un mejor comienzo para el desarrollo de una solución. Eventualmente, a medida que se van construyendo más mercados de datos, estos pueden conectarse por elementos en común (atributos, dimensiones,

indicadores, etc.) por lo que se va generando una solución que va abarcando íntegramente la organización.

La metodología Kimball es idónea cuando se está empezando a desarrollar una infraestructura de inteligencia de negocios, porque se puede desarrollar mercado de datos los cuales tienen una complejidad menor e instaurar una solución de negocios completa sobre un área determinada. Esto ayuda a que el proyecto pueda ser puesto en producción de forma rápida que si se compara con Inmon, lo cual ayuda a la organización a recuperar su inversión más rápidamente (sin contar que la inversión inicial, al ser la construcción de un mercado de datos es menor) lo que incentiva a la construcción de más mercados de datos departamentales, hasta obtener un almacén de datos conformado por muchos mercados de datos que abarca a toda la organización. La aplicación de la metodología Inmon tiene sentido cuando se requiere a priori un almacén de datos complejo, sin requerimientos claramente definidos, por lo que se necesita generar una estructura analítica base sobre la cual se generaran mercados de datos que atacarán procesos de negocios específicos.

La metodología de mejores prácticas busca sacar partido de la velocidad de implementación de la solución de la metodología de Kimball, definiendo requerimientos departamentales para la construcción de almacenes de datos departamentales a su vez que define la arquitectura, el modelo para una posterior implementación de la solución, esto garantiza que la solución sea más integral en el sentido de que los procesos al ser lineales se establece un mejor punto de comienzo para una nueva fase, así como se evitan los conflictos entre las unidades de desarrollo paralelas. La metodología de mejores prácticas adicionalmente busca adaptar su forma de implementación a las realidades de los negocios en Venezuela así como hace énfasis en la forma como se levantan los requerimientos para el desarrollo de una solución.

Como último punto, hay que establecer que la metodología de mejores prácticas efectivamente busca generar una solución de inteligencia de negocios, por lo que contempla la definición de los indicadores de gestión. Kimball e Inmon buscan en su metodología el desarrollo de un almacén de datos al que luego, de forma opcional puede incorporársele un componente de manipulación de la información. Esto genera que dichas metodologías no contemplen el desarrollo de indicadores de gestión, pues esto escapa del alcance del desarrollo de un almacén de datos.

## CAPITULO 4

### MARCO APLICATIVO

Este capítulo describe como a partir de un sistema transaccional dedicado al departamento de seguridad higiene y ambiente (SHA), es aplicada la metodología de mejores prácticas basadas en la metodología de Kimball para el desarrollo de una solución de inteligencia de negocios.

Se explica por tanto el análisis y planteamiento de los indicadores a utilizar, la arquitectura de la solución planteada, el diseño y construcción de área intermedia y almacén de datos con sus respectivos procesos de extracción, transformación y carga para la construcción de consultas analíticas y el respectivo despliegue de los indicadores de gestión utilizando herramientas de visualización propias de una solución de inteligencia de negocios.

A continuación se explica las fases y conjunto de tareas de la metodología para la construcción de la solución.

#### **4.1.-Análisis del negocio.**

Comprende el entendimiento del área de negocio en la que se llevara a cabo la solución de inteligencia de negocios para definir la arquitectura y el modelo dimensional que soportara las necesidades analíticas. Ha continuación se describe cada tarea relacionada a esta fase.

##### **4.1.1.- Identificar la forma de obtención de los requerimientos.**

Define en qué punto se pueden tener efectivamente los requerimientos analíticos y la forma como estos serán presentados. Para la solución del proceso SHA, se utilizara la base de datos transaccional para obtener los requerimientos. La obtención de requerimientos de esta forma permite poder satisfacer todos los requerimientos analíticos posibles para los datos que se manejan, lo que permite un modelo dimensional más completo, sobre todo cuando el dueño del negocio no tiene definida de forma clara todos los posibles requerimientos analíticos que existen en el sistema y el estudio del proceso de negocio puede ser trabajoso.

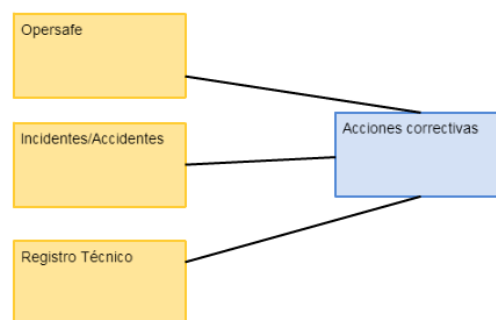
Esta forma de obtención de requerimientos de negocio a su vez, genera que los requerimientos analíticos solo puedan establecerse de forma clara una vez implementado el

almacén de datos, punto en el que se pueden realizar consultas y ser visualizadas con las distintas herramientas de manipulación que provee la solución. Sin embargo, de antemano observando el modelo relacional del sistema, pueden observarse algunos requerimientos analíticos, como por ejemplo:

- Cantidad de acciones correctivas por departamento.
- Estado de las acciones correctivas en un instante de tiempo.
- Evaluar la evolución de las acciones correctivas en el tiempo en función de su estado.
- Identificar responsables de acciones correctivas vencidas o por vencer según el tipo de evento.

#### 4.1.2.- Identificar fuentes de datos.

Como fuente de datos se posee la base de datos transaccional que es utilizada por la organización en el proceso SHA. La base de datos relacional está estructurada en 4 módulos como se observa en la figura 34. Cada evento es un módulo particular y dichos eventos se relacionan únicamente con el módulo de acciones correctivas, el cual sirve como núcleo del sistema puesto que en dicho módulo es donde se registran las operaciones a realizar según un evento o detalle de evento registrado. Para efectos prácticos, la fuente de datos será llamada “Relacional”, al mencionar algún atributo de la fuente de datos se especificará de la forma Relacional.Nombre\_tabla.Nombre\_atributo.



**Figura 34 Módulos del sistema relacional.**

**Fuente: Sistema transaccional de proceso SHA.**

Debido a las características del modelo relacional, la solución estará basada en el módulo de acciones correctiva porque representa el núcleo de sistema, adicionalmente serán utilizado

algunos datos de los módulos de los eventos para generar el nombre del detalle, el cual es la concatenación de una serie de atributos.

A continuación se muestra y explica el modelo relacional de cada uno de los módulos:

- **Acciones correctivas.**

Es el módulo eje del proceso de SHA. En él se registran todas las operaciones realizadas a partir de un evento. Una acción correctiva puede estar relacionada al detalle de un evento o a evento de forma directa, a su vez, como se observa en la figura 35 la acción correctiva tiene un conjunto de elementos que la caracterizan como el estatus, la prioridad o el responsable.

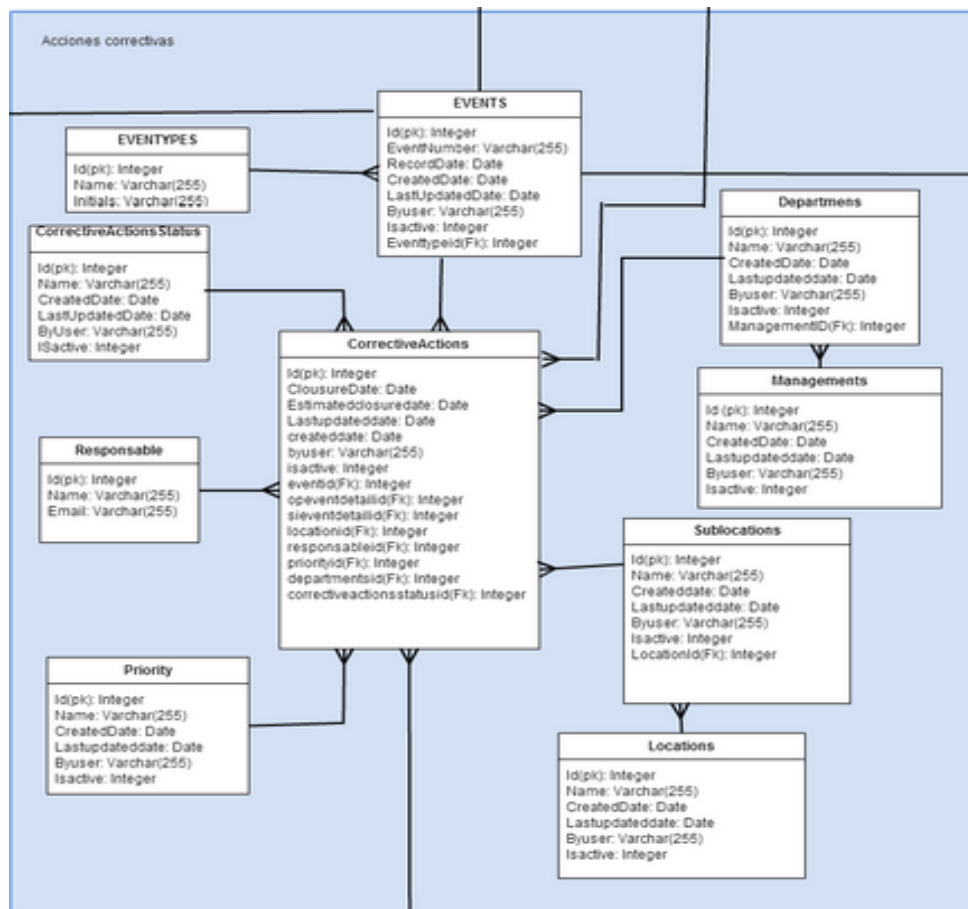


Figura 35 modelo relacional del módulo de acciones correctivas.

Fuente: Sistema transaccional de proceso SHA.

- **Opsafe.**

Modulo que registra eventos de tipo operasafe, un evento operasafe representa un hallazgo en materia de seguridad, higiene o ambiente. Son eventos de tipo anual, en donde, por la creación de un evento, cada detalle representara la ocurrencia de un hecho. Como se observa en la figura 36, el detalle de un evento operasafe es el que tiene relacionado el conjunto de elementos como la evaluación, el estado, etc. El módulo operasafe se conecta al módulo de acciones correctivas de dos formas. Existe una relación entre la tabla *opevents* y *events* y otra relación entre *opeventsdetails* y *correctiveactions*.

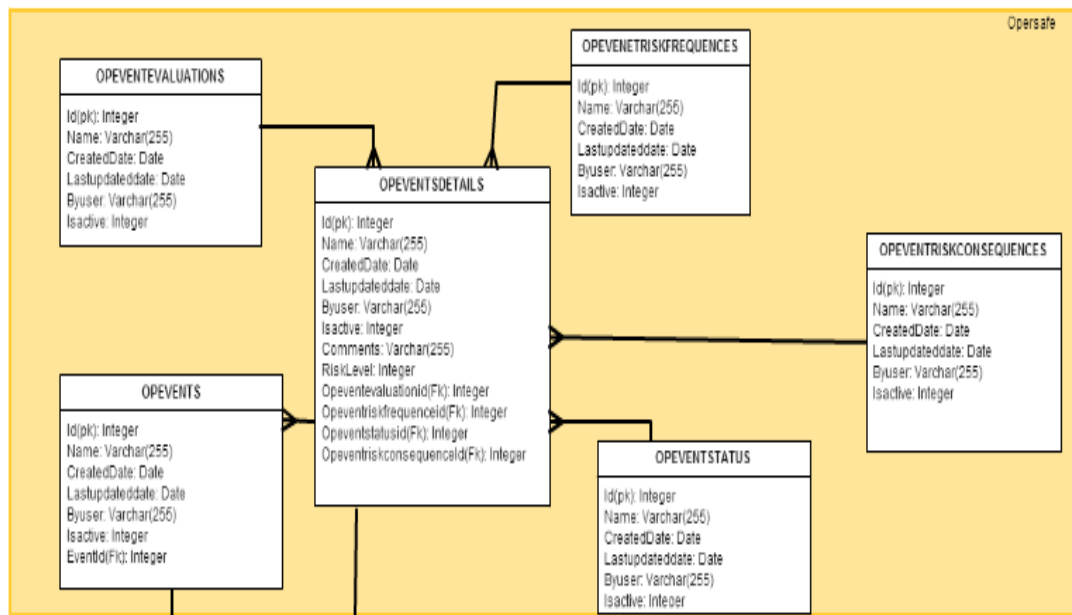


Figura 36 Modelo relacional del módulo Operasafe.

Fuente: Sistema transaccional de proceso SHA.

- **Registro Técnico.**

Registra las múltiples inspecciones realizadas en la organización. Un evento tipo registro técnico es anual y cada detalle representa una inspección específico, por esto, como se ve en la figura 37 la tabla *sieventsdetails* se asocia al conjunto de elementos que identifica el tipo de prueba, el equipo que se la ha realizado la inspección y el resultado de la misma. Se conecta al módulo de acciones correctivas de dos formas. Existe una relación entre la tabla *sievents* y *events* y otra relación entre *sieventsdetails* y *correctiveactions*.

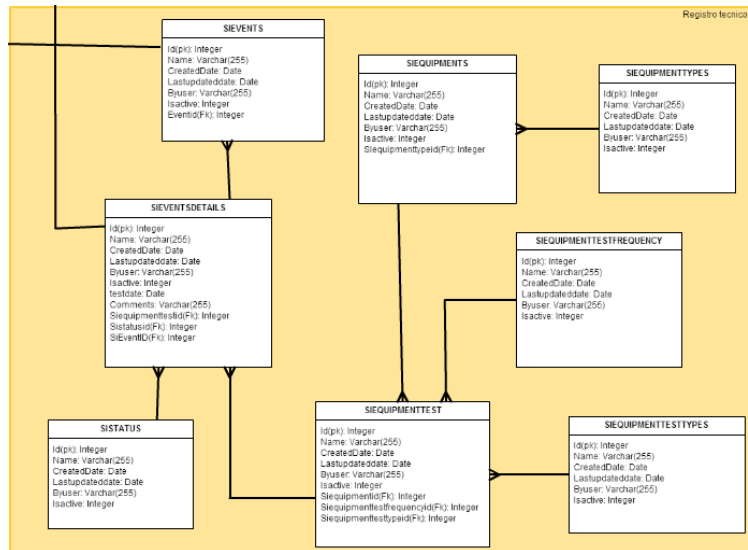


Figura 37 Modelo relacional del módulo Registro técnico.

Fuente: Sistema transaccional de proceso SHA.

- **Incidentes/Accidentes.**

Registra cualquier anomalía ocurrida en la organización, esto puede ser por un incidente particular o un accidente ocurrido en la planta.

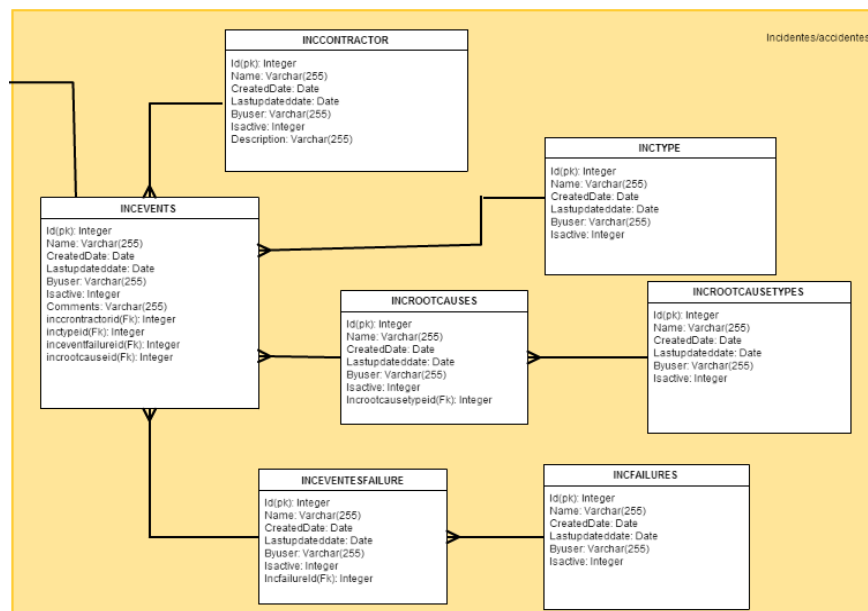


Figura 38 Modelo relacional del módulo Incidentes y Accidentes.

Fuente: Sistema transaccional de proceso SHA.

Como se observa en la figura 38, es el único tipo de evento que posee detalle, esto porque la creación de un evento de este tipo no es de forma anual o periódica, sino que efectivamente un evento tipo incidente/accidente registra una ocurrencia particular, esto hace que solo exista una relación entre este módulo y el de acciones correctivas. Esta relación se da entre la tabla *incentents* y *events*.

Luego de identificado los elementos a utilizar del modelo relacional (módulo de acciones correctivas y elementos adicionales de los demás módulos para construir atributos particulares del modelo dimensional), se procede a generar el modelo dimensional.

#### **4.1.3.- modelado dimensional.**

En esta tarea busca generar un modelo dimensional que permita satisfacer las necesidades analíticas, específicamente las necesidades analíticas asociadas a las acciones correctivas del proceso SHA. A partir de esto, se realizan las fases del modelado dimensional, descritas a continuación.

- **Escoger área temática.**

Como se ha establecido previamente, el proceso es el de seguridad, higiene y ambiente (SHA), específicamente el módulo de acciones correctivas que sirve como núcleo de dicho proceso. La forma de establecer y documentar los requerimientos vendrá dado una vez este implementado el modelo dimensional, debido a que se abarcan todos los requerimientos analíticos posibles de las acciones correctivas al utilizar el modelo relacional.

- **Granularidad.**

Permite identificar el detalle de lo que se desea medir en función de las características de un evento particular. Esto genera como resultado saber que datos o elementos son requeridos en nuestro modelo dimensional.

En este caso, se busca tener la cantidad de acciones correctivas de un evento (o detalle de evento) generadas por un responsable en un tiempo determinado, según un estatus y prioridad definidos, realizadas en una localidad y departamento específico.

- **Identificar dimensiones.**

Una vez establecida la granularidad del modelo, se procede a identificar las dimensiones. Las dimensiones caracterizan lo que se quiere medir, es la cantidad de acciones correctivas, por tanto las dimensiones son los demás elementos, en este caso: evento, responsable, tiempo, estatus, prioridad, sublocalidad y departamento. La tabla 3 describe cada una de las dimensiones establecidas.

Dimensión.	Descripción.
<b>Eventos.</b>	Representa el evento que genera la acción correctiva
<b>Responsables.</b>	Es la encargada de la acción correctiva.
<b>Tiempos.</b>	El momento en que se crea, se estima que se cierre y se cierra una acción correctiva. La dimensión tiempo toma múltiples significados en el modelo.
<b>Estatus.</b>	El estado de la acción correctiva.
<b>Prioridades.</b>	El nivel de importancia de la acción correctiva.
<b>Sublocalidades.</b>	El lugar físico donde ocurre la acción correctiva.
<b>Departamento.</b>	Área organizacional a la cual está asociada la acción correctiva-

**Tabla 3 Descripción de las dimensiones.**

Una vez establecida las dimensiones, puede establecerse las distintas jerarquías asociadas. Entendiendo que las jerarquías van a permitir navegar sobre los distintos niveles lógicos existentes en una dimensión a partir de la utilización de los distintos operadores OLAP. En la figura 39 se observan las jerarquías establecidas para cada una de las dimensiones.



**Figura 39 Representación de las dimensiones y sus respectivas jerarquías.**

Para las dimensiones responsables, estatus y prioridades, solo existe un nivel lógico de sus atributos.

La dimensión Eventos posee las jerarquías tipo evento (Registro técnico, operasafe, incidentes/accidente), el evento y sus detalles.

