



UNIVERSIDAD CENTRAL DE VENEZUELA

FACULTAD DE CIENCIAS

ESCUELA DE COMPUTACIÓN

SISTEMAS DE INFORMACIÓN

**Implementación de Indicadores de Gestión
para monitorear el rendimiento del proceso
de Revisión Editorial, utilizando una
plataforma de Inteligencia de Negocios,
desarrollada en Software Libre**

Trabajo Especial de Grado

Presentado ante la Ilustre

Universidad Central de Venezuela

por el bachiller

Fabián Eduardo Moreno Contreras

(C. I.: V.- 18.031.166)

Para optar al título de:

Licenciado en Computación

Tutor:

Prof. Franky Uzcátegui

Caracas, Mayo de 2012

Acta

Quienes suscriben, miembros del jurado designado por el Consejo de la Escuela de Computación de la Facultad de Ciencias de la Universidad Central de Venezuela, para examinar el Trabajo Especial de Grado titulado: **Implementación de Indicadores de Gestión para monitorear el rendimiento del proceso de Revisión Editorial, utilizando una plataforma de Inteligencia de Negocios, desarrollada en Software Libre**, presentado por el bachiller Fabián E. Moreno C., portador de la cédula de identidad V.- 18.031.166, a los fines de optar por el título de Licenciado en Computación, dejan constancia de lo siguiente:

Dicho trabajo, leído por cada uno de los miembros del jurado, se fijó el día 25 de mayo de 2012, a las 9:00 am, para que su autor lo defendiera en forma pública en la Escuela de Computación, mediante una presentación oral de su contenido, luego de lo cual se respondieron las preguntas formuladas. Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió aprobarlo con la nota de _____ puntos.

En fe de lo cual se levanta la presente Acta, en Caracas a los veinticinco (25) días del mes de mayo del año dos mil doce (2012), dejando constancia de que actuó como coordinador del jurado el Profesor Franky Uzcátegui.

Prof. Franky Uzcátegui
(Tutor)

Prof. Concettina Di Vasta
(Jurado)

Prof. Rhadames Carmona
(Jurado)

Dedicatoria

A Dios Todopoderoso,
por hacerme un hombre inteligente, dedicado y luchador.

A mi madre,
por la educación impartida y por ser ejemplo de mujer luchadora.

A mi hermana,
por estar siempre a mi lado y combatir conmigo.

A mi abuela Carlota,
que desde el cielo sé que está orgullosa de este triunfo.

A toda mi familia,
por su cariño, afecto y por contar siempre con su apoyo.

Fabían Moreno

Agradecimientos

Primeramente, quiero agradecer a Dios Todopoderoso, por convertirme en el hombre que soy hoy, por siempre escuchar mis oraciones y ser el único testigo de mis sueños. Gracias por ayudarme a cumplir cada uno de ellos. Sigue iluminando mi camino y mis pasos.

A mi madre, mujer que no se cansa de luchar. A ella agradezco todo lo que soy y le regalo este triunfo. Gracias por el amor, las caricias y la paciencia que me tienes. Eres la base de mi vida y sin tu apoyo este logro no sería posible. Te amo, eres la mejor.

A mi hermana, mi segunda mitad, quien incansablemente me apoya y me ayuda a luchar contra todo obstáculo que se me presente. Gracias por tu amor, por cuidarme y por tus sabias palabras que siempre las escucho. Te amo.

A mi padre, por estar siempre pendiente y querer el bien para mí. Por su paciencia y cariño. Te quiero mucho.

A mi tía Yoly, por ser mi segunda madre. Porque a pesar de la distancia, siempre está pendiente de mí y es cómplice de mis pasos hacia el éxito. Gracias por el amor y la paciencia. Te amo.

A mi abuela Carlota, quien desde el cielo sé que me está cuidando y celebrando este triunfo conmigo. Sé que estás orgullosa de mí y tienes esa gran sonrisa tan bella que te caracterizaba.

A mi abuela Carmen y mi abuelo Canon, por su amor incondicional, por sus caricias y por las sonrisas que me sacan en el rostro cada vez que los veo. Gracias por estar pendientes de mí.

A mis tías Magaly, Beatriz y Tahiz por su apoyo en cualquier momento. Por estar siempre dispuestas a ayudarme y por ser en mi vida un ejemplo a seguir. Gracias por todo el amor que me tienen y sé que soy como un hijo para ustedes. Las amo.

A mis tíos y tías por ser un ejemplo de unión y fuerza. A los Moreno, por todas esas risas que juntos compartimos y que no nos cansamos de estar juntos. A los Contreras, por todo el amor y el cariño con el que me reciben cada año. Los amo.

A Jhoanna, por ser como una hermana. Por apoyarme y darme ánimos para no decaer nunca. Por los momentos inolvidables que hemos vivido y por ser un ejemplo a seguir.

A todos mis primos, por sus sonrisas en todo momento y el apoyo incondicional que me prestan. Espero que a los más pequeños, este éxito les sirva de estímulo.

A mis amigos de la UCV, Angélica, Rosa, Mery, Marián, Marieva, Augusto, Loy, Karina y los que faltan, gracias por compartir conmigo esta bella experiencia de estudiar en “La Casa que vence las sombras”. Gracias por lo feliz que me hicieron vivir durante estos años.

A ti, que viviste conmigo cada segundo de lo que fue la recta final de este sueño. Gracias por estar conmigo en las alegrías y en las lágrimas, por ser un factor fundamental de este logro y por jamás dudar de lo que puedo llegar a ser. Te amo.

A todos mis amigos de la vida, del teatro y de la televisión, gracias por los momentos que compartí con ustedes. Gracias por las sonrisas que me sacan día a día y por representar un factor indispensable en mi vida.

Gracias a la grandiosa e ilustre Universidad Central de Venezuela, por ser testigo de mis angustias, de mis lagrimas, de mis risas, de mis secretos y de todo lo que pude vivir en ella. Gracias por acobijarme en tus brazos y hacerme parte de ti.

Gracias a mi tutor Franky por la confianza depositada en mí. Por darme la oportunidad de realizar este proyecto a pesar de las circunstancias. Gracias por todo el apoyo incondicional, por orientarme a lo largo de este camino y por enseñarme a ser un mejor profesional.

A aquellas personas que no están físicamente, sé que desde el cielo me están cuidando y guiándome. Hoy el cielo se viste de lujo.

Por último, gracias a aquellos que quisieron obstaculizar mi camino, porque me enseñaron a ser más fuerte y jamás decaer. Hacen falta mil como ustedes para intentar detenerme.

¡GRACIAS TOTALES!

Resumen

El presente Trabajo Especial de Grado tiene como objetivo implementar indicadores de gestión que permitan monitorear el rendimiento del proceso de Revisión Editorial de un sistema de publicación, a través de una plataforma de Inteligencia de Negocios, desarrollada bajo Software Libre, denominada *Pentaho Business Intelligence*. Esta solución aprovecha las ventajas que ofrece la plataforma de Inteligencia de Negocios, para así, dar soporte a la generación y monitoreo de la revisión de artículos y apoyar en la toma de decisiones, por medio de información fundamental. Para lograr el objetivo se utilizó el enfoque de Tian Consultores, que es una adaptación de los métodos: ascendente, propuesto por Ralph Kimball y descendente, presentado por Bill Inmon; y que además, sintetiza y procedimenta sus mejores prácticas. La unión del uso del enfoque metodológico y la plataforma tecnológica de *Pentaho Business Intelligence*, ofrece la herramienta descrita.

Palabras Claves:

Sistemas de Información, Inteligencia de Negocios, Proceso de Revisión Editorial, Pentaho Business Intelligence, Indicadores de Gestión.

Índice General

INTRODUCCIÓN	1
CAPÍTULO I. PROBLEMA DE LA INVESTIGACIÓN	3
1.1 TÍTULO.....	3
1.2 SITUACIÓN ACTUAL.....	3
1.3 PLANTEAMIENTO DEL PROBLEMA	4
1.4 SOLUCIÓN PROPUESTA	6
1.5 OBJETIVO GENERAL	7
1.6 OBJETIVOS ESPECÍFICOS.....	8
1.7 ALCANCE	8
1.8 IMPORTANCIA Y JUSTIFICACIÓN	9
CAPÍTULO II. MARCO CONCEPTUAL	10
2.1 SISTEMAS DE INFORMACIÓN	11
2.1.1 <i>Definición</i>	11
2.1.2 <i>Tipos de Sistemas de Información</i>	12
2.2 BASE DE DATOS.....	14
2.2.1 <i>Definición</i>	14
2.2.2 <i>Modelo de Base de Datos</i>	14
2.2.3 <i>Tipos de Base de Datos según su propósito</i>	17
2.2.4 <i>Sistemas Manejadores de Base de Datos</i>	18
2.3 ALMACÉN DE DATOS Y BODEGA DE DATOS	19
2.3.1 <i>Almacén de Datos (Data Warehouse)</i>	19
2.3.2 <i>Bodega de Datos (Data Mart)</i>	22
2.3.3 <i>Extracción, Transformación y Carga (ETL)</i>	23
2.3.4 <i>Modelo Dimensional</i>	24
2.4 INTELIGENCIA DE NEGOCIOS	31
2.4.1 <i>Definición</i>	31
2.4.2 <i>Características</i>	32
2.4.3 <i>Arquitectura</i>	33

2.4.4 Pirámide de Información	34
2.4.5 Componentes.....	35
2.4.6 Ventajas.....	38
2.5 INDICADORES DE GESTIÓN	39
2.5.1 Atributos.....	40
2.5.2 Componentes.....	40
2.5.3 Clasificación.....	41
2.5.4 Ventajas.....	42
2.5.5 Metodología General.....	43
2.6 PLATAFORMAS DE INTELIGENCIA DE NEGOCIOS SOFTWARE LIBRE	45
2.6.1 Plataforma Pentaho Business Intelligence	46
2.7 MÉTODO DE DESARROLLO DE APLICACIONES.....	55
2.7.1 Levantamiento de la Información e Indicadores.....	56
2.7.2 Modelamiento Dimensional	56
2.7.3 Desarrollo del Almacén de Datos	56
2.7.4 Análisis de Consistencia de Datos del Almacén de Datos.....	57
2.7.5 Desarrollo de Indicadores.....	57
2.7.6 Análisis de Consistencia de los Cuadros de Mando	57
2.8 PROCESO DE REVISIÓN EDITORIAL	57
CAPÍTULO III. MARCO APLICATIVO	61
3.1 CONTEXTO DE DESARROLLO	61
3.2 PROYECTO.....	61
3.3 FASES DEL MÉTODO DE DESARROLLO.....	62
3.3.1 Fase1: Levantamiento de la Información e Indicadores.....	62
3.3.2 Fase 2: Modelamiento Dimensional	67
3.3.3 Fase 3: Desarrollo del Almacén de Datos	74
3.3.4 Fase 4: Análisis de Consistencia de Datos del Almacén de Datos	84
3.3.5 Fase 5: Desarrollo de Indicadores.....	90
3.3.6 Fase 6: Análisis de Consistencia de los Cuadros de Mando.....	103
CONCLUSIONES.....	110

REFERENCIAS BIBLIOGRÁFICAS..... 112

Índice de Figuras

Figura 1. Arquitectura de la solución planteada	6
Figura 2. Sistemas de Información.....	12
Figura 3. Tipos de Sistemas de Información	12
Figura 4. Modelo Jerárquico	15
Figura 5. Modelo Red	15
Figura 6. Modelo Relacional.....	16
Figura 7. Modelo Orientado a Objetos	17
Figura 8. Arquitectura de un Almacén de Datos.....	20
Figura 9. Orientación al tema.....	21
Figura 10. Variación en el tiempo	21
Figura 11. Almacenes no volátiles.....	22
Figura 12. Integridad del almacén.....	22
Figura 13. Cubo Representativo de Dimensiones	27
Figura 14. Tabla de Hechos Aditivos	28
Figura 15. Tabla de Hechos Semi-aditivos	29
Figura 16. Tabla de Hechos No aditivos	29
Figura 17. Tabla de Hechos y Dimensiones.....	29
Figura 18. Esquema Estrella	30
Figura 19. Esquema Copo de Nieve	31
Figura 20. Arquitectura de una solución de Inteligencia de Negocios.....	33
Figura 21. Pirámide de la Información	34
Figura 22. Esquema de Cuadros de Mando Integrales	38
Figura 23. Mapa de Factores Claves de éxito de la Gestión	41
Figura 24. Metodología general para establecimiento de Indicadores de Gestión.....	43
Figura 25. Plataforma Business Intelligence utilizando Dashboard	48
Figura 26. Arquitectura de Pentaho Business Intelligence	50
Figura 27. Componentes de Pentaho Business Intelligence	51
Figura 28. Arquitectura de Pentaho Data Integration	54
Figura 29. Fases del Método de Desarrollo de la Aplicación	55

Figura 30. Gráfico del Sistema de Publicación Científica	58
Figura 31. Diagrama de Actividades de un Sistema de Publicación.....	59
Figura 32. Modelo Entidad-Relación del ambiente intermedio.....	66
Figura 33. Jerarquía de la Información	67
Figura 34. Modelo Dimensional	68
Figura 35. Niveles jerárquicos y medidas.....	69
Figura 36. Algunos operadores de Pentaho Data Integration	76
Figura 37. Creación de las tablas del Modelo Entidad-Relación.....	77
Figura 38. Mapa para la creación de la dimensión artículo	78
Figura 39. Mapa para la creación de la dimensión revista	78
Figura 40. Mapa para la creación de la dimensión usuario	79
Figura 41. Mapa para construir la dimensión Tiempo	79
Figura 42. Creación de la tabla del cubo publicación.....	80
Figura 43. Creación del esquema del cubo	81
Figura 44. Creación de la Dimensión Tiempo	82
Figura 45. Creación de la dimensión Usuario.....	82
Figura 46. Creación de la dimensión Revista	83
Figura 47. Creación de la dimensión Artículo	83
Figura 48. Creación de los hechos a medir	84
Figura 49. Consulta 1 en el sistema transaccional de usuario	85
Figura 50. Consulta 1 de la dimensión Usuario.....	86
Figura 51. Consulta 2 en el sistema transaccional de usuario	86
Figura 52. Consulta 2 en la dimensión Usuario.....	86
Figura 53. Consulta 1 en el sistema transaccional de revista	87
Figura 54. Consulta 1 en la dimensión Revista.....	87
Figura 55. Consulta 2 en el sistema transaccional de ejemplar	88
Figura 56. Consulta 2 en la dimensión Revista.....	88
Figura 57. Consulta 1 del sistema transaccional en artículo.....	88
Figura 58. Consulta 1 en la dimensión Artículo	89
Figura 59. Consulta 1 en el sistema transaccional en artículos publicados	89
Figura 60. Consulta 1 en el cubo Publicación.....	90