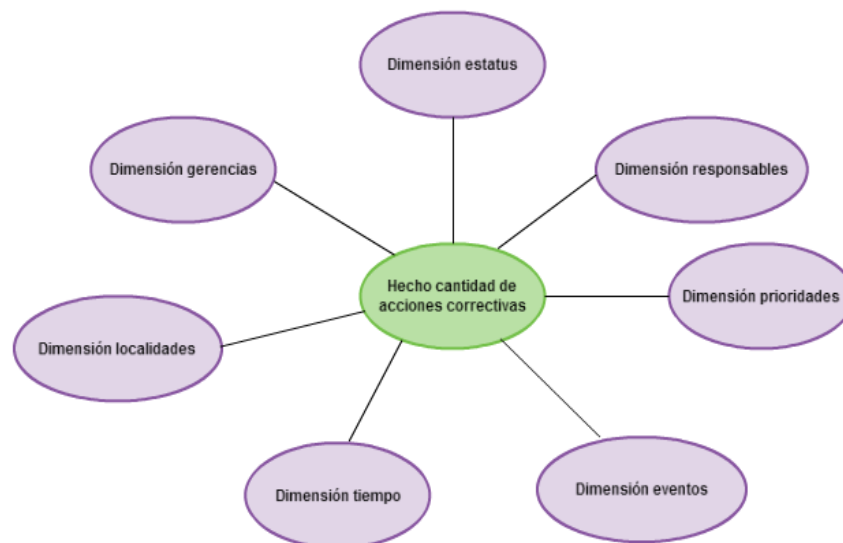
La dimensión tiempo ha sido estructurada en base a tres jerarquías lógicas: año, mes y día.

La dimensión sublocalidades está estructurada en sublocalidades que forman parte de una localidad. La dimensión departamentos por su parte está estructurada en departamentos que forman parte de una gerencia.

- **Identificar hechos.**

El conjunto de dimensiones descritas anteriormente buscan identificar la cantidad de acciones correctivas. Por tanto el hecho en su menor nivel de granularidad es la generación de una acción correctiva.

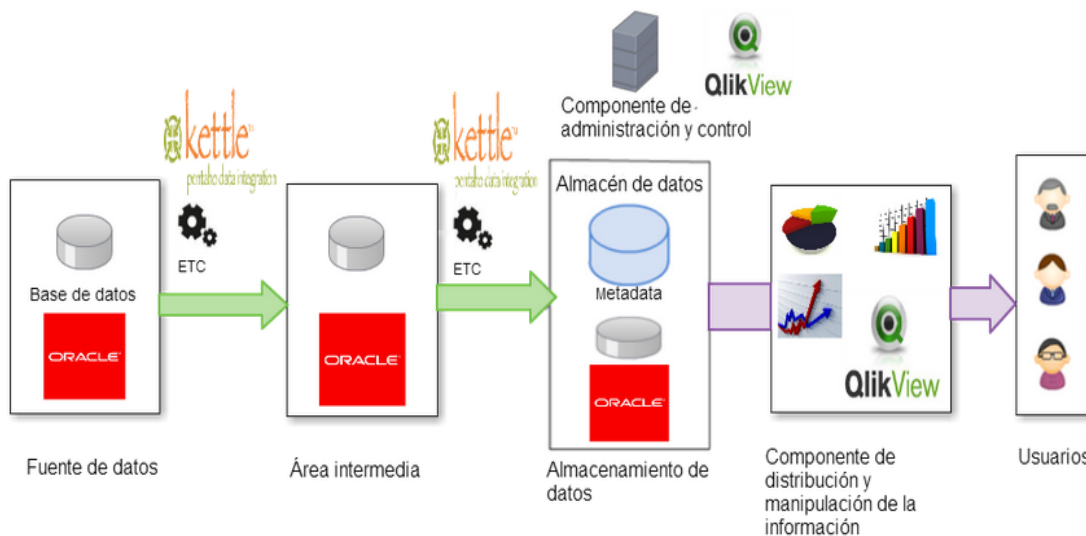
Una vez identificado el hecho y las dimensiones, se puede generar el modelo dimensional, el cual se observa en la figura 40.



**Figura 40 Modelo dimensional del proceso SHA.**

#### **4.1.4.-Identificar arquitectura.**

Para la creación de la solución de inteligencia de negocios será utilizada la arquitectura propuesta en la figura 41.



**Figura 41 Arquitectura de la solución.**

Como se ha propuesta en la solución (figura 4, capítulo 1), La solución estará conformada por:

- Fuente de datos: la base de datos transaccional Oracle del sistema SHA de la organización. Donde se hace énfasis en el módulo de acciones correctiva.
- Área intermedia: desarrollada en motor de base de datos Oracle, permitirá integrar y hacer transformaciones a los datos para ser ubicados en el almacén de datos.
- Almacén de datos: es la implementación del modelo dimensional para la ayuda en la toma de decisiones utilizando el motor de bases de datos Oracle.
- Componente de distribución y manipulación de la información: permitirá la construcción de los indicadores de gestión que permita a los usuarios satisfacer sus necesidades analíticas. Sera utilizada la herramienta Qlikview
- Componente de distribución y control: permitirá administrar y gestionar los elementos del sistema. Sera utilizado Qlikview, enfocado en la gestión de la distribución y manipulación de la información.

- ETC: son los procesos que permiten pasar de un repositorio de datos a otro adaptando e integrando los datos a un nuevo modelo de datos. Será utilizada la herramienta de Pentaho, Kettle.

#### 4.2.- Desarrollo del almacén de datos.

Una vez se posee el modelo dimensional y se ha identificado la arquitectura a utilizar, se procede a implementar el modelo dimensional en un motor de base de datos para poder ser accedido de forma que provea información analítica a la organización. A continuación, se describen las tareas requeridas para el desarrollo del almacén de datos.

##### 4.2.1.- Crear área intermedia.

Como punto medio para la integración y transformación de las fuentes de datos para obtener un modelo propicio para ser trasladado al almacén de datos. En el caso de la solución propuesta, al poseer una sola fuente de datos, no se requiere integrar distintas fuentes de datos. Por otro lado la construcción de la mayoría de las dimensiones puede hacerse de forma directa, utilizando la fuente de datos sin necesidad de un área intermedia. Para efectos prácticos, el área intermedia como base de datos será llamada “Intermedia” y el acceso a algún atributo tendrá el siguiente formato, Intermedia.Nombre\_tabla.Nombre\_atributo.

Sin embargo la dimensión evento requiere un proceso previo por sus características. La necesidad de generar un área intermedia para el evento viene dada por el detalle del evento y su forma de relacionarse con las acciones correctivas. La forma en cómo se trabaja el detalle de cada evento y la forma como se crea el nombre de cada detalle del evento requiere que se tenga un proceso particular para cada tipo de evento de forma que se consoliden al final, todos los tipos de eventos en la dimensión evento para luego ser utilizada en el almacén de datos.

La tabla 4 muestra como se trata el detalle de un evento según el tipo de evento.

Tipo evento.	Detalle evento.	Detalle nombre.
OperSAFE.	Si.	Conformado por el nombre del evento y el número que identifica ese detalle en particular.

Incidentes/Accidentes.	No.	No aplica.
Registro técnico.	Si.	Conformado por la fecha de la inspección, tipo de equipo y prueba involucradas en la misma.

**Tabla 4 Tipo de eventos.**

Debido a que el área intermedia solo trabajara con los eventos, la misma solo estará conformada por la tabla evento, la cual está estructurada como se observa en la figura 42.

Eventos
Id(pk): Integer
Tipo_evento_id: Integer
Tipo_evento_nombre: varchar(255)
Evento_id: Integer
Evento_nombre: varchar(255)
Evento_detalle_id: Integer
Evento_detalle_nombre: varchar(255)

**Figura 42 Tabla eventos.**

Para la creación del área intermedia, se utilizó una herramienta por Oracle para el manejo de base de datos, SQLdeveloper. SQLdeveloper permite realizar operaciones sobre la base de datos desde una interfaz gráfica sencilla, con un conjunto de menús y elementos que permiten realizar operaciones sin la necesidad de utilizar la línea de comandos, lo que hace que el proceso sea más simple y rápido. En la tabla 5 se describe la tabla de eventos.

Atributo	Tipo	Descripción	Procedencia
Id	Integer	Atributo que identifica unívocament e una instancia.	Propio de la generación de la tabla
Tipo_evento_id	Integer	Identificador de la fuente de datos para un tipo de evento.	Del atributo Relacional.EventsType.Id
Tipo_evento_nombre	Varchar(255)	Nombre del tipo de evento.	Del atributo Relacional.EventsType.Name

Evento_id	Integer	Identificador de la fuente de datos para un evento.	Del atributo Relacional.Events.Id
Evento_nombre	Varchar(255)	Nombre del evento.	Del atributo Relacional.Events.Eventnumber
Evento_detalle_id	Integer	Identificador de la fuente de datos para un detalle de un evento.	<p>Depende del tipo de evento:</p> <p><b>Opersafe:</b> Relacional.opeventsDetails.Id.</p> <p><b>Registro Técnico:</b> Relacional.SieventsDetails.Id.</p> <p><b>Incidentes/accidentes:</b> No posee, se le coloca -1 por defecto.</p>
Evento_detalle_nombre	Varchar(255)	Nombre del detalle del evento.	<p>Depende del tipo de evento:</p> <p><b>Opersafe (concatenación de atributos):</b> Relacional.Events.Eventnumber Relacional.opeventsDetail.Name.</p> <p><b>Registro Técnico (concatenación de atributos):</b> Relacional.SieventsDetails.Testdate</p> <p>Relacional.SiEquipmenttype.name Relacional.SiEquipment.name.</p> <p><b>Incidentes/accidentes:</b> No posee, se le coloca no valido por defecto.</p>

Tabla 5 Características de los atributos de la tabla eventos.

En función de esto se procede a crear la tabla eventos en el área intermedia, paso que se observa realizado en la figura 43.

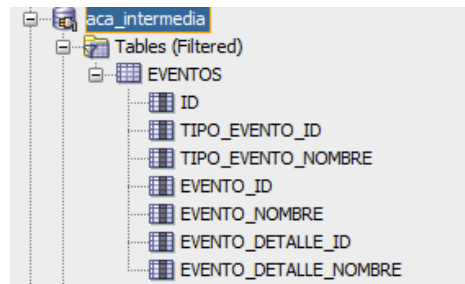


Figura 43 Tabla eventos generada en Oracle y visualizada en SQL Developer.

#### 4.2.2.- Detallar proceso ETC para área intermedia.

El área intermedia de la solución planteada posee una sola tabla, sin embargo el llenado de esta tabla requiere de tres procesos ETC. Debido a que existen tres tipos de eventos en donde, el tipo de evento define la forma como un atributo es creado, por cada tipo de evento se realiza un proceso ETC particular.

Para la realización de los ETC se utiliza la herramienta de Pentaho, Kettle. En este sentido, se asume que la base de datos y la tabla está creada, aunque dicho proceso de creación de tabla puede realizarse desde la herramienta.

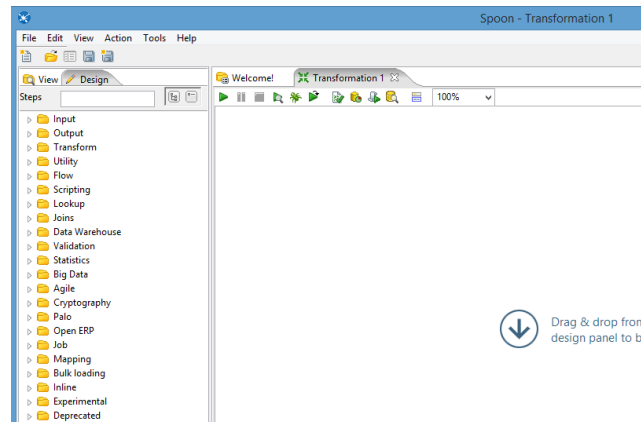
Kettle divide los tipos de elementos en transformaciones (transformation en inglés) y trabajos (Jobs en inglés). Una transformación consiste en un proceso de ETC concreto. Un trabajo permite ejecutar un conjunto de transformaciones u otros trabajos.

Para el desarrollo del área intermedia es necesario crear tres transformaciones, uno por cada tipo de evento y un trabajo, que ejecutara dichas transformaciones.

A continuación se mostrará cómo se realiza el proceso ETC para el tipo de eventos registro técnico, de forma similar, cada proceso ETC ha sido realizado para cada tipo de evento.

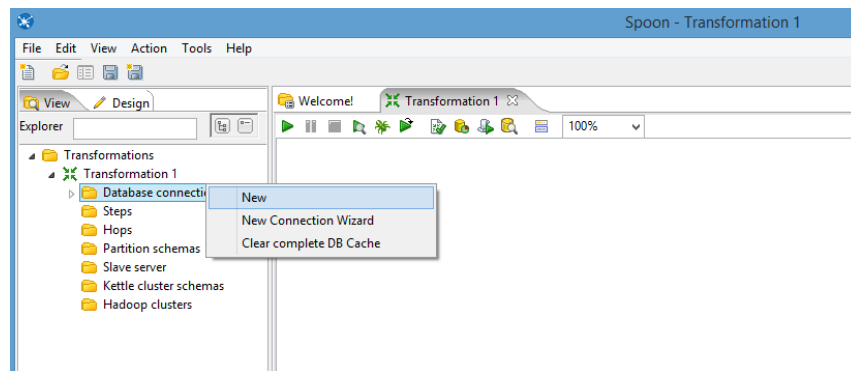
Pentaho Kettle no necesita ser instalado en el computador, debido a que utiliza la máquina virtual de java para ejecutarse (por lo que tener java instalado en el computador es un requisito). Para acceder a Kettle se requiere ejecutar el archivo Spoon.bat que se encuentra en la carpeta principal de Kettle, esto nos abrirá la interfaz de inicio de Kettle. Para la creación de la transformación se procede a clicar sobre nueva transformación. Como se observa en la figura 44, una transformación posee dos elementos, elementos para el diseño, en donde se

agrupan distintas utilidades que son utilizadas moviéndolas al panel de la transformación y la sección de vista.



**Figura 44** Interfaz de inicio de Pentaho Data Integrator.

La sección de vista como se ve en la figura 45 permite establecer conexión a las distintas bases de datos, observar los pasos que posee nuestra transformación entre otros. Desde este menú seleccionaremos “Database connection” para establecer una conexión a la base de datos transaccional donde tendremos los datos de entrada, y otra conexión a la base de datos intermedia, donde ubicaremos los datos una vez realizado el proceso ETC.

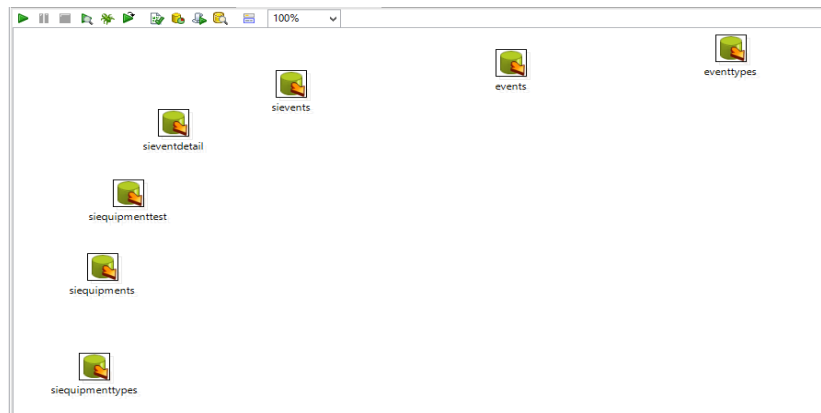


**Figura 45** Creación de una conexión a una base de datos.

Para la realización de todas las transformaciones fueron utilizados elementos de diseño, específicamente elementos de entrada (table input), de salida (table output), distintos elementos para transformar, como secuencias, concatenación, ordenamiento, etc, elementos de tipo unión (joiner) y scripts en java para la realización de la dimensión tiempo.

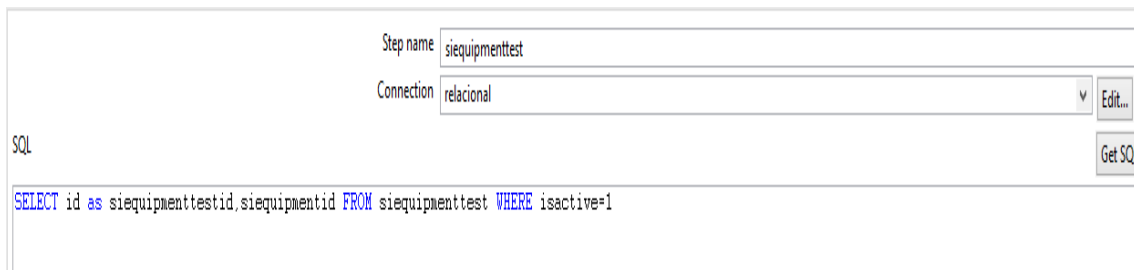
En el caso del tipo de evento registro técnico, es necesario extraer datos de las siguientes tablas: *Events*, *Eventstype*, *SiEvents*, *Sieventsdetails*, *Siequipment*, *siequipmenttest*,

*Siequipmenttesttype*. Para cada tabla es requerido utilizar un objeto TableInput, por lo que, como se observa en la figura 46, son utilizados 7 elementos de este tipo.



**Figura 46 Representación de las tablas de la base de datos en Kettle.**

Adicionalmente, cada table input debe ser configurado para establecer la fuente de datos a utilizar, la tabla, atributos y condiciones requeridas. En el caso *Siequipmenttest*, como se ve en la figura 47, se establece la conexión y la sentencia sql para obtener los datos necesarios.



**Figura 47 La extracción de elementos de una tabla se hace a partir de una consulta SQL.**

Una vez se han definido las tablas y como serán obtenidos los datos, es requerido unir las tablas para generar todos los datos de una instancia de la tabla eventos del área intermedia. Para esto son requeridos dos elementos que provee Kettle. Primero un objeto tipo Sort, porque Kettle requiere que los datos a utilizar en la unión estén ordenados por el atributo a utilizar en dicha unión, y un elemento de tipo unión. Una vez realizados, todos los ordenamientos y uniones respectivas, se tiene como resultado la figura 48.

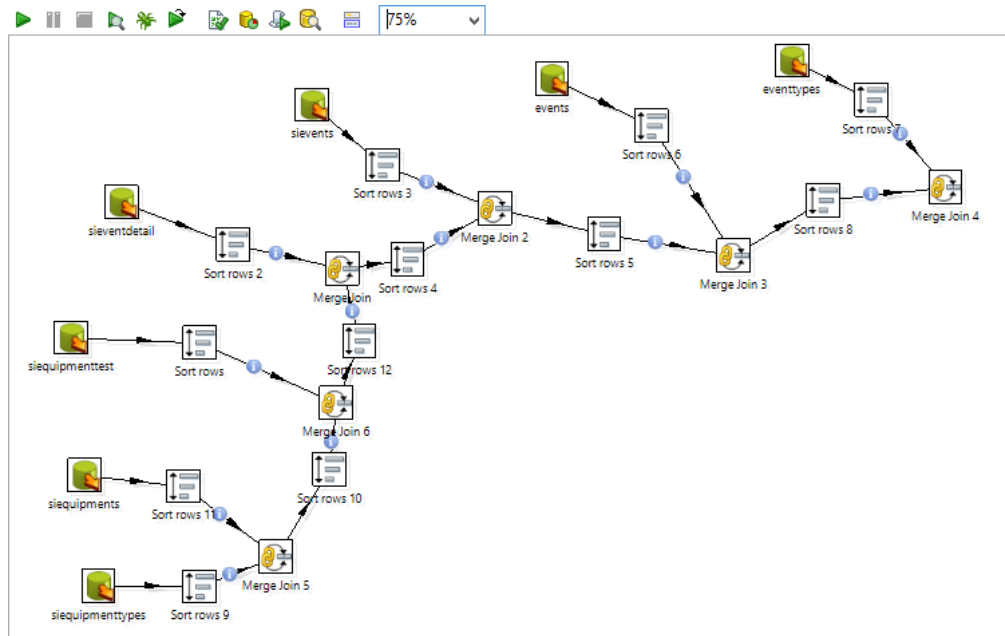


Figura 48 Flujo de trabajo para generar el proceso ETC de un tipo de eventos.