Figura 61. Visualización de un cubo en Pentaho Analyzer.....	91
Figura 62. Consulta generada por el Pentaho Analyzer.....	91
Figura 63. Creación de la plantilla en CDE	92
Figura 64. Creación de los componentes en CDE.....	92
Figura 65. Creación de los orígenes de datos en CDE	93
Figura 66. Plantilla de los Cuadros de Mando.....	93
Figura 67. Consulta para filtro del año.....	95
Figura 68. Consulta para filtro de mes	95
Figura 69. Consulta para filtro de revista	95
Figura 70. Consulta para filtro de ejemplar	95
Figura 71. Consulta SQL para la tabla de Tiempo de asignación de revisión.....	96
Figura 72. Consulta MDX para el tacómetro de Tiempo de asignación de revisión	96
Figura 73. Consulta MDX para el gráfico de Tiempo de asignación de revisión	97
Figura 74. Consulta SQL para la tabla de Tiempo de revisión de un artículo	97
Figura 75. Consulta MDX para el tacómetro de Tiempo de revisión de artículo.....	98
Figura 76. Consulta MDX para el gráfico de Tiempo de revisión de artículo.....	98
Figura 77. Consulta SQL para la tabla de Tiempo de decisión final	98
Figura 78. Consulta MDX para el tacómetro de Tiempo de decisión final.....	99
Figura 79. Consulta MDX para el gráfico de Tiempo de decisión final	99
Figura 80. Consulta SQL para la tabla de Tiempo de publicación	100
Figura 81. Consulta MDX para el tacómetro de Tiempo de publicación	100
Figura 82. Consulta MDX para el gráfico de Tiempo de publicación	100
Figura 83. Consulta SQL para la tabla de Tiempo de publicación total	101
Figura 84. Consulta MDX para el tacómetro de Tiempo de publicación total.....	101
Figura 85. Consulta MDX para el gráfico de Tiempo de publicación Total	102
Figura 86. Consulta SQL para la tabla de comparación de indicadores.....	102
Figura 87. Consulta MDX para indicador de asignación	102
Figura 88. Consulta MDX para indicador de revisión.....	103
Figura 89. Consulta MDX para indicador de decisión final	103
Figura 90. Consulta MDX para indicador de publicación	103
Figura 91. Vista de la pestaña de Tiempo de asignación de revisión.....	104

Figura 92. Vista de la pestaña de Tiempo de revisión de artículo	105
Figura 93. Vista de la pestaña de Tiempo de decisión final	106
Figura 94. Vista de la pestaña de Tiempo de publicación.....	107
Figura 95. Vista de la pestaña de Tiempo total de publicación	108
Figura 96. Vista de la pestaña de Comparación de indicadores por revista	109

Índice de Tablas

Tabla 1. Cuadro Comparativo de Sistema Manejadores de Base de Datos.....	18
Tabla 2. Conceptos erróneos de Bodega de Datos	23
Tabla 3. Atributos de un Indicador de Gestión	40
Tabla 4. Cuadro Comparativo entre Plataformas BI Software Libre analizadas y estudiadas	46
Tabla 5. Indicadores contruidos con su estado, umbral y rango.....	64
Tabla 6. Indicadores de Gestión.....	65
Tabla 7. Dimensión Tiempo.....	70
Tabla 8. Dimensión Usuario	71
Tabla 9. Dimensión Revista	72
Tabla 10. Dimensión Artículo	73
Tabla 11. Cubo Publicación	74

Introducción

En la actualidad, las organizaciones han incrementado el uso de sistemas de información para obtener, además de los datos asociados a la operación del negocio, información relevante en forma de indicadores que miden hechos del negocio, mejorando así la eficiencia de sus procesos y la efectividad de sus resultados mediante la toma oportuna de decisiones, obteniendo mejoras visibles en su rentabilidad y mostrando una ventaja competitiva en el contexto donde se desenvuelve.

Generalmente, se ha recurrido a los sistemas transaccionales que llevan un registro detallado de los datos que se producen en los procesos de la organización, para la elaboración de informes de gestión, sobrecargando de esta manera su actividad. Estos informes son fundamentales ya que registran el resultado de los indicadores, el análisis experto de estos resultados y el conjunto de acciones recomendadas para corregir las desviaciones. Este informe es la base para la toma de decisiones. Mientras más datos tengan que analizarse, más evidente es la necesidad de contar con herramientas automatizadas de procesamiento analítico de la información para la obtención de estos indicadores.

Las plataformas de Inteligencia de Negocios traen consigo una variedad de componentes que permiten extraer los datos, ya sea de fuentes internas o externas, transformarlos y cargarlos en almacenes o bodegas de datos, que son estructuras sencillas pero poderosas en cuanto a la forma de sintetizar la información. Asimismo, permiten ver los indicadores desde distintas perspectivas o criterios de clasificación.

El presente documento del Trabajo Especial de Grado tiene como objetivo principal implementar indicadores de gestión, a través de una plataforma de Inteligencia de Negocios, orientados al monitoreo del rendimiento del proceso de Revisión Editorial, de un sistema de publicación. Adicionalmente, explica las bases teóricas investigadas y estudiadas para lograr una visión más amplia y contribuir a la comprensión del ambiente que emplea este tipo de soluciones.

El presente documento consta de cuatro capítulos, descritos a continuación.

En el primer capítulo se describe el Problema de Investigación, presentando la situación actual de una organización que se dedica a la generación de publicaciones informativas y el contexto del problema. Asimismo, se establece y se define el objetivo principal y cada uno de los objetivos específicos que componen dicho Trabajo Especial de Grado (T. E. G), junto con el alcance, importancia y justificación de la solución planteada.

El segundo capítulo está constituido por el Marco Conceptual, en donde se describen los conocimientos utilizados como base teórica para la elaboración de la solución propuesta. Inicialmente, se detalla el concepto de Sistemas de Información y sus diferentes tipos, el término de base de datos y su clasificación. Luego, se define almacenes de datos, bodega de datos y sus características. Se define el concepto de Inteligencia de Negocios e Indicadores de Gestión, detallando sus características y los distintos componentes que conforman una plataforma BI. Seguidamente, se explica la plataforma utilizada como herramienta tecnológica para el desarrollo de la solución, junto con sus características, componentes y criterios de selección de la misma. Por último, se describe el método de desarrollo empleado y sus diferentes fases junto con la descripción del proceso de Revisión Editorial.

El tercer capítulo está compuesto por el Marco Aplicativo, donde se describe la aplicación del método utilizado por la empresa Tian Consultores, para analizar, diseñar, construir e implementar soluciones de Inteligencia de Negocios, donde estas soluciones parten de la documentación apropiada de los indicadores a desarrollar. Dicho método está basado en la integración de dos métodos: el desarrollado por Beltran (2000) para la definición y documentación de indicadores de gestión; y el de desarrollo de soluciones de inteligencia de negocios, considerado como una adaptación de los métodos ascendente (Kimball) y descendente (Inmon), tomando las mejores prácticas de cada uno. Finalmente, se detallan cada una de las fases que componen el método de desarrollo empleado.

Por último, se presentan las conclusiones, recomendaciones y las referencias utilizadas para el desarrollo del presente T. E. G.

Capítulo I. Problema de la Investigación

En este primer capítulo se explica el contexto del problema, se describe el proceso de Revisión Editorial, se especifica el problema a resolver y la solución al mismo; así como también el alcance, justificación del problema, el objetivo general y los objetivos específicos del presente trabajo.

1.1 Título

“Implementación de Indicadores de Gestión para monitorear el rendimiento del proceso de Revisión Editorial, utilizando una plataforma de Inteligencia de Negocios, desarrollada en Software Libre.”

1.2 Situación Actual

Antes de la era internet, la impresión y distribución de publicaciones como periódicos, revistas y libros, eran realizadas por empresas llamadas Editoriales, a través de sistemas de publicación basados en la imprenta. Hoy en día el avance de las tecnologías, permite a las Editoriales, evolucionar a sistemas de publicación digital gestionados a través de internet.

Los sistemas de publicación son aquellos encargados de divulgar, hacer notorio y patente una variedad de contenidos que se desea propagar como noticia o información a las personas. No obstante, a pesar de las publicaciones impresas, también han surgido las publicaciones digitales. Asimismo, ambos llevan a cabo el mismo objetivo, diferenciándose en el medio de propagación (López & Estrada, 2006).

De igual forma, López & Estrada (2006) aseguran, de una forma más pragmática, que existen dos tipos de publicaciones: las de divulgación, dirigidas a un público no especializado para informarlo e inducirlo a un tema; y las especializadas, dirigidas a un público específico, ya que contienen artículos especializados sobre investigaciones, donde se da a conocer información sobre una rama académica.

Para Barruecos (2000), el sistema de publicación especializado se ha dividido en comunicación formal e informal. Por su parte, Barruecos afirma que la comunicación informal es aquella donde

no existe ningún método de control o certificación de calidad, por lo que se considera una tipología de información heterogénea, que cuenta con documentos como actas de congreso, pre publicaciones, así como también el contacto entre especialistas. Por el contrario, CINDOC & CSIC (2004) establecen que a partir de las primeras publicaciones científicas, y debido a la gran cantidad de artículos que se producían, fue necesario asumir funciones de selección, revisión y edición de los artículos enviados por los investigadores. Así, se obtiene la comunicación formal, donde el artículo final de la investigación sufre un proceso de revisión antes de ser publicado.

Asimismo, Barruecos (2000) asevera que a pesar de las modificaciones que han sufrido, tanto el proceso de elaboración como el de edición de una publicación digital, dicho sistema de publicación de artículos ha permanecido inalterable hasta nuestros días.

Actualmente, para algunos sistemas de publicación, tanto impresos como digitales, es de suma importancia el proceso de revisión de artículos, o bien llamado proceso de revisión editorial, que se realiza previamente a la publicación. Según el ICSU (2001), el proceso de revisión es fundamental, ya que el artículo debe alcanzar ciertos criterios de control y calidad, previo la llegada a sus lectores.

Generalmente, en los sistemas de publicación, el proceso editorial recae sobre el editor o el consejo editorial, de tal forma que al llegar un nuevo artículo el editor selecciona el número de árbitros, entre un grupo, necesarios para la evaluación del mismo, antes de ser publicado. Estos árbitros, serán los encargados de analizar, revisar y evaluar los artículos asignados, con el fin de asegurar ciertas criterios de calidad (Campanario, 2002).

Este proceso de Revisión Editorial descrito, requiere el manejo eficiente y efectivo de los tiempos en cada subproceso, para garantizar la oportunidad y disponibilidad de la publicación.

1.3 Planteamiento del Problema

Los sistemas de publicación que hacen uso del medio electrónico para la realización del proceso de divulgación de artículos, desde el envío por parte del autor, pasando por la revisión del mismo, hasta su publicación; cuentan con una base de datos transaccional, donde se almacena toda la información de los artículos que se envían.

Todo artículo antes de ser publicado pasa por una serie de subprocesos que componen el proceso de Revisión Editorial. Primeramente, el autor envía el artículo, a dicho artículo se le debe asignar un revisor y éste debe aceptar la asignación. Una vez que el revisor acepta la revisión, procede a realizar la misma y cuando finalice, envía su evaluación al editor, quien será el encargado de tomar una decisión final de publicación. Por último, dependiendo de la decisión, el artículo se publica.

Como vemos, los actores que interactúan en el proceso de Revisión Editorial, generan distintos tipos de relaciones que hacen uso de uno de los recursos más críticos: el tiempo.

Es por esto, que la principal razón para llevar a cabo el control del rendimiento del proceso de Revisión Editorial está relacionada con el tiempo, es decir, realizar la medición de cada uno de los subprocesos que componen al mismo, con el fin de identificar la cantidad de tiempo en que se divulga un artículo en un sistema de publicación digital.

En general, las organizaciones encargadas de la propagación de artículos no cuentan con herramientas que les permitan obtener información oportuna, relacionada con el tiempo en el proceso de Revisión Editorial, lo que trae consigo una demora en la toma de decisiones de la gerencia, para optimizar dicho recurso.

Los indicadores o mediciones son fundamentales para el apoyo de la toma de decisiones por la gerencia, ya que permiten mejorar la eficiencia del proceso y la efectividad de los resultados obtenidos. Por esta aseveración, se hace necesario obtener información oportuna que permita medir continuamente la funcionalidad del proceso.

Para la problemática planteada, se genera la siguiente interrogante:

¿Cómo proveer información relevante, automatizada y oportuna a la Gerencia de un Sistema de Publicación Digital, que le permita tomar decisiones relacionadas al tiempo de duración del proceso de Revisión Editorial?

1.4 Solución Propuesta

Una vez conocido el problema que tienen los sistemas de publicación digital para llevar a cabo el control y seguimiento del proceso de revisión de artículos, por medio de mediciones desarrolladas en un menor tiempo y con datos precisos que representen la situación de dicha publicación, se propone implementar un conjunto de estrategias y herramientas claramente definidas que permitan obtener información relevante para la toma de decisiones.

Tomando en cuenta lo expuesto anteriormente, se propone implementar una serie de indicadores de gestión, por medio de una plataforma de Inteligencia de Negocios, donde se realice una representación gráfica orientada a los requerimientos establecidos, se obtengan los datos de los sistemas transaccionales, se carguen en un almacén de datos y de donde finalmente, se extraiga la información necesaria que será representada en dichos indicadores, con el fin de ser utilizados por la gerencia del sistema de publicación digital, a la hora de tomar decisiones.

Para lograr esto, en la figura 1, se plantea una arquitectura con dos ambientes: un ambiente transaccional y un ambiente BI (*Business Intelligence*, por sus siglas en inglés).

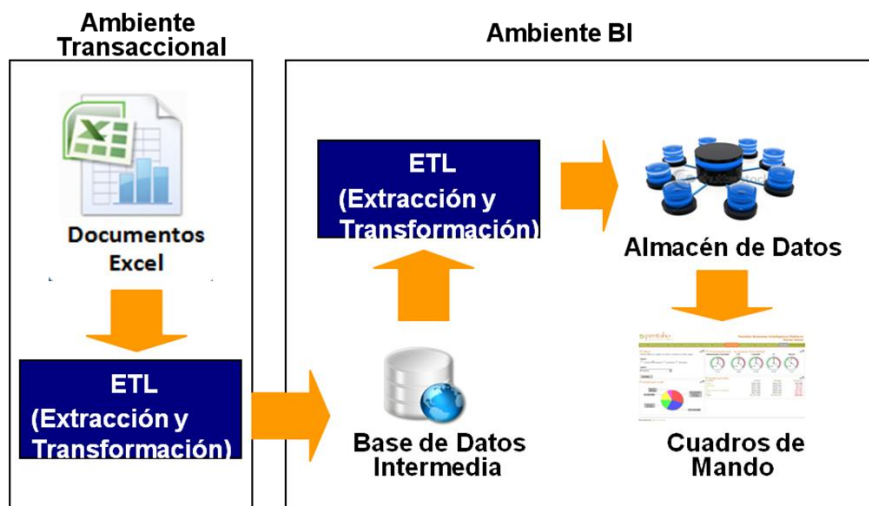


Figura 1. Arquitectura de la solución planteada

El ambiente transaccional, según se indica en la figura 1, parte de información transaccional que está contenida en hojas Excel. Esto se debe a que la solución que se presenta es genérica y puede adaptarse a cualquier plataforma de publicación digital que contenga por lo menos, la

estructura de datos que se propone con la hoja Excel. A estos archivos, se le aplica un proceso de extracción y transformación para construir un ambiente de datos intermedio, llamado Base de Datos Intermedia, que obtendrá los datos que cumplan con los requerimientos mínimos para realizar un modelo de Inteligencia de Negocios.