Realizado este proceso, es requerido concatenar los atributos que conforman el nombre del detalle. Para eso se utiliza un objeto del tipo concat fields, donde se ubican los distintos atributos a utilizar y genera como resultado, dichos atributos concatenados utilizando un separador establecido (Figura 49).

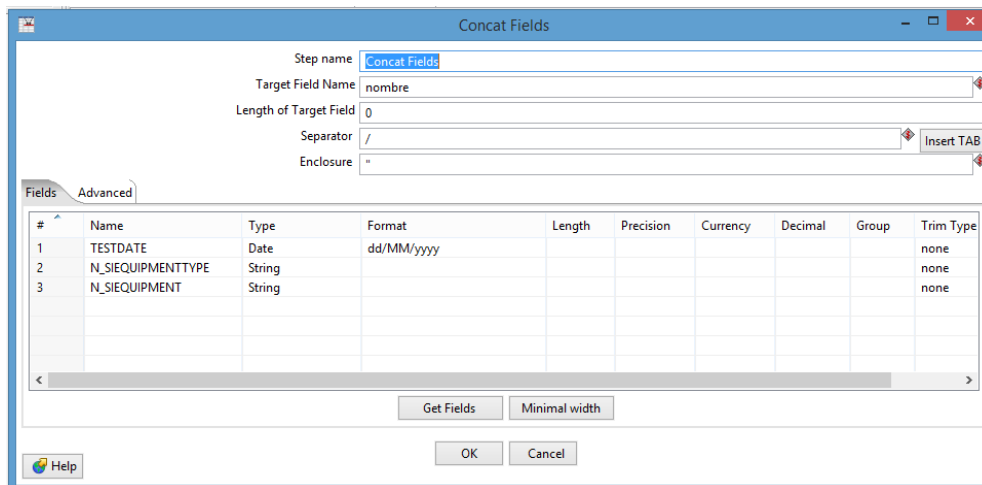
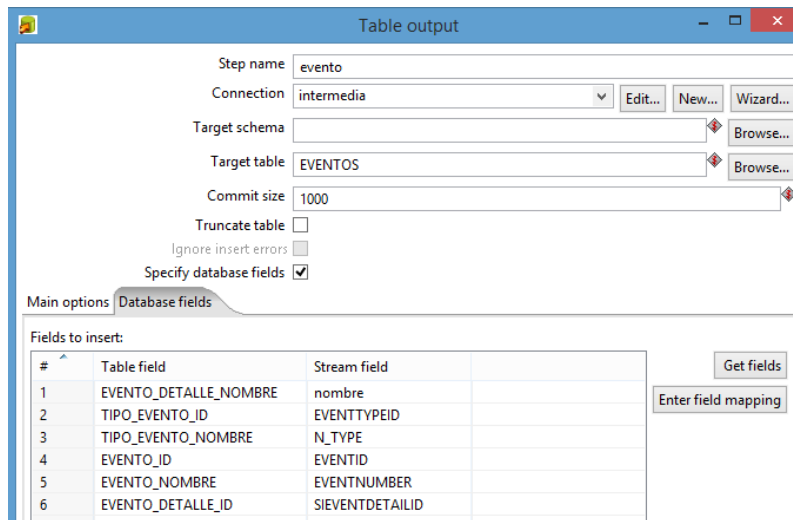


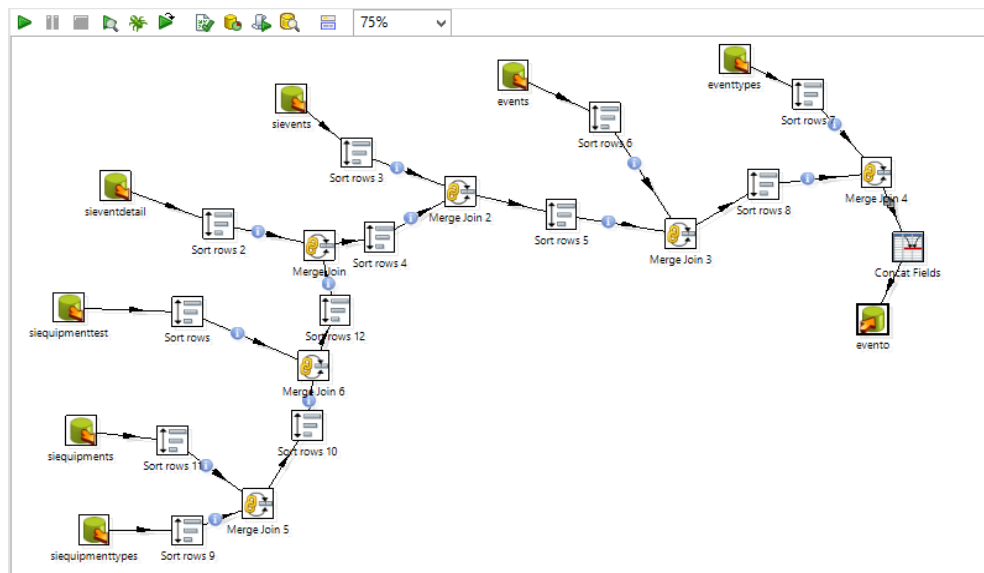
Figura 49 Forma de concatenar distintas columnas en Kettle.

Por último, se genera un elemento de tipo table output, que permitirá ubicar los datos en la tabla eventos del área intermedia, para esto, se establece la conexión al área intermedia, la tabla que será el destino y como serán llenado los atributos de dicha tabla (figura 50).



**Figura 50** En Kettle se requiere establecer como sera llenada cada columna de la tabla destino.

Con estos dos últimos pasos, el proceso ETC termina como se observa en la figura 51. Para ejecutar el proceso de ETC, se pulsa sobre el botón de iniciar (flecha verde), con lo que se ejecuta el proceso de ETC y se cargan los datos del evento tipo registro técnico.



**Figura 51** Proceso ETC para la tabla eventos del área intermedia.

Una vez realizado las tres transformaciones se procede a realizar el trabajo (job). Un job permite ejecutar un conjunto de transformaciones establecidas, lo que facilita y automatiza el proceso. Adicionalmente, un job permite pasar parámetros o variables específicos para el inicio de una transformación. Sin embargo, los Jobs generados en esta solución solo servirán para poder agrupar la ejecución de un conjunto de transformaciones.

Al igual que las transformaciones, un job está compuesto de un menú de vistas y un menú de diseño. El menú de vistas es el mismo que el de las transformaciones y en este caso, solo se utiliza para establecer la conexión a la base de datos. El menú de diseño sin embargo, como se observa en la figura 52, es distinto, esto porque un job permite ejecución de procesos o scripts concretos, mientras que una transformación está orientada a la manipulación de la fuente de datos.

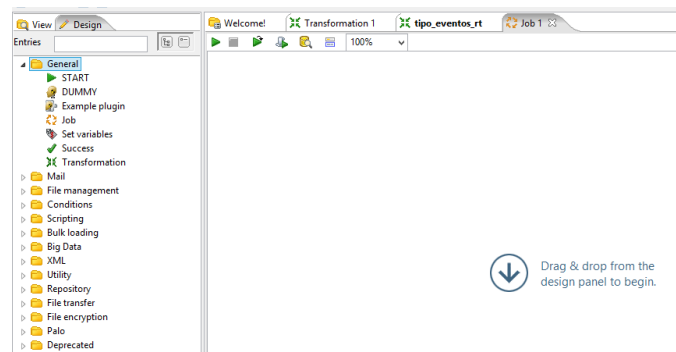


Figura 52 Pantalla para la creación de trabajo de Kettle.

Para llenar el área intermedia es requerido utilizar las tres transformaciones, ejecutándose una detrás de otra. Por lo que se utiliza un objeto del tipo Start (inicio en inglés) para establecer el punto de inicio del job y las tres transformaciones que representaran cada tipo de archivo. Como resultado se obtendrá la figura 53.

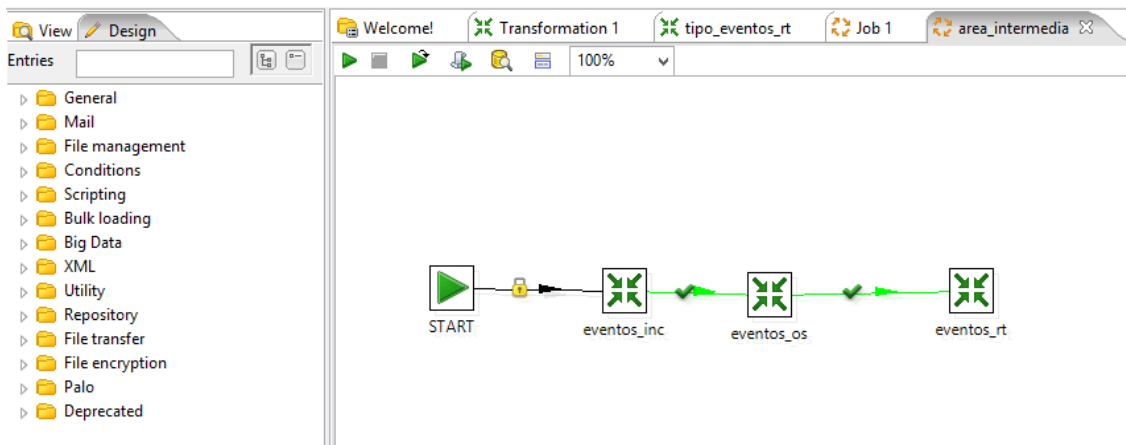


Figura 53 Trabajo que permite cargar los datos en el área intermedia.

Al ejecutar el job, todos los datos correspondientes al área intermedia son cargados desde la fuente de datos.

### 4.2.3.- Crear almacén de datos.

Para la construcción del almacén de datos se ha utilizado el modelo dimensional planteado previamente, utilizando la base datos Oracle, a partir de esto, la construcción del almacén de datos ha sido desarrollada utilizando la herramienta sqldeveloper y utilizando la siguiente implementación del modelo dimensional en función a un motor de base de datos relacional (figura 54).

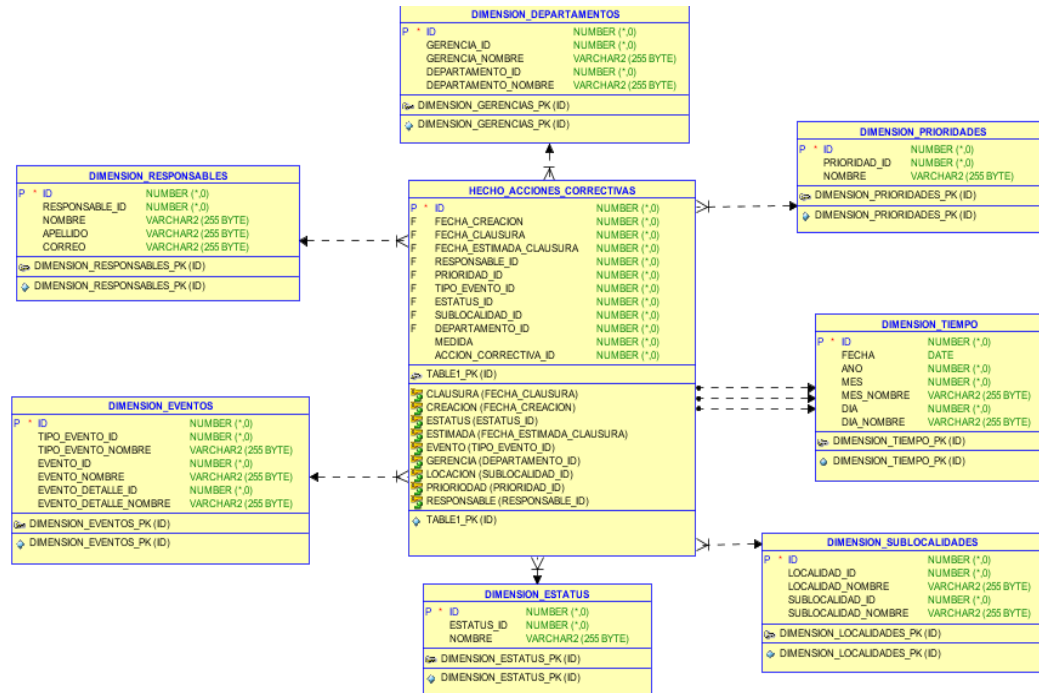


Figura 54 Modelo dimensional preparado para su implementación en la base de datos.

Con sqldeveloper se crearon las tablas requeridas así como el establecimiento del conjunto de restricciones (claves primarias y foráneas) para crear la implementación del modelo relacional (figura 55). Para efectos prácticos, la base de datos que soporta el modelo dimensional será llamada dimensional y la referencia de un atributo de una tabla será del tipo dimensional.nombre\_tabla.nombre\_atributo.

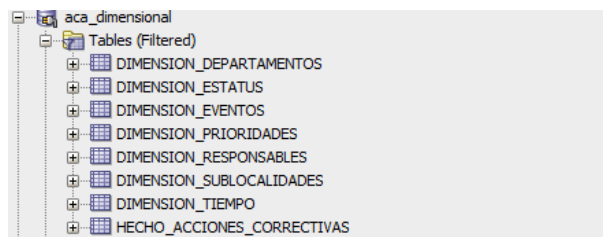


Figura 55 Base de datos creada para soportar el modelo dimensional.

A continuación se describirán todas las dimensiones y la tabla de hechos contemplada para la solución de inteligencia de negocios para el proceso de seguridad, higiene y ambiente:

- **Dimensión Estatus.**

Permite definir un estado específico a una acción correctiva. (Tabla 6).

Atributo	Tipo	Descripción	Procedencia
Id.	Integer (clave primaria).	Identificador unívoco de una instancia.	Generado en el proceso ETC.
Estatus_Id.	Integer.	Identificador unívoco utilizado en el ambiente transaccional.	Relacional.correctiveactionsstatus.Id.
Nombre.	Varchar(255 Byte).	Nombre del estatus.	Relacional.correctiveactionsstatus.name.

**Tabla 6 Detalle de la dimensión estatus.**

- **Dimensión prioridades.**

Permite establecer el nivel de prioridad a una acción correctiva para ser realizada y por tanto, la importancia o impacto de la realización de la misma (tabla 7).

Atributo	Tipo	Descripción	Procedencia
Id.	Integer (clave primaria).	Identificador unívoco de una instancia.	Generado en el proceso ETC.
Prioridad_Id.	Integer.	Identificador unívoco utilizado en el ambiente transaccional.	Relacional.Priority.Id.
Nombre.	Varchar(255 Byte).	Nombre del nivel de prioridad.	Relacional.Priority.name.

**Tabla 7 Detalle de la dimensión prioridades.**

- **Dimensión responsables.**

Permite gestionar los responsables de las acciones correctivas. El responsable es la persona encargada de velar que la realización de una acción correctiva sea exitosa (tabla 8).

Atributo	Tipo	Descripción	Procedencia
Id.	Integer (clave primaria).	Identificador unívoco de una instancia.	Generado en el proceso ETC.
Responsable_Id.	Integer.	Identificador unívoco utilizado en el ambiente transaccional.	Relacional.Responsable.Id.
Nombre.	Varchar(255 Byte).	Nombre de un responsable.	Relacional.Responsable.name.
Apellido.	Varchar(255 Byte).	Apellido de un responsable.	Relacional.Responsable.name.
Correo.	Varchar(255 Byte).	Correo del responsable.	Relacional.Responsable.email.

**Tabla 8 Detalle de la dimensión responsables.**

- **Dimensión departamentos.**

Establece a nivel organizacional, donde ocurre la acción correctiva en el proceso SHA.

Dicha ocurrencia se genera en un departamento, que forma parte de una gerencia (tabla 9).

Atributo	Tipo	Descripción	Procedencia
Id.	Integer (clave primaria).	Identificador unívoco de una instancia.	Generado en el proceso ETC.
Gerencia_Id.	Integer.	Identificador unívoco utilizado en el	Relacional.Managements.Id.

		ambiente transaccional.	
Gerencia_Nombre.	Varchar(255 Byte).	Nombre de una gerencia del sistema.	Relacional.Managements.name.
Departamento_Id.	Integer.	Identificador único utilizado en el ambiente transaccional.	Relacional. Departments.id.
Departamento_Nombre.	Varchar(255 Byte).	Correo de un departamento del sistema.	Relacional.Departments.name.

**Tabla 9 Detalle de la dimensión departamentos.**

- **Dimensión sublocalidades.**

Establece a nivel físico, donde ocurre la acción correctiva en el proceso SHA. Dicha ocurrencia se genera en una sublocación, que forma parte de una locación (tabla 10).

Atributo	Tipo	Descripción	Procedencia
Id.	Integer (clave primaria).	Identificador único de una instancia.	Generado en el proceso ETC.
Localidad_Id.	Integer.	Identificador único utilizado en el ambiente transaccional.	Relacional.Locations.Id.
Localidad _Nombre.	Varchar(255 Byte).	Nombre de una gerencia del sistema.	Relacional.Locations.name.
Sublocalidad_Id.	Integer.	Identificador único utilizado en el ambiente transaccional.	Relacional.Sublocations.id.

Sublocalidad_Nombre.	Varchar(255 Byte).	Correo de un departamento del sistema.	Relacional.Sublocations.name.
----------------------	--------------------	--	-------------------------------

Tabla 10 Detalle de la dimensión sublocalidad.

- **Dimensión eventos.**

La tabla dimensión y sus características son detalladas en la creación del área intermedia. Es la única dimensión que sus datos no provienen del sistema transaccional de forma directa, esto porque se requiere un procesamiento previo por los tipos de eventos y la forma de trabajar el detalle de los mismos (tabla 11).

Atributo	Tipo	Descripción	Procedencia
Id.	Integer (clave primaria).	Identificador único de una instancia.	Generado en el proceso ETC.
Tipo_evento_Id.	Integer.	Identificador único utilizado en el ambiente transaccional y pasado al área intermedia.	Intermedia.Eventos.Tipo_evento_Id.
Tipo_evento_nombre.	Varchar(255 Byte).	Nombre de un tipo de evento.	Intermedia.Eventos.Tipo_evento_nombre.
evento_Id.	Integer.	Identificador único utilizado en el ambiente transaccional y pasado al área intermedia.	Intermedia.Eventos.evento_Id.
evento_nombre.	Varchar(255 Byte).	Nombre de un evento.	Intermedia.Eventos.evento_nombre.

Evento_detalle_Id.	Integer.	Identificador unívoco utilizado en el ambiente transaccional y pasado al área intermedia.	Intermedia.Eventos.evento_detalle_Id.
Evento_detalle_nombre	Varchar(255 Byte).	Nombre del detalle de un evento.	Intermedia.Eventos.evento_detalle_nombre.

**Tabla 11 Detalle de la dimensión eventos.**

- **Dimensión tiempo.**

Permite analizar las acciones correctivas en función del momento en que se crea, se estima una clausura o se clausura una acción correctiva. Por esto, dicha dimensión tomara distintos roles en el modelo dimensional en función del criterio a utilizar. Una característica particular de la dimensión tiempo es que son generadas todas las fechas acorde a un período de tiempo en el proceso de ETC, para luego asociar las fechas específicas a las ocurrencias en las instancias. Esta forma de trabajo permite generar en los cuadros de mandos calendarios porque se poseen todas las fechas, para posteriormente solo obtener las acciones correctivas a partir de un rango de fechas establecido. La tabla 12 muestra una descripción de la dimensión tiempo.

Atributo	Tipo	Descripción	Procedencia
Id.	Integer (clave primaria).	Identificador unívoco de una instancia.	Generado en el proceso ETC.
Fecha.	Date.	Una fecha del calendario	Generado en el proceso ETC.
Ano.	Integer.	Año de la fecha.	Generado en el proceso ETC.
Mes.	Integer.	Mes de la fecha.	Generado en el proceso ETC.

Mes_nombre.	Varchar(255 Byte).	Nombre del mes.	Generado en el proceso ETC.
Día.	Integer.	Día de la fecha en la semana. Por ejemplo, el día domingo es el día 1.	Generado en el proceso ETC.
Día_nombre.	Varchar(255 Byte).	Nombre del día	Generado en el proceso ETC.

**Tabla 12 Detalle de la dimensión tiempo.**

- **Tabla hechos acciones correctivas.**

El modelo dimensional busca la representación de la ocurrencia de una acción correctiva caracterizada por las dimensiones. Por tanto la medida en el modelo dimensional descrito anteriormente es la incidencia de un hecho (tabla de hechos sin hechos) y no medidas específicas. Por tanto la tabla de hechos, como se observa en la tabla 13 está conformada por un conjunto de claves foráneas que enlazan al hecho con las dimensiones, una clave que identifica unívocamente ese hecho y un atributo medida con valor por defecto 1, que ayuda a comprender que cada instancia tiene el mismo valor.

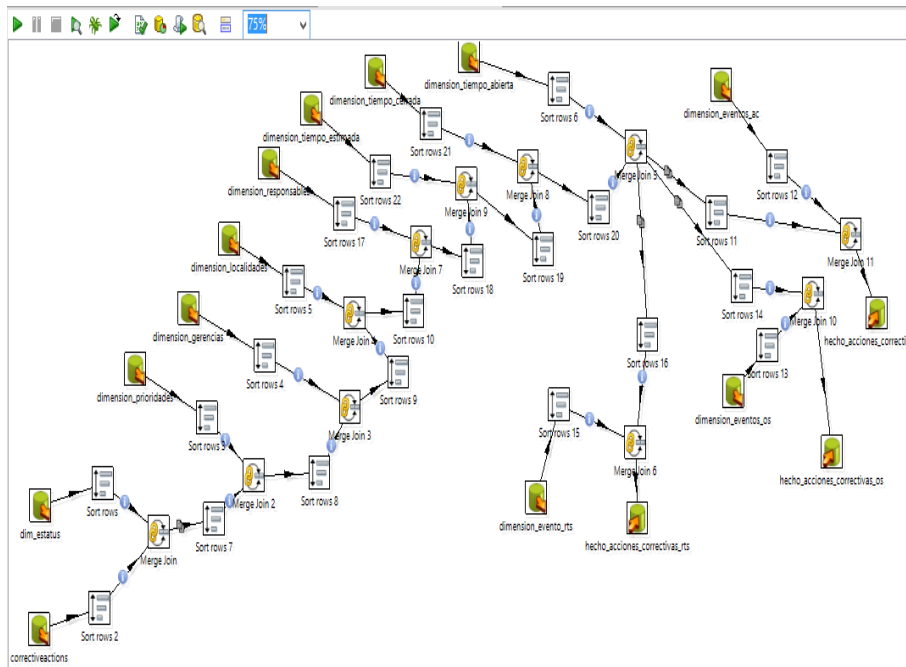
Atributo	Tipo	Descripción	Procedencia
Id	Integer (clave primaria).	Identificador unívoco de una instancia.	Generado en el proceso ETC.
Acción_correctiva_id	Integer.	Identificador unívoco utilizado en el ambiente transaccional.	
Fecha_creación	Integer (clave foránea).	Asociación con la dimensión tiempo.	Generado en el proceso ETC.
Fecha_clausura	Integer (clave foránea).	Asociación con la dimensión tiempo.	Generado en el proceso ETC.

Fecha_Estimada_clusura	Integer (clave foránea).	Asociación con la dimensión tiempo.	Generado en el proceso ETC.
Responsable_id	Integer (clave foránea).	Asociación con la dimensión responsables.	Generado en el proceso ETC.
Prioridad_id	Integer (clave foránea).	Asociación con la dimensión prioridades.	Generado en el proceso ETC.
Evento_id	Integer (clave foránea).	Asociación con la dimensión eventos.	Generado en el proceso ETC.
Estatus_id	Integer (clave foránea).	Asociación con la dimensión estatus.	Generado en el proceso ETC.
Sublocación_id	Integer (clave foránea).	Asociación con la dimensión sublocalidades.	Generado en el proceso ETC.
Departamento_id	Integer (clave foránea).	Asociación con la dimensión departamentos.	Generado en el proceso ETC.
Medida	Integer.	Forma de esclarecer que la medida de un evento es igual para cualquier instancia (1 por defecto).	Generado en el proceso ETC.

**Tabla 13 Detalle de la tabla de hecho de acciones correctivas.**

#### 4.2.4.- Detallar proceso ETC para almacén de datos.

Para los procesos ETC del área dimensional, también ha sido utilizado Pentaho Kettle. La realización de un proceso ETC para una dimensión es similar al mostrado en el punto 2.2. Sin embargo la realización del proceso ETC para la tabla de hechos conlleva un conjunto de elementos adicionales que serán descritos a continuación.



**Figura 56 Proceso ETC para la tabla de hechos.**

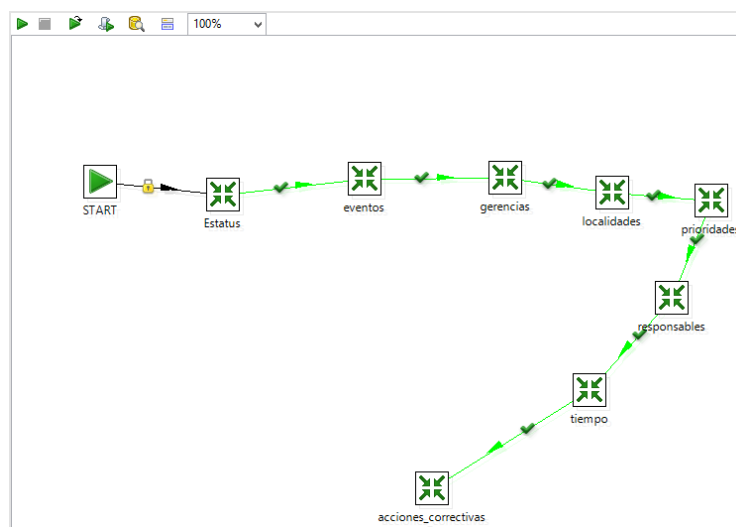
Para la realización del proceso de la tabla de hecho es requerido previamente ejecutar el conjunto de procesos ETC asociado a todas las dimensiones. La figura 56 muestra el proceso de ETC generado para la tabla de hechos.

Como se observa, a partir del uso de la tabla relacional *correctiveactions*, se van realizando una serie de uniones para asociar el conjunto de dimensiones. Esto porque una instancia de la tabla *correctiveactions* representa el nivel de granularidad asociado al modelo dimensional, lo que implica que se requiere ir agregando las claves asociadas a las dimensiones que caracterizan esa acción correctiva en particular.

Otro punto que se destaca es que al referenciarse de distintas formas, la dimensión tiempo es utilizada en tres distintos contextos, en el contexto de apertura, clausura y estimación de clausura.

Por último la necesidad de dividir el flujo del proceso de ETC se debe a la forma en como la tabla *correctiveactions* se asocia a los eventos según su tipo. Para el tipo incidentes/accidentes se usa el atributo *eventsid*. Para el tipo registro técnico se utiliza el atributo *Sieventsdetailsid* y para el tipo opsafes se utiliza el atributo *opeventsdetailsid*.

Debido a que existen atributos distintos para generar la relación acción correctiva evento en función de que una acción correctiva puede asociarse directamente a un evento o a un detalle de evento, una sola unión no permite ejemplificar esta situación. Esta situación requiere que se replique los datos que se tienen acumulados y se realicen uniones específicas según el tipo de evento.



**Figura 57 Trabajo para implementar el modelo dimensional.**

Una vez realizado este proceso se cargan los datos a la tabla de hechos. Una vez realizado todos los procesos de ETC, se genera el job que ejecutara todos estos procesos (figura 57).

#### 4.2.5.- Verificar calidad de los datos.

Una vez se ha realizado la inserción de los datos en el almacén de datos, es requerido comprobar que la carga ha sido correcta y que el almacén de datos representa de forma

acertada los datos del modelo relacional, para esto se realiza verificación en cada dimensión y en la tabla de hechos. Dicha verificación consiste en la aplicación de un conjunto de consultas en ambos ambientes y evaluar los resultados para establecer si no existen diferencias.

A continuación se presenta la verificación por cada elemento del modelo dimensional.

- **Dimensión prioridades.**

Debido a la sencillez de la dimensión, la validez de los datos puede realizarse visualizando los datos de la dimensión y de la tabla *priority* del ambiente transaccional (figura 58).

ID	PRIORIDAD_ID	NOMBRE
1	27	1 baja
2	28	2 media
3	29	3 alta

ID	NAME
1	1 baja
2	2 media
3	3 alta

Figura 58 Verificación de datos de la dimensión prioridades.

- **Dimensión estatus.**

Debido a la sencillez de la dimensión, la validez de los datos puede realizarse visualizando los datos de la dimensión y de la tabla *correctivaactionsstatus* del ambiente transaccional (figura 59). Los datos adicionales observados en el ambiente relacional no son tomados en cuenta porque no agregan valor al modelo dimensional.

ID	NAME	CREATEDDATE	LASTUPDATEDDATE	BYUSER	ISACTIVE
1	1 Abierto	14-JAN-07	14-JAN-07	vordonez	1
2	3 Cerrado	14-JAN-07	14-JAN-07	vordonez	1

ID	ESTATUS_ID	NOMBRE
1	61	1 Abierto
2	62	3 Cerrado

Figura 59 Verificación de datos de la dimensión estatus.

- **Dimensión responsables.**

Pa ra el caso de la dimensión responsables, se muestran los primeros registros ambas tablas (figura 60).

ID	RESPONSABLE_ID	NOMBRE	APELLIDO	CORREO	ID	NAME	EMAIL	CREATEDDATE	LASTUPDATEDDATE	BYUSER	ISACTIVE
1	123	Abrahan	Flores	aflores@ypergas.com	1	123Abrahan Flores	aflores@ypergas.com	14-OCT-15	14-OCT-18	vordonez	1
2	124	Abrahan	Gonzalez	agonzalez@ypergas.com	2	124Abrahan Gonzalez	agonzalez@ypergas.com	14-OCT-15	14-OCT-15	vordonez	1
3	126	Adolfo	Diaz	adiarz@ypergas.com	3	126Adolfo Diaz	adiarz@ypergas.com	14-OCT-15	14-OCT-15	vordonez	1
4	127	Adolfo	José	amoreno@ypergas.com	4	127Adolfo José Moreno Rojas	amoreno@ypergas.com	14-OCT-15	14-OCT-15	vordonez	1
5	128	Agustin	Guillen	aguillen@ypergas.com	5	128Agustin Guillen	aguillen@ypergas.com	14-OCT-15	14-OCT-15	vordonez	1
6	129	Alba	Martinez	amartinez@ypergas.com	6	129Alba Martinez	amartinez@ypergas.com	14-OCT-15	14-OCT-15	vordonez	1
7	130	Alberto	De	adefreitas@ypergas.com	7	130Alberto De Freitas	adefreitas@ypergas.com	14-OCT-15	14-OCT-15	vordonez	1
8	131	Alberto	Solana	asolana@ypergas.com	8	131Alberto Solana	asolana@ypergas.com	14-OCT-15	14-OCT-15	vordonez	1
9	132	Alejandro	Arrieta	aarrieta@ypergas.com	9	132Alejandro Arrieta	aarrieta@ypergas.com	14-OCT-15	14-OCT-15	vordonez	1
10	133	Alejandro	Diaz	adiarz@ypergas.com	10	133Alejandro Diaz	adiarz@ypergas.com	14-OCT-15	14-OCT-15	vordonez	1
	134	Alva	Teón	aleon@ypergas.com	11	134Alva Teón	aleon@ypergas.com	14-OCT-15	14-OCT-15	vordonez	1

Figura 60 Verificación de datos de la dimensión responsables.

Adicionalmente se consulta la cantidad de valores existentes en cada una de las tablas. Puede observarse además como el atributo nombre de la base de datos transaccional ha sido separado en nombre y apellido en la implementación del modelo dimensional (figura 61).

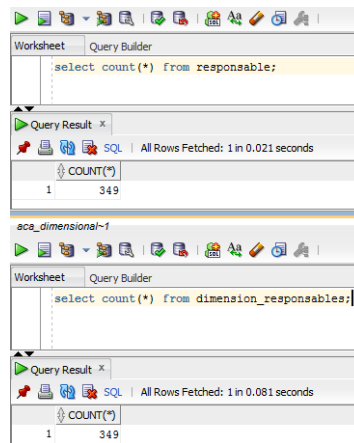


Figura 61 Verificación de la cantidad de responsables.

- **Dimensión departamentos.**

Para la verificación de datos de la dimensión departamentos, es realizada una consulta entre las tablas *departaments* y *managements*, que son las tablas que generan los datos de dicha dimensión (figura 62).

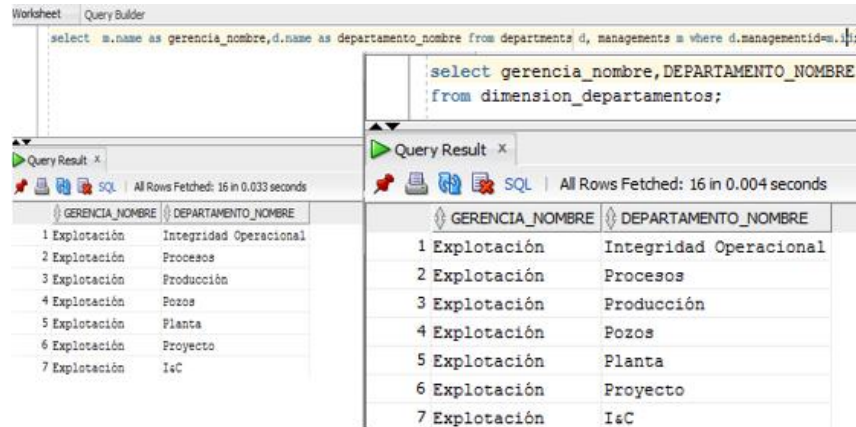


Figura 62 Verificación de datos de la dimensión departamentos.

- **Dimensión sublocalidades.**

Para la verificación de datos de la dimensión sublocalidades, es realizada una consulta entre las tablas *locations* y *sublocations*, que son las tablas que generan los datos de dicha dimensión (figura 63).

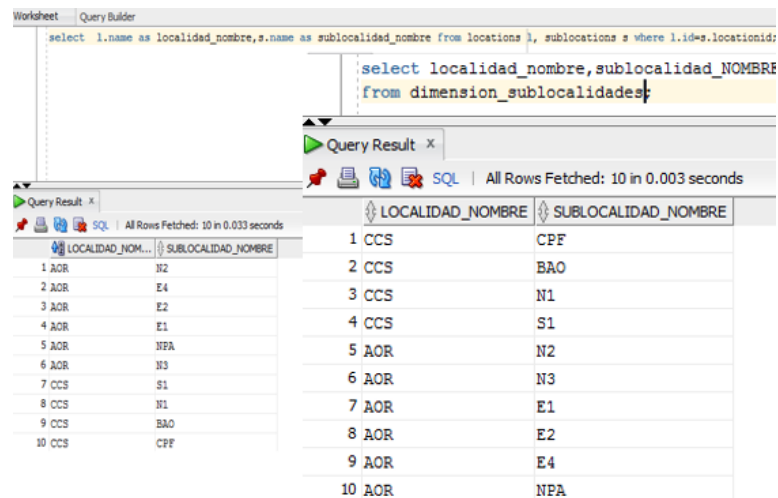


Figura 63 Verificación de datos de la dimensión sublocalidades.

- **Dimensión eventos.**

La validación de datos se ha realizado observando los eventos en función de su nombre y tipo (figura 64).

EVENTO_NOM...	TIPO_EVENTO_NOMBRE
1 OPER-2014	Operasafe
2 OPER-2015	Operasafe
3 RIA-1	Incidentes / Accidentes
4 RIA-10	Incidentes / Accidentes
5 RIA-11	Incidentes / Accidentes
6 RIA-12	Incidentes / Accidentes
7 RIA-13	Incidentes / Accidentes
8 RIA-14	Incidentes / Accidentes
9 RIA-15	Incidentes / Accidentes
10 RIA-16	Incidentes / Accidentes
11 RIA-17	Incidentes / Accidentes
12 RIA-18	Incidentes / Accidentes
13 RIA-19	Incidentes / Accidentes
14 RIA-2	Incidentes / Accidentes

ID	REFERENCE	EVENTNUM...	RECORDDATE	CREATEDDATE	LASTUPATEDDAT
1	1 (null)	OPER-2014	01-DEC-14	01-DEC-14	31-MAY-15
2	11 (null)	OPER-2015	05-FEB-15	30-MAY-15	30-MAY-15
3	6 (null)	RIA-1	09-FEB-15	09-FEB-15	31-MAY-15
4	17 (null)	RIA-10	29-MAY-15	30-MAY-15	31-MAY-15
5	18 (null)	RIA-11	11-MAY-15	30-MAY-15	31-MAY-15
6	19 (null)	RIA-12	10-MAY-15	30-MAY-15	31-MAY-15
7	20 (null)	RIA-13	21-MAY-15	30-MAY-15	31-MAY-15
8	21 (null)	RIA-14	26-MAY-15	30-MAY-15	31-MAY-15
9	22 (null)	RIA-15	18-MAY-15	30-MAY-15	30-MAY-15
10	23 (null)	RIA-16	25-MAY-15	30-MAY-15	31-MAY-15
11	24 (null)	RIA-17	25-MAY-15	30-MAY-15	31-MAY-15
12	25 (null)	RIA-18	20-MAY-15	30-MAY-15	31-MAY-15
13	26 (null)	RIA-19	31-MAY-15	30-MAY-15	31-MAY-15
14	7 (null)	RIA-2	20-MAY-15	06-MAY-15	31-MAY-15

Figura 64 Verificación de datos de la dimensión eventos.

- **Dimensión tiempo.**

La dimensión no requiere una verificación de datos con respecto a la base de datos transaccional, sin embargo puede confirmarse que los datos son consistentes en función de una fecha (figura 65).

ID	FECHA	ANO	MES	MES_NOMBRE	DIA	DIA_NOMBRE
1	9574 15-MAR-19	2019	3	Marzo	15	Viernes
2	9575 16-MAR-19	2019	3	Marzo	16	Sabado
3	9576 17-MAR-19	2019	3	Marzo	17	Domingo
4	9577 18-MAR-19	2019	3	Marzo	18	Lunes
5	9578 19-MAR-19	2019	3	Marzo	19	Martes
6	9579 20-MAR-19	2019	3	Marzo	20	Miercoles
7	9580 21-MAR-19	2019	3	Marzo	21	Jueves
8	9581 22-MAR-19	2019	3	Marzo	22	Viernes
9	9582 23-MAR-19	2019	3	Marzo	23	Sabado
10	9583 24-MAR-19	2019	3	Marzo	24	Domingo
11	9584 25-MAR-19	2019	3	Marzo	25	Lunes
12	9585 26-MAR-19	2019	3	Marzo	26	Martes
13	9586 27-MAR-19	2019	3	Marzo	27	Miercoles
14	9587 28-MAR-19	2019	3	Marzo	28	Jueves

Figura 65 Verificación de datos de la dimensión tiempo.