Una vez asegurado los datos necesarios en la base de datos intermedia, se dispone a un proceso de extracción y transformación para realizar la carga en el almacén de datos modelado en el modelo dimensional. Por último, se usarán los datos del almacén, desde las herramientas especializadas de la solución BI para presentarse a los usuarios finales. En este caso, los indicadores serán visualizados sobre cuadros de mando. Los cuadros de mando son herramientas que representan visualmente indicadores numéricos, permitiendo mantener el seguimiento y monitoreo de los mismos a lo largo de un período.

Se utilizan un conjunto de herramientas que forman parte de la plataforma *Pentaho Business Intelligence*:

- *PostgreSQL 9.1.*
- *Pentaho Data Integration (Kettle).*
- *Pentaho Schema Workbench.*
- *Pentaho BI Server 3.9.*
- *Pentaho Analyzer.*
- *Pentaho Report Designer.*
- *Pentaho Community Dashboard Editor 1.0.*

1.5 Objetivo General

Implementar indicadores de gestión, que permitan llevar a cabo el monitoreo del rendimiento del proceso de Revisión Editorial, utilizando una plataforma de Inteligencia de Negocios de Software Libre.

1.6 Objetivos Específicos

- Estudiar el modelo operativo del proceso de Revisión Editorial de un sistema de publicación digital.
- Realizar el levantamiento de información y requerimientos.
- Definir los indicadores de gestión.
- Implementar las fuentes de datos, donde se extraerán los datos necesarios para modelar la solución.
- Diseñar el modelo dimensional.
- Diseñar y Construir los procesos de extracción, transformación y carga (ETL).
- Implementar el almacén de datos, que contendrá los datos de las fuentes, con el fin de integrarlos y estandarizarlos.
- Realizar las pruebas necesarias al almacén para verificar la consistencia de los datos.
- Implementar los indicadores necesarios para modelar los requerimientos.
- Construir los cuadros de mandos necesarios.
- Realizar las pruebas necesarias para validar el correcto funcionamiento de la solución.

1.7 Alcance

La solución presentada implementa un conjunto de indicadores para hacer seguimiento y asegurar el uso eficiente del tiempo, que como hemos dicho, es un recurso esencial en el proceso de Revisión Editorial para la publicación de artículos. Adicionalmente, la visualización de dichos indicadores será bajo una plataforma de Inteligencia de Negocios de Software Libre. Así, se genera información relevante que apoye en la toma de decisiones de la gerencia del sistema de publicación.

Adicionalmente, se puede afirmar que el desarrollo de la misma permite a los usuarios finales:

- Integrar y estandarizar los datos desde las fuentes y almacenarlos en un repositorio centralizado.
- Realizar consultas, en tiempo real, de la información almacenada en el repositorio.
- Generar cuadros de mandos parametrizables con la información necesaria del almacén.

- Realizar comparaciones del valor de un indicador a través del tiempo, con el fin de detectar tendencias.
- Facilitar la creación de nuevos indicadores en el cuadro de mando.

1.8 Importancia y Justificación

Generalmente, los sistemas de publicación digital requieren de información relevante y oportuna, para llevar a cabo el monitoreo del proceso de Revisión Editorial. Así, se permite mantener el control y seguimiento del proceso para mejorar su eficiencia y la efectividad de los resultados obtenidos.

Es por esto que en el presente T. E. G. se plantea una posible alternativa que solvete los inconvenientes mencionados, con el fin de ayudar en la toma de decisiones de la gerencia, y además, tener la posibilidad de respaldar el monitoreo de la revisión de artículos enviados por los autores.

Por lo tanto, se describe la importancia de implementar una serie de indicadores de gestión, bajo una plataforma de Inteligencia de Negocio, para llevar a cabo el monitoreo del proceso de Revisión Editorial, ahorrando tiempo y generando información oportuna, y así, poder contribuir de una manera adecuada a la toma de decisiones con datos reales y en poco tiempo.

Esta solución planteada trae consigo algunos beneficios, tales como:

- Automatización en la elaboración de indicadores que se utilizan para monitorear la revisión de artículo, disminuyendo el esfuerzo y tiempo.
- Facilitar el proceso de análisis de dicho monitoreo con el fin de ayudar en la toma de decisiones.
- Permitir realizar comparaciones en diferentes períodos de tiempo, con el fin de detectar tendencias y variaciones.

Capítulo II. Marco Conceptual

Este segundo capítulo tiene como objetivo principal describir los conocimientos utilizados como base teórica para el desarrollo del presente Trabajo Especial de Grado. En el presente capítulo se distinguen ocho secciones que engloban todo el desarrollo del Sistema de Inteligencia de Negocios, para el monitoreo del proceso de Revisión Editorial de un sistema de publicación digital.

En la primera sección se explica detalladamente la definición de Sistemas de Información y los diferentes tipos que se pueden distinguir en una organización o empresa.

En la segunda sección se define el término de base de datos, junto con los diferentes tipos de modelado de las mismas y su clasificación. De igual forma, se realiza un cuadro comparativo sobre algunos sistemas manejadores de base de datos, con el fin de detectar aquel que se adapte a las necesidades del presente trabajo.

La tercera sección está compuesta por la descripción del tipo de base de datos que existen para almacenar la información proveniente de fuentes externas o internas. Es por esta razón que se definen conceptos de almacén de datos, bodega de datos y el proceso de extracción, transformación y carga necesario para llevar los datos desde las fuentes hacia los almacenes o bodegas.

Seguidamente se detalla la definición de Inteligencia de Negocios, sus características, su arquitectura, las herramientas utilizadas para la representación de datos modelados y los beneficios o ventajas que trae consigo un sistema de Inteligencia de Negocios para una empresa u organización. Así, se compone la cuarta sección del presente capítulo.

En la quinta sección se especifica el término de indicadores de gestión, explicando los diferentes tipos que existen y detallando la metodología planteada por Beltran (2000) para la definición de los mismos.

En la sexta sección se establece un cuadro comparativo entre tres plataformas de Inteligencia de Negocios, bajo Software Libre. Además, se explica detalladamente la plataforma utilizada para

el desarrollo de la solución, que mejor se adapta a las necesidades iniciales, describiendo cada uno de los componentes que la conforman.

Luego, en la séptima sección, se describe el método de desarrollo de la aplicación y cada una de las fases contempladas en el mismo para llevar a cabo con éxito un sistema de Inteligencia de Negocios.

Por último, se desarrolla una breve descripción del área de trabajo hacia la cual se enfoca la solución de Inteligencia de Negocios. Por lo tanto, se explica el funcionamiento del proceso de Revisión Editorial de un sistema de publicación digital.