- **Tabla hechos acciones correctivas.**

Para la comprobación de que la tabla de hechos ha sido cargada efectivamente, primero se comprueba que la cantidad de instancias corresponde a la cantidad de acciones correctivas registradas en la base de datos transaccional (figura 66).

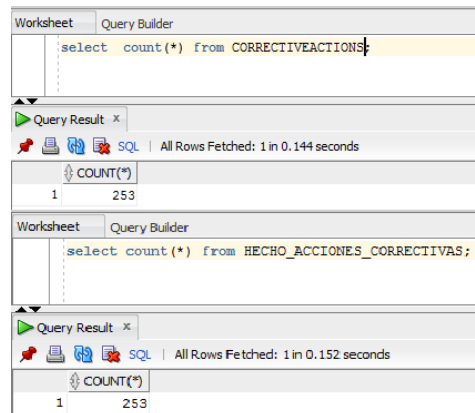


Figura 66 Verificación de la cantidad de acciones correctivas.

También se ha realizado verificación para comprobar que la unión del hecho con las dimensiones ha sido efectiva. En este caso se ha comprobado en función de un rango de fechas de creación de acciones correctivas (figura 67).

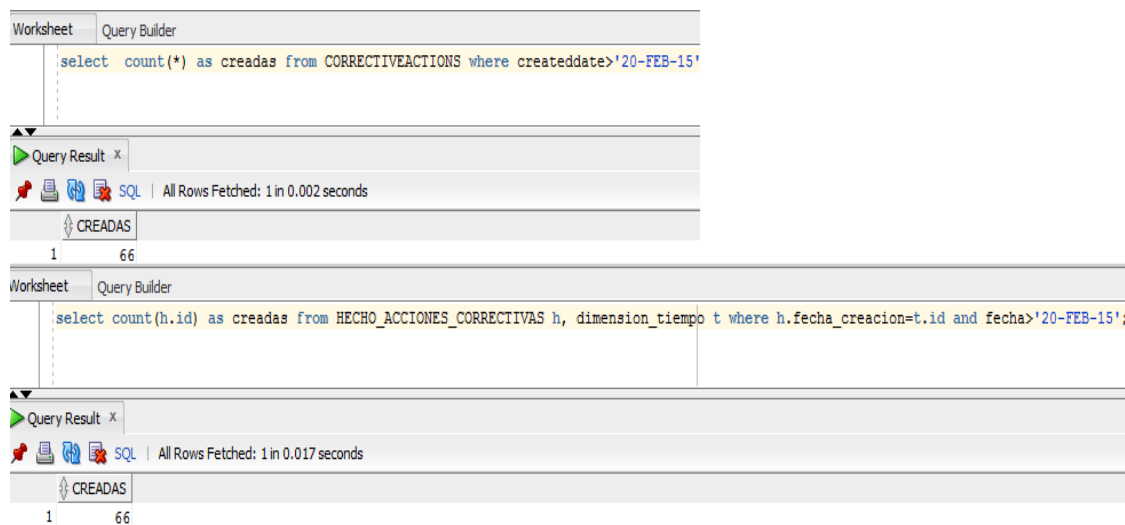


Figura 67 Verificación de la tabla de hechos a partir de una consulta.

#### 4.2.6.- Verificación de la calidad de los datos de los componentes del modelo implementado.

Adicionalmente a la verificación de los datos por tabla, se ha realizado una verificación para comprobar que efectivamente los datos del modelo relacional han sido transformados y cargados en el modelo dimensional generando el modelo estrella deseado.

Para esto hay que comprender que las realización del modelo estrella y la tabla de hecho y dimensiones, ocasiona que se generen nuevas claves primarias (y por ende foráneas) para el

registro de cualquier elemento. Por lo que, la inserción de dicho elemento en la tabla de hechos conlleva a que se haga una relación a partir de estas claves generadas en el proceso ETC. Por tanto, la verificación del modelo implementado implica comprobar que las distintas claves foráneas en la tabla de hecho, corresponden efectivamente a los datos que se quiere enlazar en las distintas dimensiones. Por tanto la verificación se realiza del modo tabla de hecho una dimensión, llegando a evaluar todas las dimensiones del modelo.

- **Verificación tabla de hechos con dimensión tiempo.**

Para la comprobación de que efectivamente se están representando los datos del modelo relacional con respecto a la dimensión tiempo. Se ha observado un registro de la tabla de hechos y comprobado las claves foráneas, observando que datos muestran en la dimensión tiempo, luego se ha observado la tabla *correctiveactions* que posee esta información en el modelo operacional (figura 68). Se observa además que se visualizan tres claves porque la dimensión tiempo es multivaluada en el modelo.

Tabla de hechos acciones correctivas

ID	ACCION_CORRECTIVA_ID	FECHA_CREACION	FECHA_CLAUSURA	FECHA_ESTIMADA_CLAUSURA
21760	64	18345	18367	18343
21762	69	18337	18354	18354
21764	70	18284	18287	18287
21767	71	18343	(null)	18345
21769	73	18323	(null)	18352
21771	75	18343	(null)	18352
21773	76	18343	18351	18351
21775	74	18344	18351	18351
21777	77	18346	(null)	18352
21778	78	18350	18351	18351
21781	79	18339	18369	18369
21594	232	18226	18232	18232

Dimensión tiempo						Tabla modelo relacional correctiveactions			
ID	FECHA	ANO	MES	MES_NOMBRE	DIA	DIA_NOMBRE	CREATEDDATE	ESTIMATEDCLOSUREDATE	CLOSUREDATE
18348	07-MAR-15	2015	3	Marzo	7	Sabado	76 02-MAR-15	10-MAR-15	10-MAR-15
18349	08-MAR-15	2015	3	Marzo	8	Domingo	77 05-MAR-15	11-MAR-15	(null)
18350	09-MAR-15	2015	3	Marzo	9	Lunes	78 09-MAR-15	10-MAR-15	10-MAR-15
18351	10-MAR-15	2015	3	Marzo	10	Martes	79 26-FEB-15	28-MAR-15	28-MAR-15
18352	11-MAR-15	2015	3	Marzo	11	Miercoles	80 30-MAR-15	02-APR-15	(null)
18353	12-MAR-15	2015	3	Marzo	12	Jueves	81 22-JAN-15	03-FEB-15	(null)
...	...	...	...	...	...	...	82 25-MAR-15	01-APR-15	(null)

Figura 68 Comprobación de datos del modelo con la dimensión tiempo.

- **Verificación tabla de hechos con dimensión estatus.**

En este caso, debido a que el nombre estatus solo permite dos nombres (abierto, cerrado) se ha verificado dentro de la tabla de hecho dos registros que posean estos estatus, a partir del *estatus\_id*, se ha verificado que corresponden a los dos nombres posible. Luego se ha

realizado una consulta en el modelo relacional que permita ver que efectivamente los datos corresponden (figura 69).

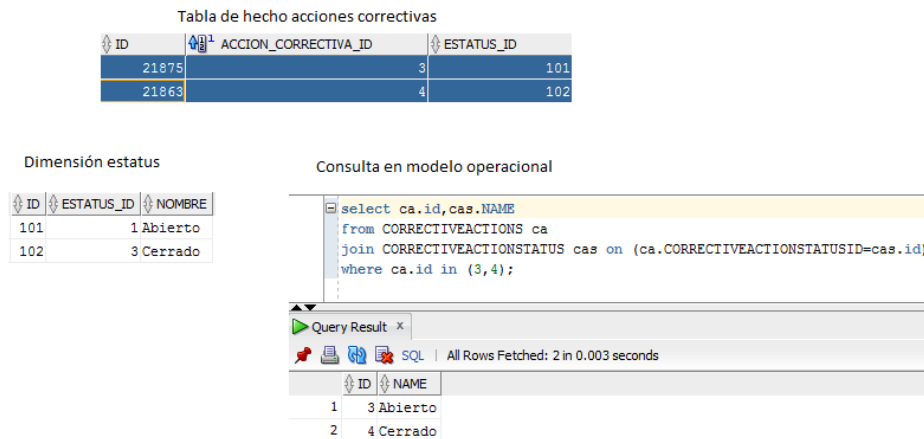


Figura 69. Comprobación de datos del modelo con la dimensión tiempo.

- **Verificación tabla de hecho con dimensión prioridades.**

Como en el caso de la dimensión estatus, debido a la poca variabilidad de la dimensión, la verificación se realiza observado acciones correctivas que cumplan las distintas prioridades existentes. Una vez seleccionadas las acciones correctivas se visualizan los datos en la dimensión prioridades y se realiza una consulta sobre el modelo operacional para evaluar si efectivamente los datos corresponden (figura 70).

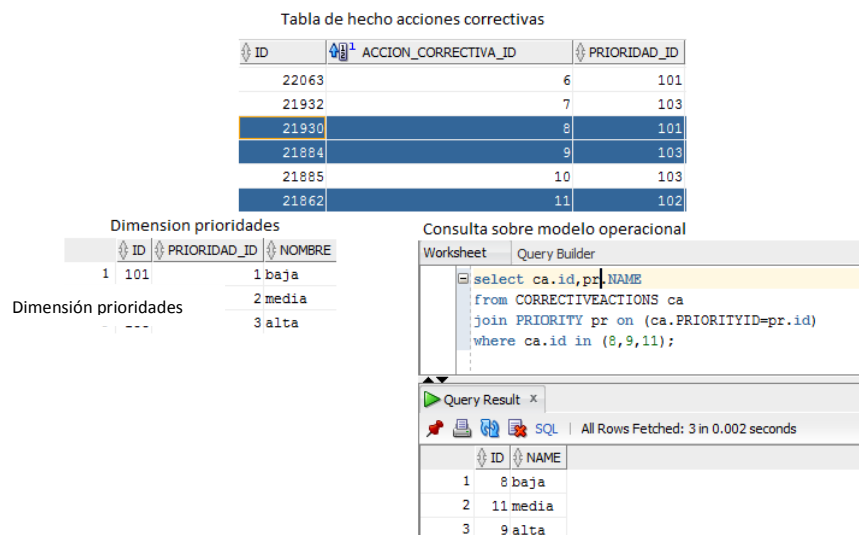


Figura 70. Comprobación de datos del modelo con la dimensión prioridades.

- **Verificación tabla de hechos con dimensión sublocalidades.**

Para la dimensión sublocalidades se comprueba dado un registro, sus valores en la dimensión localidades y su veracidad evaluando el modelo operacional según la acción correctiva seleccionada (figura 71).

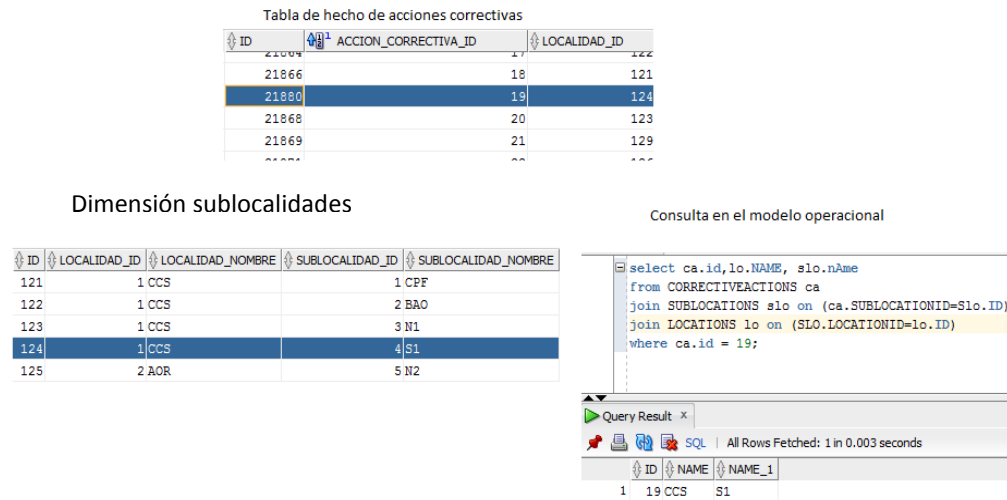


Figura 71. Comprobación de datos del modelo con la dimensión sublocalidades.

- **Verificación tabla de hechos con dimensión departamentos.**

De la misma forma que se ha realizado en la dimensión departamentos se ha comprobado a partir de la selección de un registro de la tabla de hechos, navegando a la dimensión departamentos para visualizar los datos y compararlos con lo que se poseen en el modelo operacional (figura 72).

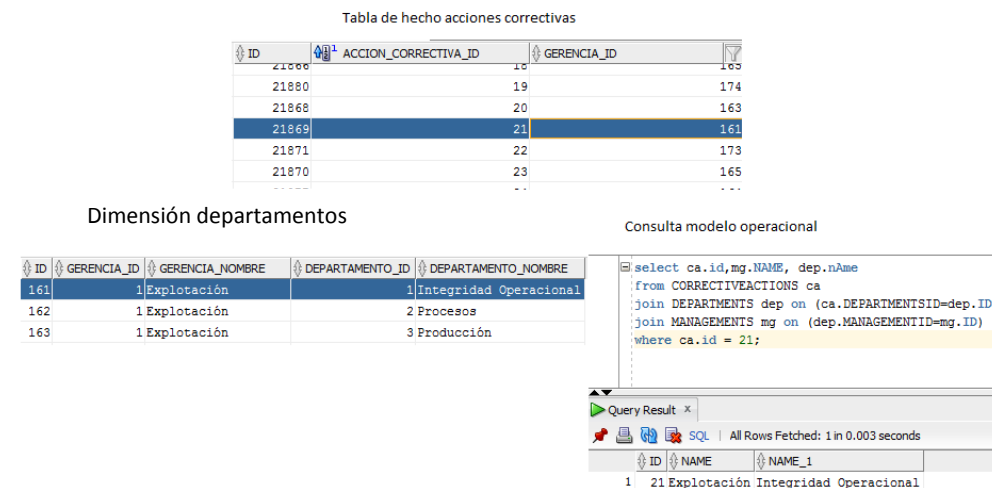


Figura 72. Comprobación de datos del modelo con la dimensión departamentos.

- **Verificación tabla de hechos con dimensión eventos.**

Para la dimensión eventos se comprueba a partir de una acción correctiva los datos que posee la dimensión eventos y se compara con lo que se tiene en el sistema operacional (figura 73). Se resalta que la forma de relacionarse con la tabla *correctiveactions* varía en función del tipo de evento, por lo que consulta de comprobación varía en función del tipo.

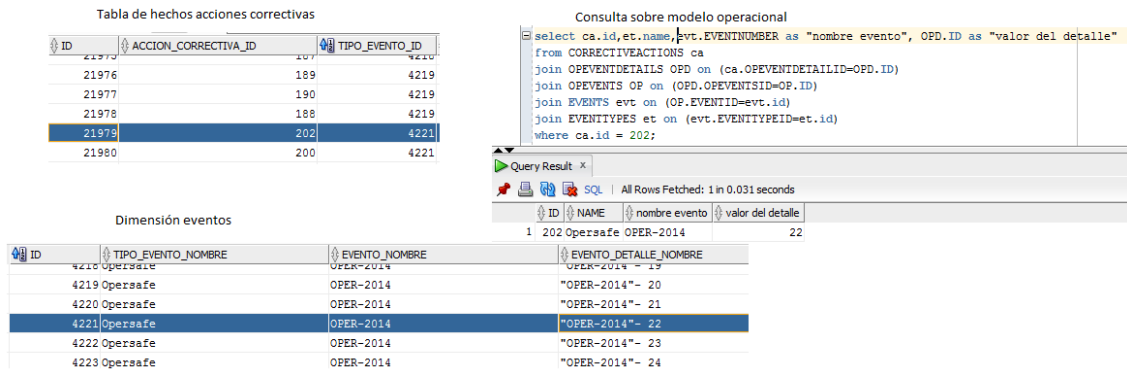


Figura 73. Comprobación de datos del modelo con la dimensión eventos.

- **Verificación tabla de hechos con dimensión responsables.**

Se comprueba que la referencia a la dimensión responsables corresponde a lo que se tiene en el sistema operacional (figura 74).

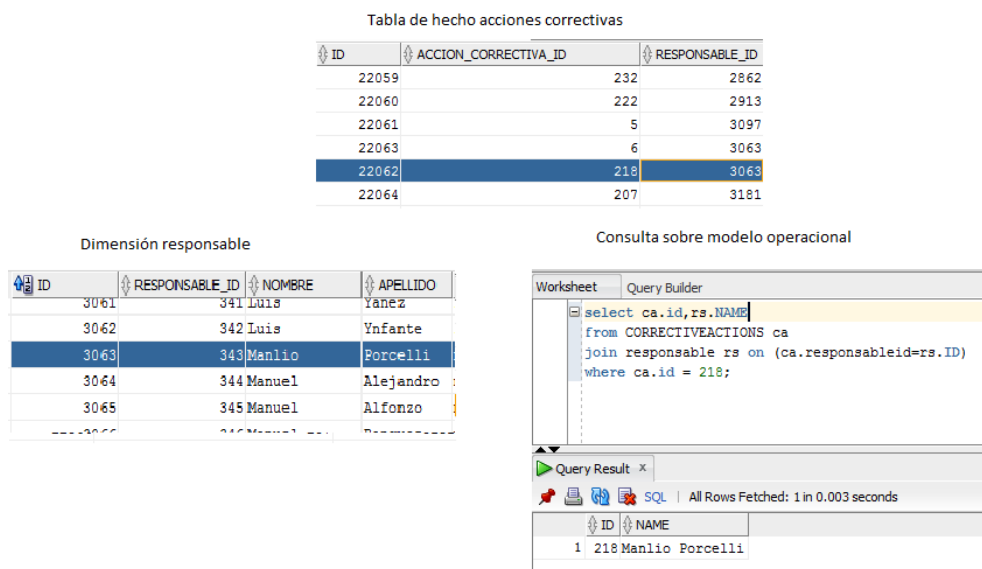


Figura 74. Comprobación de datos del modelo con la dimensión responsables.

#### 4.2.7.- Detallar requerimientos según modelo.

Desde el estudio del modelo dimensional, que plasma los datos relevantes del sistema transaccional, específicamente el núcleo de acciones correctivas del departamento de

seguridad, higiene y ambiente (SHA) y utilizando los gráficos que se estaban generando desde el sistema transaccional para buscar satisfacer las necesidades analíticas, se han propuesto un conjunto de requerimientos analíticos presentados a continuación:

1. Cantidad de acciones correctivas por gerencia.
2. Cantidad de acciones correctivas por sublocalidad.
3. Cantidad de acciones correctivas por responsable.
4. Cantidad de acciones correctivas por estado.
5. Cantidad de acciones correctivas por prioridad.
6. Cantidad de acciones correctivas por evento.
7. Cantidad de acciones correctivas a partir de la fecha de creación, fecha estimada de clausura y clausura.
8. Establecer que acciones correctivas están por vencer o han vencido.
9. Conocer la prioridad de las acciones correctivas por estado.
10. Identificar qué acciones correctivas no se cumplió con la fecha estimada de cierre.
11. Poder identificar acciones correctivas que posean un mismo patrón, por ejemplo, saber que responsables tienen más de 5 acciones correctivas de prioridad alta.
12. Distribución de las acciones correctivas por evento en el tiempo.

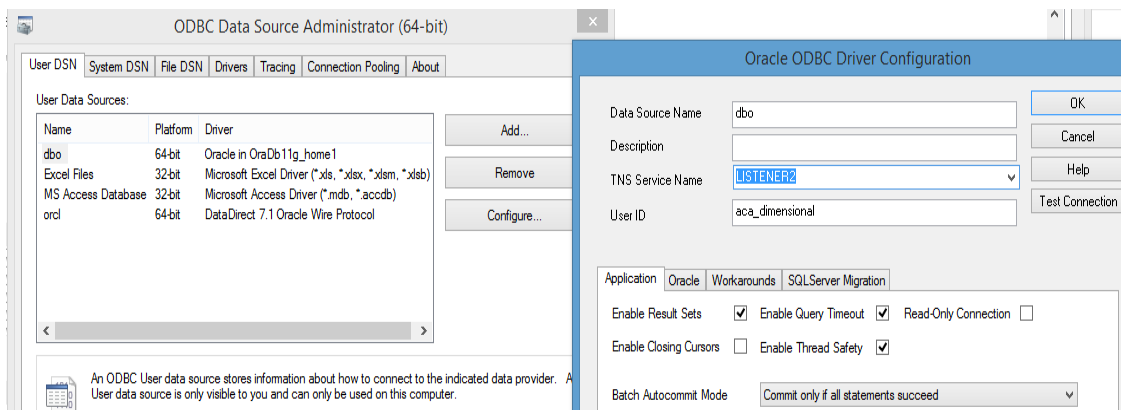
#### **4.3.- Desarrollar requerimientos de información.**

Esta fase consiste en la creación de los indicadores de gestión usando la herramienta de inteligencia de negocios (Qlikview), de forma tal que puedan satisfacerse los requerimientos analíticos planteados previamente. Para esto se debe configurar desde la herramienta a utilizar la conexión al modelo dimensional, realizar las consultas que satisfacen los requerimientos analíticos para posteriormente generar un cuadro de mando que permita de forma sencilla poder visualizar los distintos indicadores dispuesto de una forma ordenada. Por último se realiza una verificación de los datos para confirmar la exactitud de los indicadores.

### 4.3.1.- Configuración del administrador.

Implica integrar la implementación del modelo dimensional ubicado en una base de datos Oracle en Qlikview, de forma que se puedan manipular los datos dentro de la herramienta de inteligencia de negocios.

La conexión a una fuente de datos, en este caso el modelo dimensional del departamento SHA desde Qlikview implica la utilización de un driver que enlace con el motor de base de datos que se utiliza y establecer las credenciales de conexión. Figura 75.



**Figura 75 Creación de conexión con la fuente de datos desde Qlikview.**

Una vez configurada la conexión al modelo dimensional, desde Qlikview se construye un script que especifica que elementos de la base de datos serán utilizados, adicionalmente en este script pueden realizarse transformaciones a los datos.

Qlikview enlaza las distintas tablas a partir de atributos clave, un atributo clave es aquel que su nombre que se repite en dos tablas, Qlikview realiza una unión entre dichas tablas por lo que se puede navegar entre los atributos de dichas tablas. Esta forma de trabajo en Qlikview implica que dentro de nuestro script, los atributos que hacen relación entre las tablas a partir de claves foráneas, sean renombrados para ser atributos clave y la herramienta detecte que las tablas se conectan. Como observamos en la figura 76 los identificadores son renombrados a la forma como son llamados en la tabla de hechos acciones correctivas.

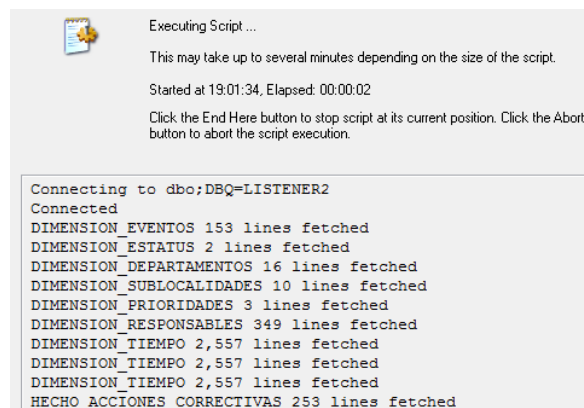
```

13 //----- Start Multiple Select Statements -----
14 SQL SELECT id as id_evento,
15 tipo_evento_id,
16 tipo_evento_nombre,
17 evento_id,
18 evento_nombre,
19 evento_detalle_id,
20 evento_detalle_nombre
21 FROM "ACA_DIMENSIONAL"."DIMENSION_EVENTOS";
22
23 SQL SELECT id as id_estatus,
24 estatus_id,
25 nombre as estatus_nombre
26 FROM "ACA_DIMENSIONAL"."DIMENSION_ESTATUS";
27
28 SQL SELECT id as id_gerencia,
29 gerencia_id,
30 gerencia_nombre,
31 departamento_id,
32 departamento_nombre
33 FROM "ACA_DIMENSIONAL"."DIMENSION_GERENCIAS";

```

**Figura 76** El script dentro de Qlikview define que datos se manejarán en el cuadro de mando.

Una vez terminado el script, se procede a su ejecución, lo que trae como resultado que el modelo dimensional y sus datos puedan ser utilizados dentro de Qlikview (figura 77). Se puede observar como la dimensión tiempo aparece tres veces, esto es debido a que dentro del modelo, la dimensión tiempo cumple distintos roles (dimensión multivaluada) y este genera que los datos de dicha dimensión se tengan que replicar para cada uno de los roles.



**Figura 77** Ejecución del script.

#### 4.3.2.- Generación de indicadores de gestión.

Disponibles los datos en la herramienta de inteligencia de negocios se procede a desarrollar los indicadores de gestión que cumplan con la propuesta piloto de requerimientos analíticos señalados previamente.

La construcción de los indicadores de gestión implicó la evaluación de las distintas formas como pueden observarse los mismos y la forma de distribuir los indicadores de gestión dentro del cuadro de mando de forma que se tuviera una lógica al ser observados

### 4.3.3.- Desarrollo y distribución de consultas en el cuadro de mando.

Dentro de Qlikview se pueden generar distintos cuadros de mando o sheets (hojas en inglés) dentro de un proyecto o archivo .qvw. De esta forma se puede organizar los datos de un modelo dimensional en bloques separados.

Para la realización de las consultas, se ha optado por generar cuatro cuadros de mando (cada hoja corresponde a un cuadro de mando dentro del proyecto). El primer cuadro de mando muestra un vistazo general del modelo implementado como muestra la figura 78

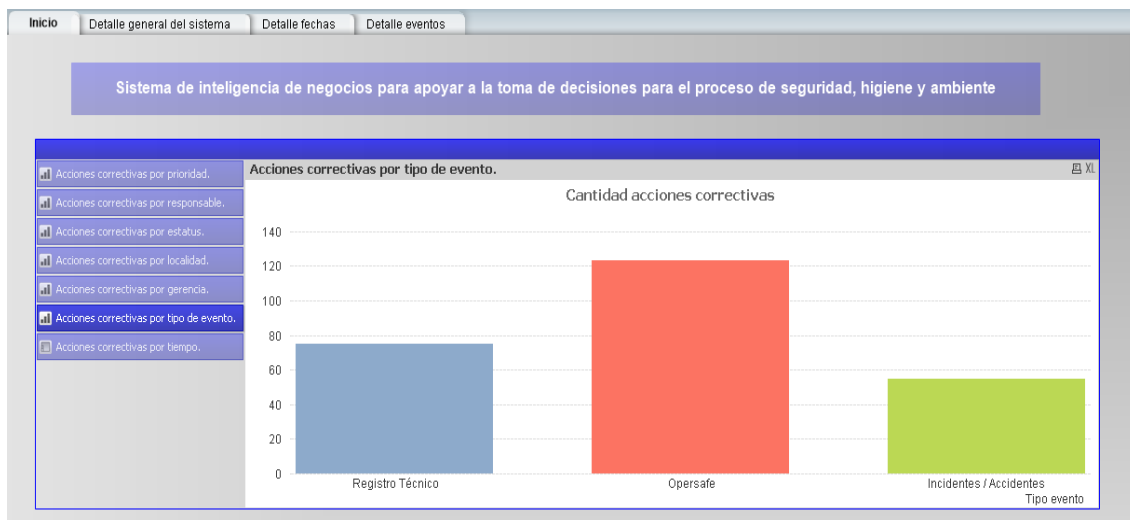


Figura 78 Cuadro de mando inicio.

Se puede navegar dentro de las distintas dimensiones observando el máximo nivel de agregación en cada una de ellas. Adicionalmente, debido a que la dimensión tiempo es multivaluada, puede navegarse entre los distintos roles que dicha dimensión toma (figura79).



Figura 79 La dimensión tiempo y sus distintos roles en el modelo.

El segundo cuadro de mando permite observar las acciones correctivas desde la perspectiva de distintas dimensiones y jerarquías de las mismas con mayor detalle (figura 80).

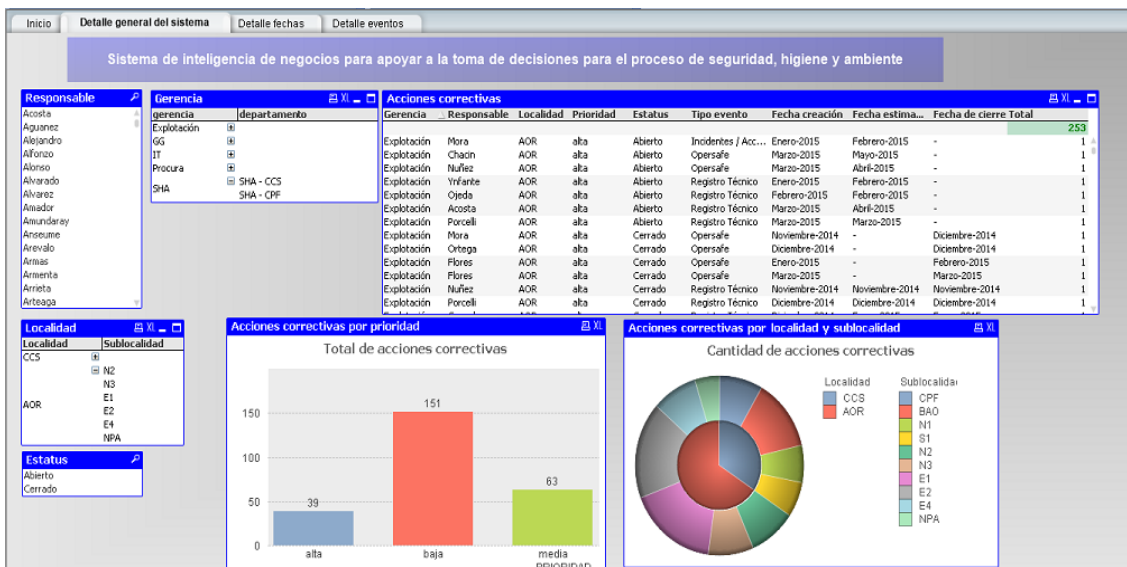


Figura 80 Cuadro de mando general.

Desde este cuadro de mando, puede observarse como se aplican distintos filtros y gráficos para poder observar los datos desde diferentes enfoques. La ventaja de agrupar distintos indicadores en un mismo cuadro de mando radica en que la selección en un elemento, altera todos los demás, por ejemplo, como se observa en la figura 81, si se selecciona localidad “CCS” con Estatus “Abierto”, los demás elementos muestran que posibles valores coinciden

con la selección, por esto, solo se muestran las gerencias donde ocurrieron acciones correctivas con la selección establecida.

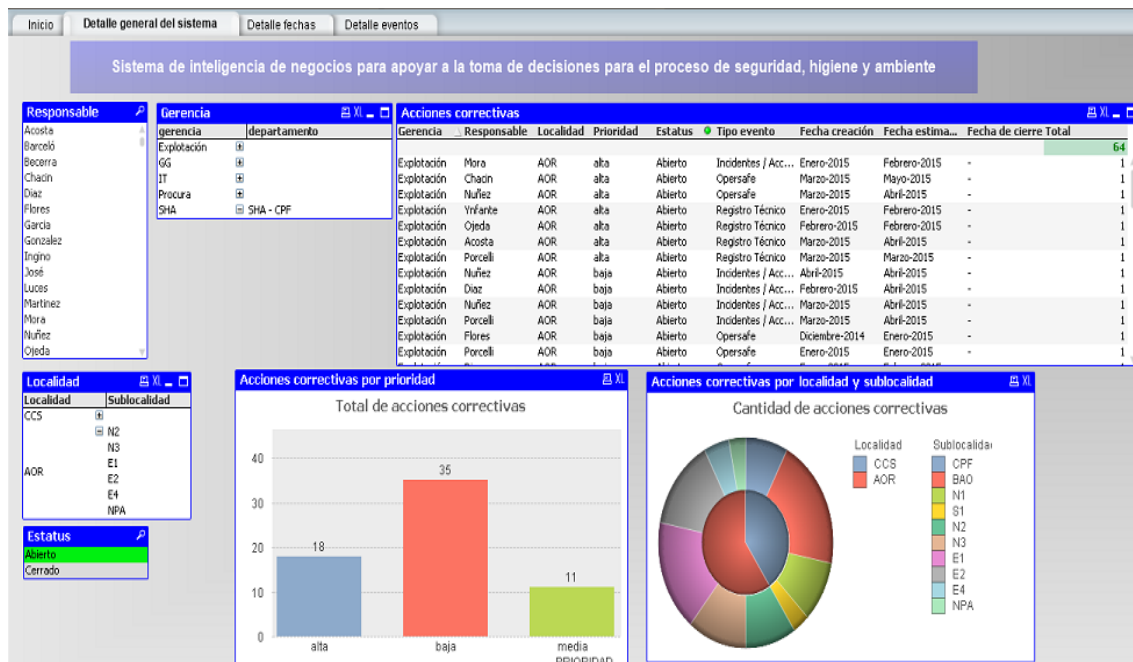


Figura 81 Los elementos del cuadro de mando se adaptan a la selección en un elemento.

A continuación se detallara los elementos que conforman este primer cuadro de mando.

- **Acciones correctivas por gerencia.**

Se puede observar desde el cuadro general (figura 82), En este caso, la figura xx permite seleccionar una gerencia específica y adicionalmente usando una operación drill down, especificar un departamento específico. Como se mencionó anteriormente, por un elemento seleccionado puede observarse la modificación en los demás elementos, por lo que este indicador específico no requiere explícitamente la cantidad de acciones correctivas, porque implícitamente otros indicadores reflejan dicha información.

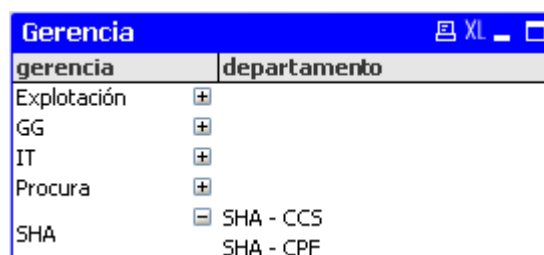


Figura 82 Dimensión Gerencias donde puede navegarse a partir de la jerarquía establecida.

- **Acciones correctivas por estatus.**

Para poder manipular la dimensión estatus en el cuadro de mando general, se provee el elemento identificado en la figura 83.



Figura 83 Forma de manipular los valores posibles de la dimensión estatus.

- **Acciones correctivas por responsable.**

Se provee una lista con los apellidos de los responsable (la dimensión responsables podrá verse desde otras perspectivas en otros cuadros de mando), adicionalmente Qlikview permite realizar una busca rápida sobre una lista de valores (figura 84).

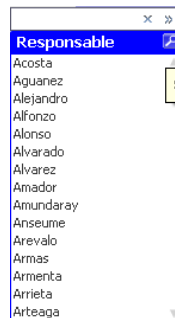


Figura 84 Lista de responsables, puede buscar por un apellido específico.

- **Acciones correctivas por prioridad.**

Para este caso, se utiliza un gráfico para visualizar la cantidad de prioridades según las acciones correctivas. Qlikview ofrece un abanico de acciones a lo que se refiere construcción de gráficos. En este caso se utiliza un gráfico de barras (figura 85).

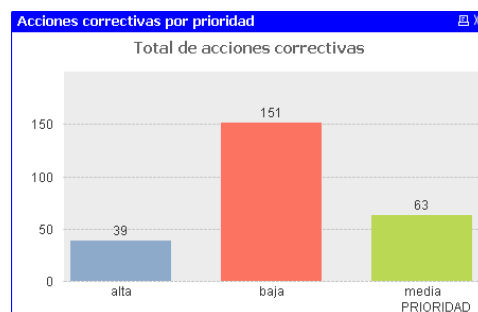


Figura 85 Representación de las prioridades de las acciones correctivas.

- **Cantidad de acciones correctivas por sublocalidad.**

En este caso se utilizan dos elementos para visualizar las sublocalidades. Como se observa en la figura 86 se tiene una tabla que permite realizar operaciones de drill down y roll up en base a la jerarquía establecida en la dimensión sublocalidades (localidad y sublocalidad) y un gráfico de torta donde se puede visualizar por localidad, la cantidad de acciones correctivas (el desplazar el curso sobre un segmento del gráfico genera que Qlikview muestre la cantidad de acciones correctivas para dicho segmento).

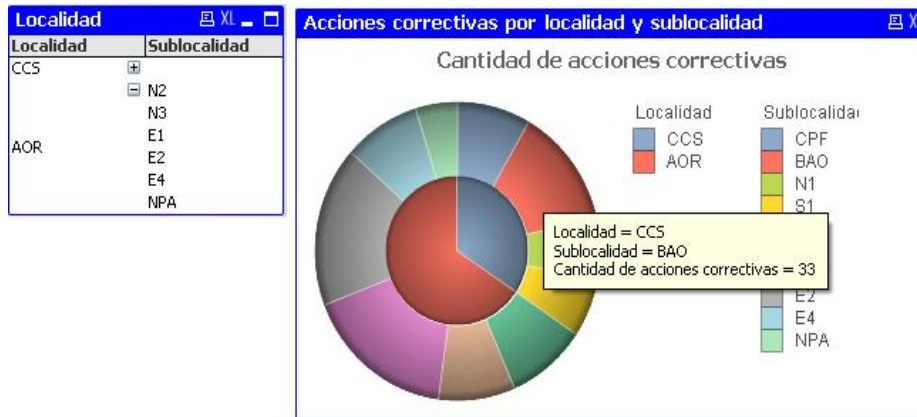


Figura 86 Formas de mostrar los valores y jerarquía de la dimensión sublocalidades.

- **Exportando datos desde qlikview.**

Qlikview permite fácilmente imprimir un elemento del cuadro de mando o exportarlo a Excel, esto se realiza con los íconos que se generan por cada elemento y se ubican en el lado superior derecho del mismo. El primer símbolo permite imprimir el elemento, el segundo exportarlo a Excel, en la figura 87, puede observarse como la tabla resumen que muestra la cantidad de acciones correctivas a partir de todas las dimensiones representadas en el cuadro de mando general puede exportarse y ser visualizada en Excel.