2.1 Sistemas de Información

2.1.1 Definición

Según Saroka (2002), un sistema de información es un conjunto de recursos humanos, materiales, financieros, tecnológicos, normativos y metodológicos, organizados para apoyar a los involucrados en la toma de decisiones de una organización, y así obtener información necesaria para desarrollar sus respectivas funciones.

Por otra parte, Laudon K. & Laudon J. (2004), definen a los sistemas de información como un conjunto de componentes interrelacionados que recogen (u obtienen), procesan, almacenan y distribuyen información para ayudar en la toma de decisiones y el control de una organización, tal como se muestra en la figura 2.

Tomando en cuenta las definiciones anteriores, se puede afirmar que un sistema de información es un conjunto de procesos que se encargan de recolectar, procesar, almacenar y distribuir información específica para actividades de dirección y control, apoyando en parte a la toma de decisiones necesarias para el desarrollo de las actividades y funciones de negocios de una organización.

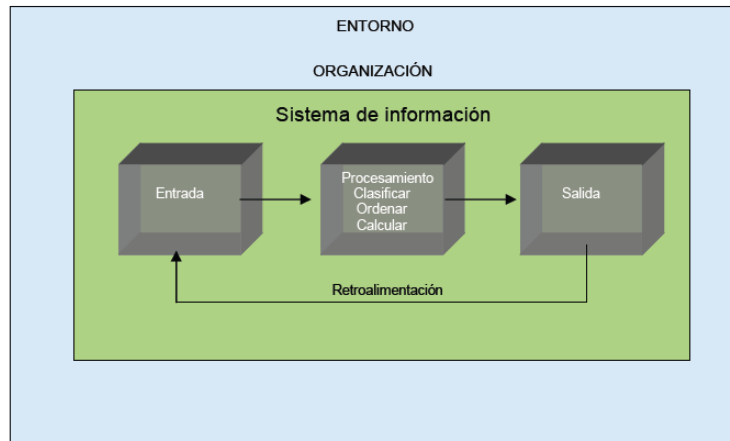


Figura 2. Sistemas de Información. (Laudon K. & Laudon J., 2004)

2.1.2 Tipos de Sistemas de Información

Según Laudon K. & Laudon J. (2004), los sistemas de información se dividen en cuanto a los niveles de la organización. En la figura 3, se pueden observar los tipos de sistemas de información, según los niveles de una organización:

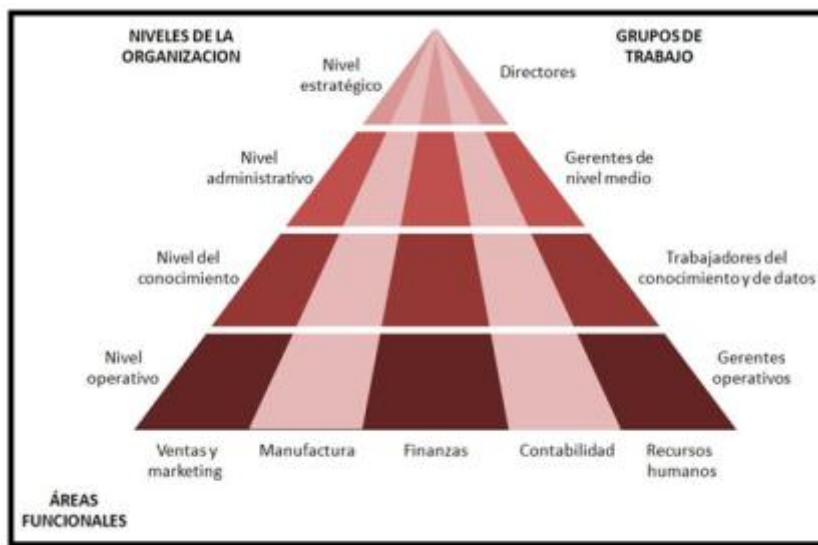


Figura 3. Tipos de Sistemas de Información (Laudon K. & Laudon J., 2004)

A continuación, se describen cada uno de los niveles:

- **Sistemas a Nivel Operativo:** Son aquellos que apoyan a los gerentes operativos en el seguimiento de las actividades y transacciones cotidianas y elementales de la organización. Este nivel cuenta con sistemas de procesamiento de transacciones (TPS, por sus siglas en inglés), que efectúan y registran transacciones diarias de los negocios.
- **Sistemas a Nivel de Conocimiento:** Son aquellos que apoyan la integración de nuevos conocimientos del negocio, ayudando a controlar el flujo de trabajo dentro del negocio. Este nivel cuenta con los Sistemas de Trabajo de Conocimiento (KWS, por sus siglas en inglés), que apoyan a trabajadores profesionales, generándoles nuevos conocimientos que contribuyen a la organización; y los Sistemas de Oficina que ayudan, principalmente, a los trabajadores de datos, quienes usan la información para analizarla y transformar datos.
- **Sistemas a Nivel Administrativo:** Son aquellos que apoyan a las actividades de supervisión, control y toma de decisiones. En este nivel se encuentran los Sistemas de Información Gerencial (MIS, por sus siglas en inglés), quienes se encargan de dar soporte a las tareas organizacionales, proporcionando informes en períodos de tiempo, que ayuden a los gerentes a la toma de decisiones, control y planeación a nivel administrativo. Adicionalmente, se encuentran los Sistemas de Apoyo a la Toma de Decisiones (DSS, por sus siglas en inglés), quienes tienen la tarea de ayudar a los gerentes a tomar decisiones exclusivas, cambiantes y no fáciles de anticipar.
- **Sistemas a Nivel Estratégico:** Son los que apoyan a las actividades de planificación de las organizaciones a largo plazo, y además, ayudan a los directores a enfrentar y resolver aspectos estratégicos; dando así, los cambios necesarios del entorno externo de la organización. En este nivel se encuentran los Sistemas de Apoyo a Ejecutivos (EES, por sus siglas en inglés), quienes dan soporte a la toma de decisiones no rutinarias que requieren juicio, evaluación y comprensión, es decir, apoyan a la toma de decisiones mediante gráficos y comunicaciones avanzadas.

2.2 Base de Datos

2.2.1 Definición

Para Oppel (2009), una base de datos no es más que una colección de datos interrelacionados que son manejados por una sola unidad. Es decir, se considera un sitio físico y lógico donde se almacenan datos, pero estos datos están relacionados entre sí, para darle un sentido al almacenamiento que estamos haciendo.

Una base de datos implica, además de estructuras y modelos, un software que soporte los mismos. Esto quiere decir que existen distintas maneras de modelar y de crear una base de datos y que en estas bases de datos intervienen distintos elementos como los datos en sí, objetos de bases de datos, archivos, entre otros. Pero las bases de datos dependen de un software el cual es el que soportará lógicamente esta estructura, a este software se le conoce como Sistema Manejador de Bases de Datos (Oppel, 2009).

Un sistema manejador de bases de datos provee herramientas y servicios que permiten ver, manipular y manejar las bases de datos. Dichos servicios permiten la manipulación de los datos que se encuentran en la base de datos física, el manejo multiusuario de las bases de datos, el manejo de las transacciones que se realizan en la base de datos, dar soporte de comandos que permiten editar y ver datos, el respaldo de las bases de datos y mecanismos de seguridad para prevenir que solo usuarios autorizados vean y manipulen datos. En el mercado existen distintos sistemas manejadores de bases de datos tales como *Oracle, MySQL, Access, Microsoft SQL Server, PostgreSQL*, entre los más destacados.

Es importante resaltar que las bases de datos cuentan con una capa física, que son los componentes en los cuales se almacenan físicamente los datos y cuentan con una capa lógica, que es el modelo en el cual se basa para almacenar y relacionar lógicamente los datos.

2.2.2 Modelo de Base de Datos

Según Oppel (2009), existen distintas estrategias para crear modelos para el almacenamiento en bases de datos, a continuación se desarrollan los modelos principales:

- **Modelo jerárquico:** Es un modelo en el que los registros se convierten en nodos, y a través de las relaciones entre nodos, se forma una jerarquía, lo cual quiere decir que lógicamente un nodo estará encima de otros, siendo este el padre de los nodos sobre los cuales esté, y siendo hijo del nodo que esté encima de él. La conexión o relación entre los nodos se hace vía apuntadores, y son relaciones uno a muchos, lo cual quiere decir que un nodo puede tener muchos hijos, pero solo puede tener un padre. En la figura 4, se puede apreciar un ejemplo de dicho modelo.

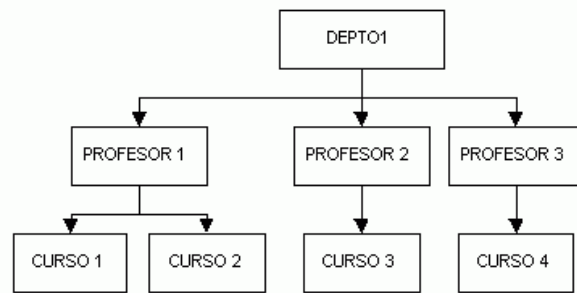


Figura 4. Modelo Jerárquico (Moraga, 2001)

- **Modelo red:** Este modelo se diferencia del anterior en que cambia de ser una relación uno a muchos a ser una relación muchos a muchos, lo cual significó un cambio muy importante ya que esto permitía navegar con más rapidez entre los nodos y relacionarlos de una mejor manera entre sí. Sin embargo este modelo sigue constituyendo una jerarquía, lo cual quiere decir que aun existe esa figura de padre-hijo entre los nodos. En la figura 5, se muestra un ejemplo de este modelo.

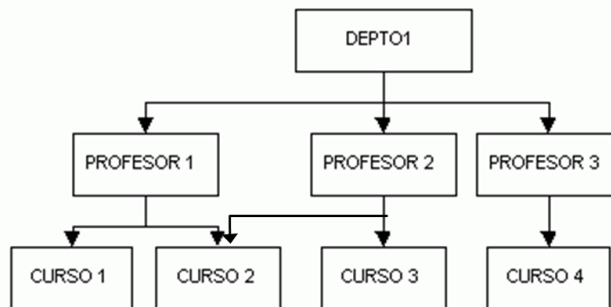


Figura 5. Modelo Red (Moraga, 2001)

- **Modelo relacional:** Es probablemente el modelo más utilizado en la actualidad, el modelo relacional se impuso a los dos anteriores por ser un sistema mucho más flexible y generalizado. Permite relaciones de uno a uno, uno a muchos y muchos a muchos, y el elemento a relacionar se denomina entidad, por esto es común escuchar que se nombra a los modelos que representan base de datos relacionales como modelos entidad-relación. Adicionalmente, se apoya en claves e índices para relacionar y poder viajar de una tabla a otra y conocer sus relaciones, en este caso se utiliza una clave primaria para reconocer al registro una clave foránea que deberá ser una clave primaria de otro registro. Si la relación es de muchos a muchos, se creará una estructura llamada relación, la cual servirá como una especie de puente o conexión entre ambas entidades. En caso de ser relaciones de uno a muchos la clave primaria del uno estará contenida en la entidad muchos, y en el caso uno a uno entonces la clave puede estar en cualquiera de las dos entidades. La figura 6 muestra un ejemplo de este tipo de modelo.

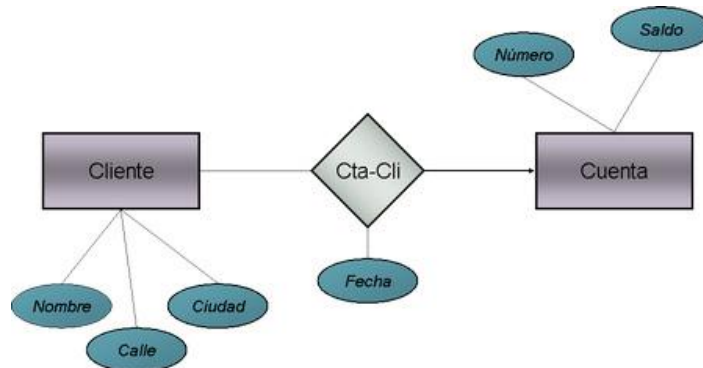


Figura 6. Modelo Relacional (Moraga, 2001)

- **Modelo orientado a objetos:** Este modelo se basa (como su propio nombre lo dice) en objetos, estos objetos agrupan datos y funciones asociadas con lo que debe ser un objeto real en el computador. En la figura 7 se puede apreciar un ejemplo de dicho modelo.

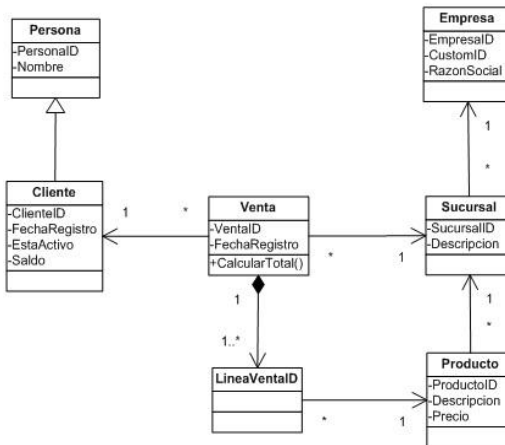


Figura 7. Modelo Orientado a Objetos (Moraga, 2001)

- Modelo objeto-relacional: Este modelo explota las virtudes del modelo relacional y el orientado a objeto. Dicho modelo surge principalmente por la necesidad de manejos de datos complejos como se mencionan anteriormente pero a su vez brinda las facilidades de consulta que brindan los modelos de base de datos relacionales.

2.2.3 Tipos de Base de Datos según su propósito

Además de tener un propósito general, el cual es el almacenar los datos, las bases de datos pueden tener distintas funciones y pueden servir para distintos propósitos, por ello se cuentan con distintas denominaciones (Oppel, 2009). Se contemplan solo los tipos de bases de datos que podrían ser útiles para el propósito de la investigación:

- Almacén de datos: Es considerado un sistema que recupera y consolida datos periódicos desde sistemas fuentes a sistemas dimensionales o almacena datos normalizados (Rainardi, 2008). Los almacenes de datos son utilizados regularmente para sistemas de Inteligencia de Negocios, por su capacidad de almacenar los datos en forma dimensional y de además cargar datos normalizados.
- Base de datos transaccionales: es una base de datos regular pero cuyo fin es realizar conexiones rápidas, es decir, es una base de datos usada frecuentemente para transacciones. Estas bases de datos cuentan regularmente con grandes volúmenes de

datos pero también con gran capacidad de conexión para que las transacciones se realicen con la rapidez que ameriten.

- Base de datos distribuida: son un conjunto de bases de datos en distintos espacios físicos y que se comunican entre sí a través de una red de computadores. Las bases de datos distribuidas trabajan en conjunto dos áreas importantes de la computación, las bases de datos y las redes, ya que almacenan datos por medio de distintas bases de datos y utilizan los protocolos de comunicación de redes para la comunicación entre ellas.