Gerencia	Responsable	Localidad	Prioridad	Estatus	Tipo evento	Fecha creación	Fecha estima de	Fecha de cierre	Total
									253
Explotación	Mora	AOR	alta	Abierto	Incidentes / Accider	Enero-2015	Febrero-2015	-	1
Explotación	Chacin	AOR	alta	Abierto	Opersafe	Marzo-2015	Mayo-2015	-	1
Explotación	Nuñez	AOR	alta	Abierto	Opersafe	Marzo-2015	Abril-2015	-	1
Explotación	Ynfante	AOR	alta	Abierto	Registro Técnico	Enero-2015	Febrero-2015	-	1
Explotación	Ojeda	AOR	alta	Abierto	Registro Técnico	Febrero-2015	Febrero-2015	-	1
Explotación	Acosta	AOR	alta	Abierto	Registro Técnico	Marzo-2015	Abril-2015	-	1
Explotación	Porcelli	AOR	alta	Abierto	Registro Técnico	Marzo-2015	Marzo-2015	-	1
Explotación	Mora	AOR	alta	Cerrado	Opersafe	Noviembre-2014	-	Diciembre-2014	1
Explotación	Ortega	AOR	alta	Cerrado	Opersafe	Diciembre-2014	-	Diciembre-2014	1

Figura 87 Se puede fácilmente exportar a Excel un elemento dentro del cuadro de mando.

El siguiente cuadro de mando permite observar las acciones correctivas tomando como punto principal la manipulación del tiempo. Es este sentido, una acción correctiva puede ser evaluada en el tiempo por su fecha de creación, clausura o estimación de clausura si posee (figura 88). Adicionalmente, se han colocado indicadores referidos al estado de las acciones correctivas, si las mismas se encuentran vencidas entre otros.

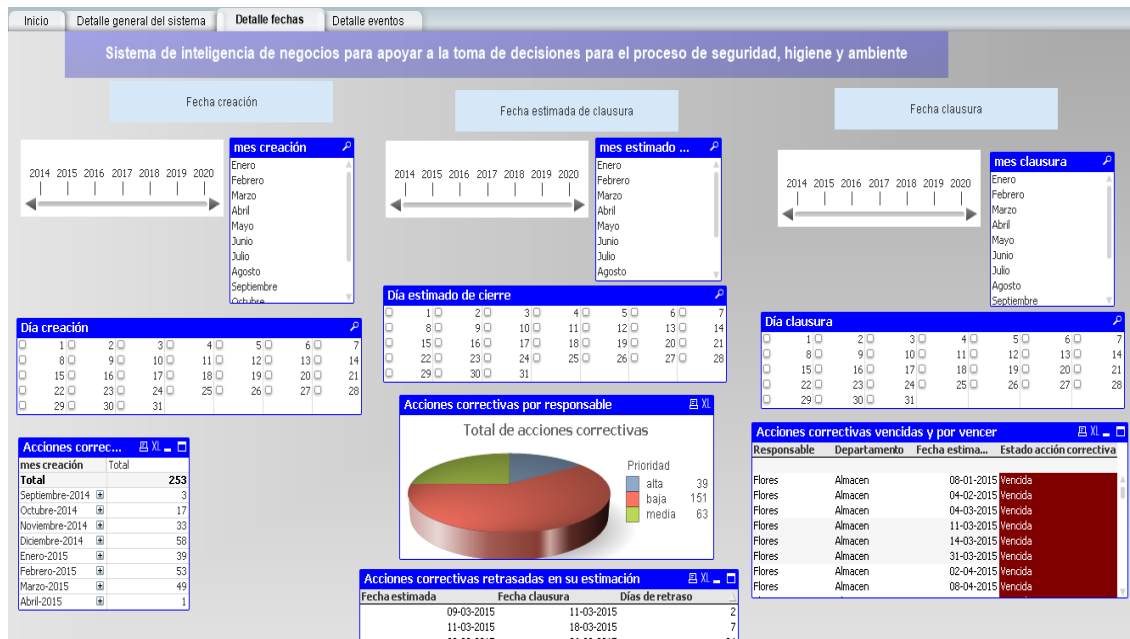


Figura 88 Cuadro de mando enfocado a fechas.

Dentro de Qlikview se presenta dificultades para establecer un calendario típico de formulario de página web, por lo que se realizó un calendario utilizando los elementos que Qlikview provee (figura 89).



Figura 89 Distintas formas de representar un calendario.

A continuación se detallan los indicadores de gestión presentados en este cuadro de mando:

- **Cantidad de acciones correctivas por fecha de creación.**

Para este requerimiento analítico se ha desarrollado un calendario como el presentado en la figura 90. De esta forma se puede manipular el año, mes y día de creación para identificar las acciones correctivas creadas en un periodo de tiempo determinado. Por lo que se pueden seleccionar todas las acciones correctivas creadas en un mes específico durante todos los años o las acciones correctivas creadas en la primera semana de cada mes, etc.

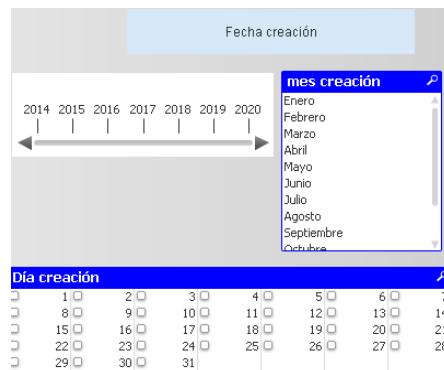


Figura 90 Calendario para gestionar la fecha de creación de las acciones correctivas.

- **Cantidad de acciones correctivas por fecha estimada de clausura.**

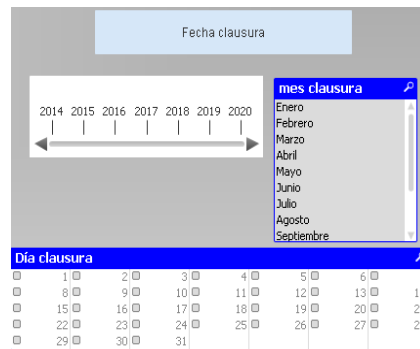
Se realiza un calendario similar al presentado para fecha de creación, con la diferencia que se trabaja con la fecha estimada de cierre. Aunque siempre se posee la misma cantidad de días, el escoger un mes con menos de treinta y un días, inhabilita seleccionar un día que de una fecha errónea, por ejemplo, como se observa en la figura 91, el seleccionar el mes de abril inhabilita seleccionar el día 31.



Figura 91 Qlikview evita crear una selección inválida dentro del calendario.

- **Cantidad de acciones correctivas por fecha clausura.**

Permite gestionar las acciones correctivas según la fecha de clausura a partir de un calendario (figura 92).



**Figura 92** Calendario para la fecha de clausura de acciones correctivas.

Adicionalmente las fechas dentro del cuadro de mando pueden manejarse dentro de la tabla presentada en la figura 93. Como se observa, la selección de un mes de creación despliega los meses de estimación de clausura para las acciones correctivas y para un mes de estimación específico pueden observarse cuál fue el mes de clausura. El valor nulo en esta tabla es representado con el carácter “-” el cual representa para la figura acciones correctivas sin mes estimado.

Acciones correctivas por mes			
mes creación	Mes estimado	Mes clausura	Total
<b>Total</b>			<b>253</b>
Septiembre-2014			3
	Noviembre-2014	Noviembre-2014	8
Octubre-2014	Diciembre-2014	-	3
	-		3
Noviembre-2014			33
Diciembre-2014			58
Enero-2015			39

**Figura 93** Otra forma de manipular las fechas en Qlikview.

- **Cantidad de acciones correctivas por prioridad.**

Un elemento importante dentro de las herramientas de inteligencia de negocios es poder representar un indicador de gestión de distintas formas. Aunque la representación de acciones correctivas por prioridad está presente en el cuadro de mando general, se ha optado por incluir dicho indicador dentro de este cuadro de mando para facilitar la visualización de las acciones correctivas según una fecha determinada y mostrar la capacidad de la herramienta de generar

un indicador de múltiples formas. En este caso, como se muestra en la figura 94, se muestran las acciones correctivas por prioridad a partir de un gráfico de torta.

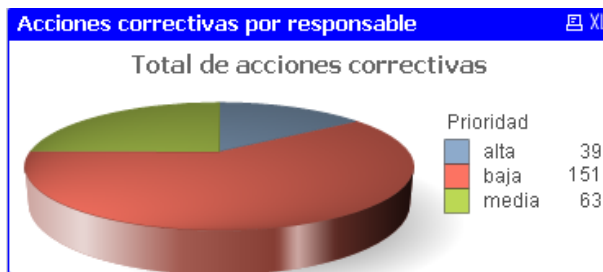


Figura 94 Cantidad de acciones correctivas por prioridad mostrado en un gráfico de torta.

- **Cantidad de acciones correctivas vencidas y por vencer.**

Un aspecto a considerar dentro de las acciones correctivas, es establecer cuales están por vencer o están vencidas. Una acción correctiva por vencer es aquella que dentro del próximo mes ocurrirá su fecha estimada de clausura. Una acción correctiva vencida es aquella que posee estado abierto y ya paso su fecha estimada de clausura.

Como se observa en la figura 96, al elemento adicionalmente se le ha colocado el responsable el departamento, fecha estimada de clausura y el estado “por vencer” o “vencida”). Dentro de Qlikview se puede resaltar un elemento dentro de una tabla a partir de los colores, como se muestra en la figura 95 donde las acciones correctivas vencidas presentan un fondo rojo.

Responsable	Departamento	Fecha estima...	Estado acción correctiva
Porcelli	Pozos	28-03-2015	Vencida
Porcelli	Proyecto	14-04-2015	Vencida
Porcelli	Planta	16-04-2015	Vencida
Porcelli	Planta	03-05-2015	Por vencer
Porcelli	Planta	06-11-2014	Vencida
Acosta	Pozos	02-04-2015	Vencida
Luces	Pozos	01-05-2015	Vencida
Diaz	Almacen	03-02-2015	Vencida

Figura 95 Acciones correctivas vencidas y por vencer.

- **Cantidad de acciones correctivas retrasadas en su estimación.**

Permite saber la cantidad de días que ocurrió de un retraso en el cierre de una acción correctiva. Un retraso es la diferencia entre la fecha estimada de clausura y la fecha real de clausura (figura 96).

Fecha estimada	Fecha clausura	Días de retraso
09-03-2015	11-03-2015	2
11-03-2015	18-03-2015	7
02-03-2015	26-03-2015	24

Figura 96 Cantidad de acciones correctivas retrasadas.

Por último se tiene un cuadro de mando asociado a los eventos. Los eventos dentro del sistema SHA cobran vital importancia porque reflejan que tipo de suceso ocurrió en la organización para que se generaran el conjunto de acciones correctivas en busca de solventar o mejorar la organización en base a ese suceso

El cuadro de mando enfocado a los eventos se observa en la figura 97. Y se describe a continuación sus indicadores de gestión..

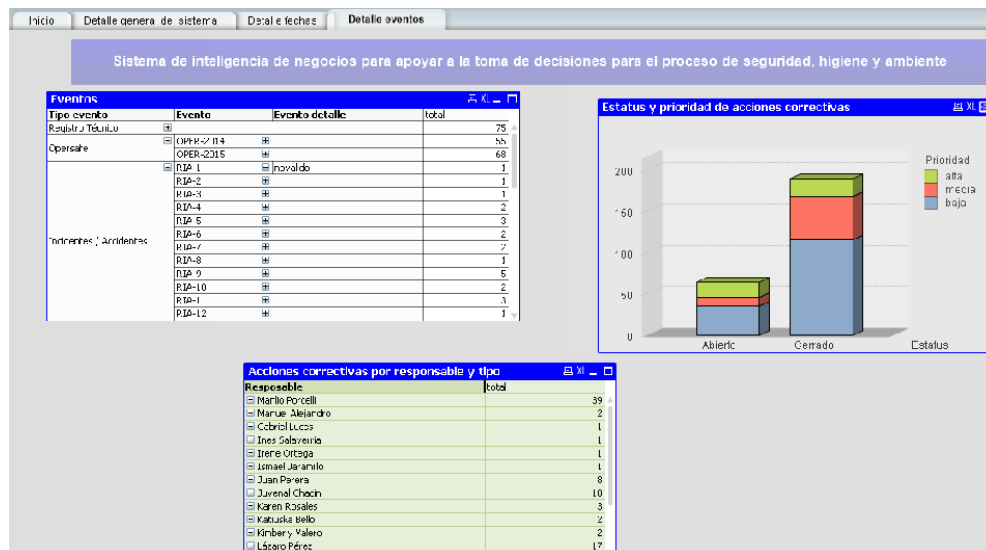


Figura 97 Cuadro de mando enfocado a los eventos.

- Cantidad de acciones correctivas por evento.

En este caso el eje del cuadro de mando es evaluar las acciones correctivas en función de los distintos eventos, por lo que se disponen de dos formas para evaluar las acciones correctivas en función de esta dimensión.

En primera forma, se posee un elemento que permite a partir del tipo de evento (figura 98), realizar operaciones de roll up y drill down sobre la jerarquía de la dimensión (tipo evento-evento-evento detalle).

Eventos			
Tipo evento	Evento	Evento detalle	total
Registro Técnico			75
Operasafe	OPER-2014		55
	OPER-2015		68
Incidentes / Accidentes			55

Figura 98 Forma de navegar por la dimensión eventos a partir de la jerarquía.

En segunda forma, se puede ver por responsable el total de acciones correctivas que tiene en función de los tipos de evento (figura 99).

Responsable	Tipo evento	Total	Incidentes / Accidentes	Opersafe	Registro Técnico
Manlio Porcelli		39	12		20
Manuel Alejandro		2	-		2
Gabriel Luces		1	1		-
Ines Salaverria		1	-		1
Irene Ortega		1	-		1
Ismael Jaramillo		1	-		1
Juan Perera		8	-		8
Juvenal Chacin		18	1		16
Karen Rosales		3	-		3
Katuska Bello		2	1		1
Kimberly Valero		2	-		2

Figura 99 Acciones correctivas por responsable y tipo.

Qlikview adicionalmente permite modificar las columnas por filas (pivot) para cambiar la presentación de un indicador. En este caso el tipo de evento donde cada columna representa un tipo de evento puede pasar a ser filas por cada tipo de responsable (figura 100).

Responsable	total
Kimberly Valero	2
Lázaro Pérez	17
Miguel Díaz	1
Oli Nuñez	33
Incidentes / Accidentes	12
Opersafe	12
Registro Técnico	9
Oswaldo Mora	12
Oswaldo Paris	1
Otto Garcia	1
Pablo Barceló	5

Figura 100 Acciones correctivas por responsable y tipo de evento aplicando

- **Cantidad de acciones correctivas por estatus y prioridades.**

Qlikview permite generar indicadores de gestión a partir de la incorporación de múltiples dimensiones dentro de un gráfico, en este caso, se puede generar un gráfico de barra utilizando las dimensiones de estatus y prioridades (figura 101).

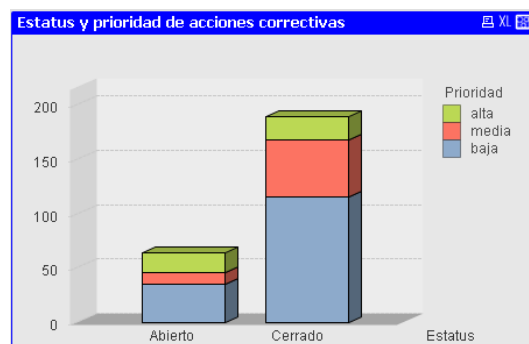


Figura 101 Cantidad de acciones correctivas por estatus y prioridad.

#### 4.3.4. Verificar la calidad de los datos en la herramienta de manipulación y visualización de datos.

Una vez generado los cuadros de mando, se procedió a realizar la verificación de los datos. En este caso fueron analizadas las distintas dimensiones usando los criterios planteados en el punto 4.2.5, lo cual implica la realización de consultas sobre la base de datos relacional ajustando los indicadores de gestión a partir de los filtros planteados en el cuadro de mando para comprobar su igualdad.

A continuación se presentan el conjunto de consultas realizadas para la verificación de los datos:

- **Cantidad de acciones correctivas por prioridad.**

Como se observa en la figura 102, la consulta al sistema transaccional muestra el mismo resultado que el gráfico dentro del cuadro de mando.

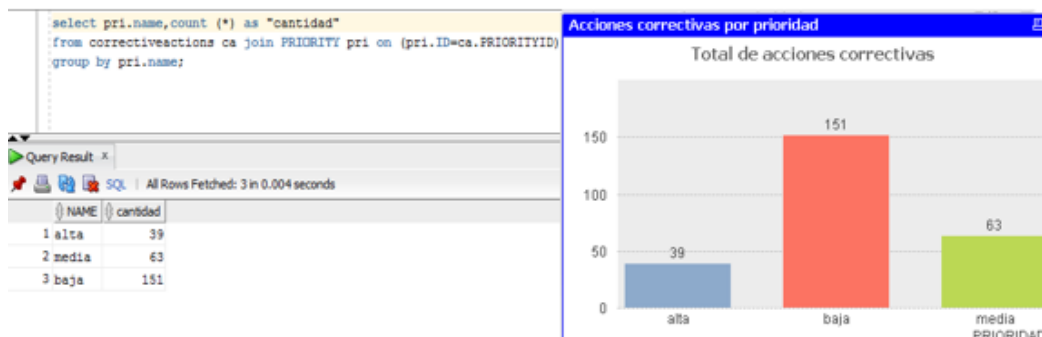


Figura 102 Verificación de datos a partir del cálculo de acciones correctivas por prioridad.

- **Cantidad de acciones correctivas seleccionando una sublocalidad.**

En este caso se evalúa la selección de una localidad específica, lo que permite evaluar también la integridad de la jerarquía en la dimensión sublocalidades (figura 103).

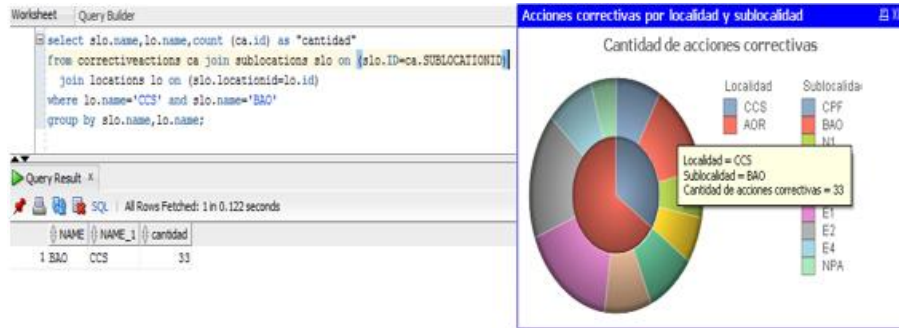


Figura 103 Verificación de los datos realizando una selección en la dimensión sublocalidades.

- Acciones correctivas retrasadas en su estimación.

Verificación de la consulta que permite identificar las acciones correctivas que fueron cerradas días después de la estimación de su clausura (Figura 104).

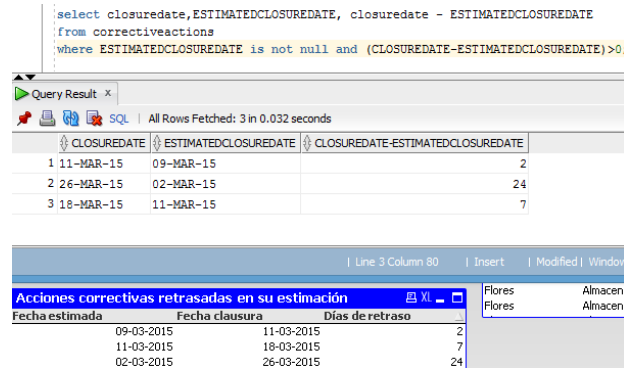


Figura 104 Verificación de concordancia en el retraso de acciones correctivas.

- Cantidad de acciones correctivas para un evento.

Verificación de datos a partir de la especificación de un tipo de evento, en este caso tipo Opersafe figura 105.

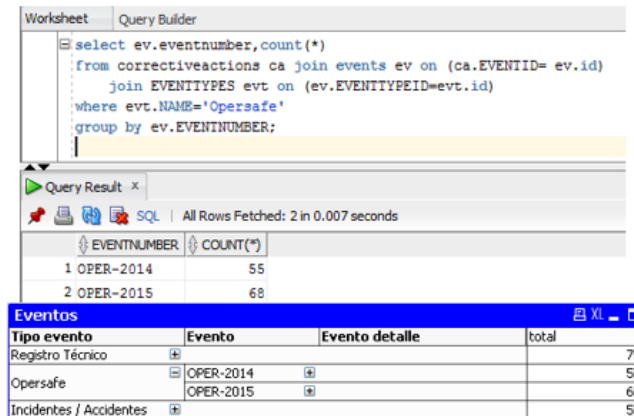


Figura 105 Verificación de cantidad de acciones correctivas para el evento Opersafe.

- **Responsables que crearon acciones correctivas en un momento determinado.**

Se puede a su vez realizar una selección para realizar verificación de datos dentro de distintos cuadros de mando (figura 106).



**Figura 106 Verificación de datos seleccionando un mes y visualizando sus responsables.**

#### **4.4.- Implementación de la solución.**

Debido a que el trabajo presentado forma parte de una solución de inteligencia de negocios para un trabajo especial de grado, su implementación dentro de la organización para su utilización en las operaciones de la empresa no fue realizada. Esto genera que dicha fase de desarrollo de la solución no sea ejecutada de forma total.

##### **4.4.1.- Ajuste de las consultas.**

Para el ajuste de consultas se realizaron pruebas de visualización de los indicadores con integrantes del grupo de tecnología de información de la organización para comprobar la viabilidad en la forma como se estaban desplegando los indicadores de gestión.

De estas reuniones con el grupo de tecnología se realizaron y distribuyeron las consultas en el cuadro de mando buscando que los elementos destacados fueran visibles, todos los elementos del cuadro de mando tuvieran concordancia, Se pudiera navegar de forma natural por las jerarquías, se cumplieren con los requerimientos planteados y su disposición fuera armónica.

Para esto se desarrollaron prototipos de interfaces de cuadros de mando, en donde se presenta en este documento la versión final aceptada por el equipo de tecnología de información.

La fase de ajuste de consulta consistió en dos ciclos de presentación a cinco usuarios que conforman parte del grupo de tecnología de la información. En el primer ciclo, como se observa en la tabla 14, el grupo de pruebas estableció algunas fallas al prototipo inicial.

Usuario	Detalle general del sistema	Detalle tiempo	Detalle evento
Usuario 1	Correcto con detalles	Correcto	Correcto
Usuario 2	Correcto	Correcto con detalles	Correcto
Usuario 3	Correcto	Correcto	Correcto
Usuario 4	Correcto	Correcto con detalles	Correcto
Usuario 5	Correcto	Correcto	Correcto

**Tabla 14 Resumen de evaluación de primer prototipo.**

A continuación se describen las observaciones realizadas en el primer ciclo de entrega:

- El usuario 1 planteó que la tabla dentro del cuadro de mando de detalle general del sistema debe poseer representación de todas las dimensiones. Como solución se agregaron todas las dimensiones
- El usuario 2 estableció que debería de existir una forma más intuitiva para observar las acciones correctivas creadas en un mes y a partir de este mes ver los meses donde cerradas dichas acciones correctivas. La solución fue la creación de un nuevo indicador de gestión, observado en la figura 93.
- Dentro del detalle del tiempo, gráficamente la tabla que mostrabas las acciones correctivas vencidas no se observada de forma destacada las acciones que estaban vencidas o por vencer, para esto como solución se agregó color para identificar de forma clara el estado de las acciones correctivas.
- Se estableció arreglar los colores de fondo y títulos para que los cuadros de mando tuviesen una mejor presentación
- Incorporar un cuadro de mando inicio donde se listen las dimensiones (el máximo nivel de jerarquía de las mismas) y se pueda observar las acciones correctivas según la selección de una dimensión.

En función de las observaciones se realizó un nuevo ciclo de desarrollo donde se generó el prototipo final de la solución de inteligencia de negocio, la tabla 15 muestra la satisfacción de los usuarios utilizados para evaluar el producto.

Usuario	Inicio	Detalle general del sistema	Detalle tiempo	Detalle evento
Usuario 1	Correcto	Correcto	Correcto	Correcto
Usuario 2	Correcto	Correcto	Correcto	Correcto
Usuario 3	Correcto	Correcto	Correcto	Correcto
Usuario 4	Correcto	Correcto	Correcto	Correcto
Usuario 5	Correcto	Correcto	Correcto	Correcto

**Tabla 15 Resumen de evaluación de primer prototipo.**

#### **4.4.2.- Establecer esquemas de seguridad.**

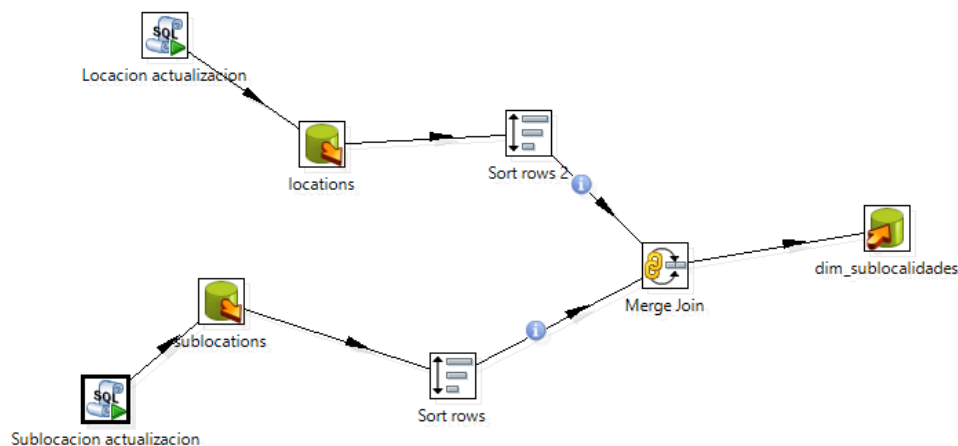
En el caso del establecimiento de esquemas de seguridad, es requerido utilizar QlikServer que permite configurar usuarios y permisos dentro de la solución o a partir de QlikServer, utilizar un frame, es decir, un despliegue del cuadro de mando dentro de una aplicación web y desde la aplicación web establecer los esquemas de seguridad. Debido a que QlikServer es una aplicación paga sin licencia comunitaria y la solución generada en este trabajo especial de grado tiene carácter experimental, se utilizó solo la herramienta Qlikview desktop la cual no provee seguridad. El establecimiento de un esquema de seguridad bajo este escenario depende del acceso a la base de datos del modelo dimensional, es decir se trabaja bajo el esquema de seguridad del sistema manejador de base de datos Oracle. Esta conexión es requerida para poder cargar los datos y recargar el script, lo que actualiza los datos en el cuadro de mando. Por tanto es la forma que se puede garantizar la restricción a los datos, debido a que el archivo generado por Qlikview, mantiene el formato de los cuadros de mando, pero no los datos, es decir, exportar la solución origina que en el sitio de destino se tenga que cargar los datos a partir de un usuario que tenga acceso a los mismos dentro de la base de datos.

#### 4.4.3.-Crear flujos de procesos automáticos.

A partir de la consulta con el departamento de seguridad, higiene y ambiente se optó por desarrollar un proceso que actualizase los datos dentro del almacén de datos cada semana o 7 días. Para esto se han generado un conjunto de procesos ETC adicionales y trabajos dentro de pentaho Kettle.

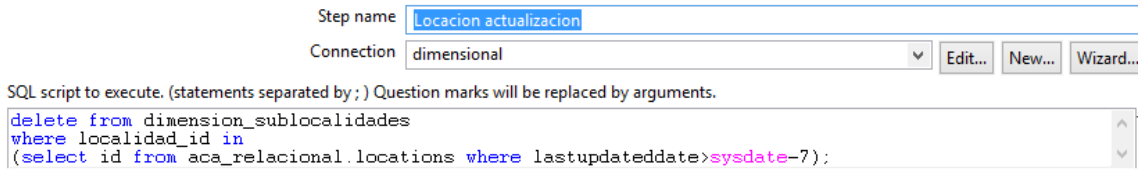
Debido a las características de las dimensiones, en donde los registros no suelen recibir modificaciones y la cantidad de valores en la mayoría de las dimensiones no supera los diez valores, el proceso de actualización de los datos consiste en eliminar los elementos modificados de la implementación del modelo dimensional y luego insertar los elementos nuevos o modificados. Por último se eliminan las acciones correctivas en donde alguna de sus dimensiones ha sido modificada y se insertan las acciones correctivas nuevas y modificadas.

Para ejemplificar dicho flujo, se muestra el proceso ETC de la dimensión sublocalidad, como se observa en la figura 107.



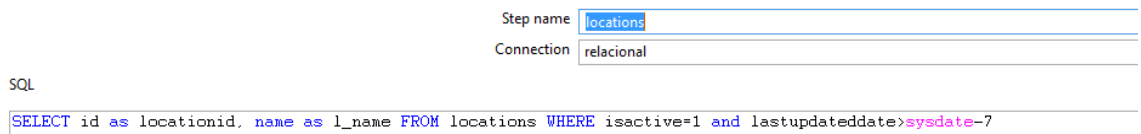
**Figura 107 Proceso ETC periódico de la dimensión sublocalides.**

Al proceso ETC de la dimensión sublocalidad en la carga inicial se le han añadido dos scripts, que verifican y eliminan los datos actualizados. Dichos scripts en todas las dimensiones son similares y como se observa en la figura 108 se revisa cuales datos fueron actualizados para eliminarlos de la implementación.



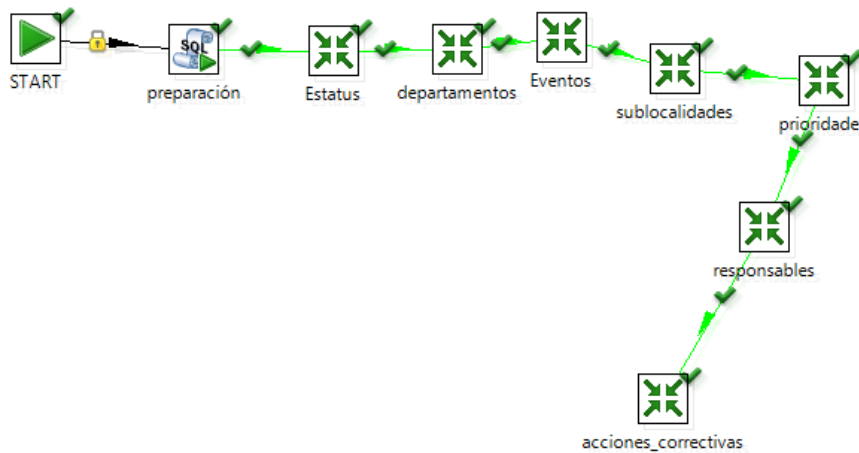
**Figura 108 Script que verifica y elimina datos actualizados.**

Luego solo son agregados los elementos recientes (menos de 7 días de su última actualización) figura 109.



**Figura 109 Inserción de elementos nuevos.**

Por último, en el job de kettle (figura 110), se incorporan los ETC periódicos, el tiempo es estático por lo que no se hace falta crear un proceso ETC. Se agrega un proceso de preparación que elimina las acciones correctivas modificadas y luego son cargadas las acciones correctivas nuevas y actualizadas en el proceso ETC de la tabla de hechos respectivamente.



**Figura 110 Job de actualización periódica de la implementación del modelo dimensional.**

## **CONCLUSIONES Y RECOMENDACIONES**

Se ha logrado exitosamente el objetivo de este trabajo especial de grado, el cual era el desarrollo de una solución de inteligencia de negocios para el área de seguridad, higiene y ambiente (SHA) para una empresa dedicada a la extracción de gas.

La construcción de la solución de negocios implica la realización de un modelo dimensional que permitiese evaluar las ocurrencias de las acciones correctivas a partir de distintas perspectivas del negocio.

A partir de la implementación de un área intermedia y el modelo dimensional previamente definido, se utilizaron procesos de extracción transformación y carga de forma que se pudiera obtener los datos del sistema transaccional para ser utilizada en un entorno analítico.

La utilización de la metodología de mejores prácticas para la construcción de la solución de inteligencia de negocios facilita el desarrollo de la misma al establecer en qué punto es idóneo el concretar los indicadores de gestión a partir de la forma en cómo se obtienen en este caso los requerimientos además de guiar en el proceso de la generación de los indicadores de gestión en la herramienta de manipulación de datos.

La flexibilidad a la hora de realizar informes o visualizar indicadores, en contraposición a la solución del departamento SHA usando gráficos predefinidos permite el análisis de datos de forma eficiente puesto que la realización de nuevas consultas, gráficos a partir de necesidades presentadas pueden ser atendidas rápidamente. Esta rapidez se traduce en que el sistema es capaz de resolver necesidades analíticas sin necesidad de que el usuario final recurra a programadores que tengan que generar a partir de la programación, un nuevo elemento para la obtención de información.

Es importante realizar procesos de verificación de la calidad de los datos para asegurar que los resultados mostrados son correctos. La realización de procesos de ETC y realización de consultas dentro de la herramienta requiere que se compruebe dicha calidad para asegurar consistencia en la solución que se está generando.

Debido a que el trabajo de investigación se realiza en un ambiente de pruebas y no en un ambiente de producción, las tareas involucradas a la cuarta fase planteada en la metodología de mejores prácticas se realizaron parcialmente.

Como recomendaciones generales tenemos:

- Se plantea realizar un proceso de adiestramiento a los usuarios, con el objetivo de ajustar las consultas en función a los usuarios de la organización e implementar efectivamente la solución en la organización. Adicionalmente utilizar QlikServer para establecer esquemas de seguridad. De esta forma se cumple de forma completa con la cuarta fase de la metodología mejores prácticas.
- Una de las ventajas de la utilización de las herramientas Qlik es la capacidad de que el cuadro de mando se ajuste a la resolución de la pantalla, lo que facilita la visualización de los datos en distintos dispositivos. Para la utilización de esta capacidad es necesaria adquirir la herramienta QlikServer, que es la que realiza el proceso de adaptación del cuadro de mando. Se aconseja adquirir a la organización dicho software para el aumento de las funcionalidades de la solución.
- El sistema maneja un conjunto definido de eventos, que forma parte de los eventos que se gestionan en el sistema transaccional, sin embargo nuevas versiones del sistema transaccional aun no puesto en producción generaran nuevos tipos de eventos (eventos del tipo auditoria, parada de planta, shutdown entre otro). Para mantener la vigencia y valor de la solución de inteligencia de negocios deben incorporarse estos nuevos eventos que puedan surgir, entendiendo que la generación de un nuevo tipo de evento implica la realización de un proceso de extracción, transformación y carga.
- La solución de inteligencia de negocios realizada se basa en el núcleo del sistema transaccional del departamento SHA, las acciones correctivas. Se plantea la generación de soluciones de inteligencia de negocios parciales realizando modelos dimensionales para los distintos tipos de eventos y conectando todos estos modelos obteniendo un esquema constelación para abarcar todos los componentes del departamento SHA, lo que implicaría tener una solución más robusta.

## **BIBLIOGRAFÍA**

Almeida, M. Ishikawy, y M.,Reinschmidt J. (1999). *Getting Started with DataWarehouse and Business Intelligence* (1ª ed.). Estados unidos de América: International Technical Support Organization.

Cano, J. L. (2007). *Business Intelligence: Competir con información. España:ESADE.*

Chiavetano, I. (2000). *Administración de Recursos Humanos. Quinta Edición. Colombia, McGRAW-HILL INTERAMERICANA, S. A.*

Gartner (2014). Gartner Magic Quadrant .Recuperado el 15 de enero de 2015 de: [http://www.gartner.com/technology/research/methodologies/research\\_mq.jsp](http://www.gartner.com/technology/research/methodologies/research_mq.jsp).

Peréz,G (2011). *Guía de estudio: Seguridad, higiene y ambiente (1ª ed.)*. Unefa.

IBM (s.f). IBM Bussines Intelligence. Recuperado el 5 de enero de 2015 de: <http://www-03.ibm.com/software/products/en/category/business-intelligence>

Inmon, I. (1992). *Building the Operational Data Store* (1ª ed.). Estados unidos de América: John Wiley and Sons.

Kimball, R. (1998). *The datawarehouse lifecycle toolkit* (1ª ed.). Estados unidos de América:Wiley.

Kimball, R., y Ross, M. (2013). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling* (4ª ed.). Estados unidos de América: Wiley Computer Publishing.

Laudon, K. y Laudon J. (2008). *Sistemas de información gerencial: Administración de la empresa digital* (10ª ed.). México: PEARSON Education.

Lunh, P. (1959). *A Business Intelligence System*. Estados unidos de América: Frederick A. Praeger.

Microsoft. (s.f.). *Microsof Power BI*. Recuperado el 28 de enero de 2015 de: <http://www.microsoft.com/en-us/powerbi/home/power-bi.aspx>.

Microsoft. (s.f). *Business Intelligence in Office and SQL Server* .Recuperado el 28 de enero de 2015 de: <http://www.microsoft.com/en-us/server-cloud/solutions/business-intelligence/>.

Molina, H, (2008). *Database system* (2ª ed.). Estados unidos de América: Pearson.

Pentaho (s.f). *Pentaho*. Recuperado el 27 de enero de 2015 de: <http://www.pentaho.com/>.

Pérez Jaramillo, Carlos Mario (2013). Los indicadores de gestión. Recuperado el 7 de marzo de 2015 de:  
<http://www.escuelagobierno.org/inputs/los%20indicadores%20de%20gestion.pdf>.

Ponniah. (2001). *Data Warehousing Fundamentals*. Estados unidos de América: John Wiley & Sons, Inc.

Qlik. (s.f). Qlik Products. Recuperado el 27 de enero de 2015 de:  
<http://www.qlik.com/us/explore/products/>

Rozo. Fabio (2013). *Indicadores de gestión para la toma de decisiones basada en Inteligencia de Negocios*. Recuperado el 5 de marzo de 2015 de:  
<http://revistas.udistrital.edu.co/ojs/index.php/tia/article/view/4639/7094>.

Seguridad Higiene Ambiental SHA (s.f). *Sensibilización a la seguridad, higiene y ambiente*. Recuperado el 11 de marzo de 2015 de:  
<http://seguridadhigieneambiental.blogspot.com/2012/12/modulo-i-sensibilizacion-la-seguridad.html>.

Smi (1999). *Data Warehouses and the Telecommunications Industry*. Recuperado el 20 de diciembre de 2014 de: [www.cs.colorado.edu](http://www.cs.colorado.edu).

Tableau.(s.f).*Tableau Products*. Recuperado el 25 de enero de 2015 de:  
<http://www.tableau.com/es-es/products/>.

TIAN CONSULTORES. (2010). “Manual Metodología de Desarrollo de Sistemas de Información de Inteligencia de Negocio”.Caracas, Venezuela.

Universidad del Zulia. (s.f).*Departamento de Higiene y Seguridad Industrial* .Recuperado el 10 de marzo de 2015 de:

[http://www.dsi.luz.edu.ve/index.php?option=com\\_content&task=view&id=111&Itemid=19](http://www.dsi.luz.edu.ve/index.php?option=com_content&task=view&id=111&Itemid=19)  
1.

Vasiliev, Y. (2010). *Oracle Business Intelligence: The Condensed Guide to Analysis and Reporting*. United Kingdom: PACKT Editorial.