2.2.4 Sistemas Manejadores de Base de Datos

Un sistema manejador de base de datos, es un software que permite la interacción entre un usuario y distintas bases de datos. Los sistemas manejadores de bases de datos proveen distintas interfaces y funcionalidades que permiten al usuario interactuar de forma natural con las bases de datos. A continuación, en la tabla 1, se muestra un cuadro comparativo entre cuatro sistemas manejadores de base de datos, de los más conocidos.

Tabla 1. Cuadro Comparativo de Sistema Manejadores de Base de Datos (Fuente: El Autor)

SMBD/ Característica	PostgreSQL	Oracle	MySQL	Fastobjects
Multiplataforma	Sí.	Sí.	Sí.	Sí.
Multiusuario	Sí.	Sí.	Sí.	Sí.
Modelo(s)	Relacional, Objeto-Relacional	Relacional, Objeto-Relacional	Relacional	O.O
Licencia	Libre	Propietaria	Propietaria	Libre

En la tabla 1 se observan las diferencias entre las características evaluadas de los sistemas manejadores de bases de datos, con la finalidad de escoger aquel sistema que cuente con los mejores beneficios y características para el desarrollo del presente T. E. G.

2.3 Almacén de Datos y Bodega de Datos

2.3.1 Almacén de Datos (*Data Warehouse*)

Según Kimball (1997), un almacén de datos, no es más que una copia de datos transaccionales, específicamente estructurados, utilizados para consulta y análisis.

Por su parte Inmon (1998), define un almacén de datos como un conjunto de datos integrados, orientados a materia, variantes con el tiempo y no transitorios, los cuales soportan el proceso de toma de decisiones de una administración.

Partiendo de estas definiciones, de dos de los pioneros en el área de almacenes de datos, se puede afirmar que ambas se centran en los datos per se. Por lo tanto, podemos definir un almacén de datos, de forma más generalizada, como un repositorio centralizado de datos, que almacena gran cantidad de información, utilizada para apoyar en la toma de decisiones de una organización. Generalmente, estos almacenes son caracterizados por ser la fuente que alimenta los procesos analíticos dentro de una organización. Adicionalmente, los almacenes de datos se encargan de guardar información histórica y permiten principalmente consultas, debido a que por lo general están conectados a herramientas que extraen la información para mostrarla a gerentes o analistas para la toma de decisiones.

2.3.1.1 Arquitectura de un Almacén de Datos

A continuación, en la figura 8, se observa la arquitectura de un almacén de datos:

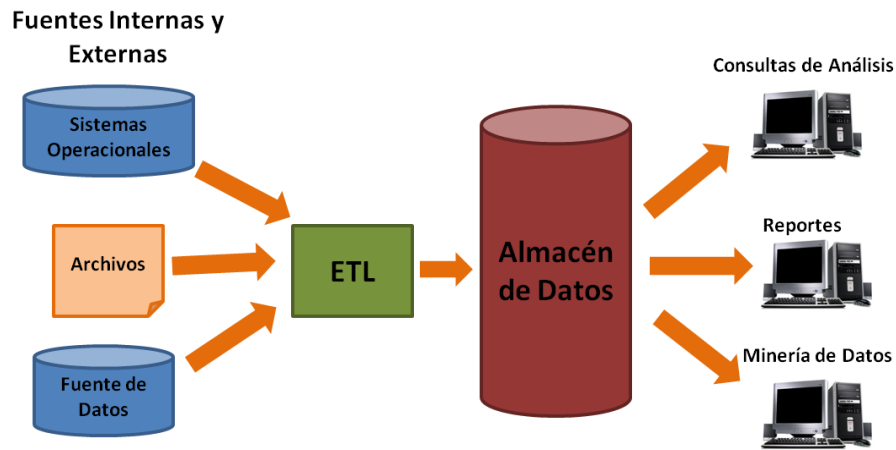


Figura 8. Arquitectura de un Almacén de Datos (EcuRed.cu, 2012)

En la figura 8, se puede apreciar como el almacén se caracteriza, por medio de un proceso de extracción, transformación y carga (ETL, por sus siglas en inglés) de extraer datos, ya sea de fuentes internas o fuentes externas, transformar y procesar dichos datos recolectados y almacenarlos en el repositorio. Finalmente, por medio de herramientas especializadas, los datos se mostrarán en reportes o gráficos para su posterior análisis.

2.3.1.2 Características

Según Immon (1998), uno de los aspectos más importantes que tienen los almacenes de datos es que éstos se han definido en términos de sus características, los cuales con el tiempo, han definido su estructura. Dichas características se describen a continuación:

- **Orientado a Temas:** Ya que la información es clasificada en base a los aspectos más importantes de la organización. Principalmente, se busca centrar la información en las actividades básicas, y no en procesos, de dicha organización, ya sea ventas, producción, distribución, etc. Así, se facilita el acceso y entendimiento de los datos por parte de los usuarios. La diferencia entre en actividades y procesos se ubica en el nivel de detalle de los datos. En la figura 9, se aprecia esta característica.

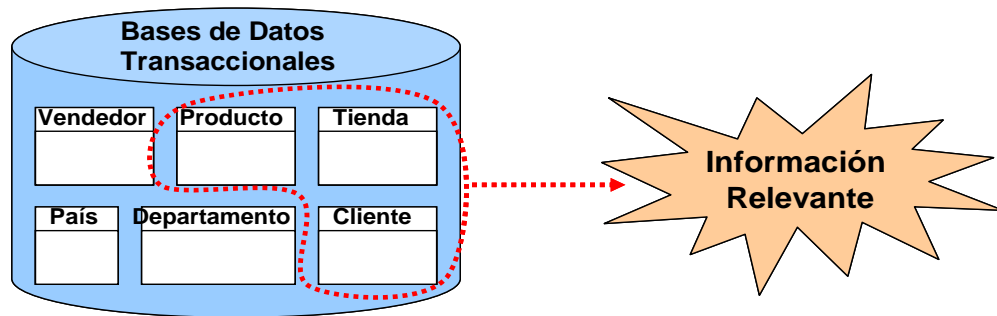


Figura 9. Orientación al tema (Herrera, 2007)

- Variante en el tiempo: Esto se debe a que los datos del almacén son relativos a un período de tiempo, como se observa en la figura 10. Las variaciones que se produzcan en los datos, a lo largo del tiempo, deben quedar registradas periódicamente. De esta forma, se permite generar reportes que reflejen el comportamiento de dichas variaciones.

	Bases de Datos Operacional	Almacén de Datos
	Valor actual de los datos	Datos instantáneos
Horizonte de tiempo	60-90 días	5-10 años
Clave	Puede o no tener un elemento de tiempo	Tiene un elemento de Tiempo
Datos	Pueden ser actualizados	Una vez que la foto (snapshot), no se puede actualizar

Figura 10. Variación en el tiempo (Herrera, 2007)

- No Volátil: Ya que la información que se encuentra en el almacén, no puede ser modificada, ni eliminada, solamente se permite agregar. En la figura 11 se puede observar esta característica, en donde los datos que son cargados en el almacén sólo podrán servir de lectura. Por lo tanto, se puede afirmar que los almacenes de datos permiten dos operaciones: carga inicial de los datos y acceso a los mismos. Así, se elimina la necesidad de tener mecanismos de control de concurrencia y recuperación de datos.

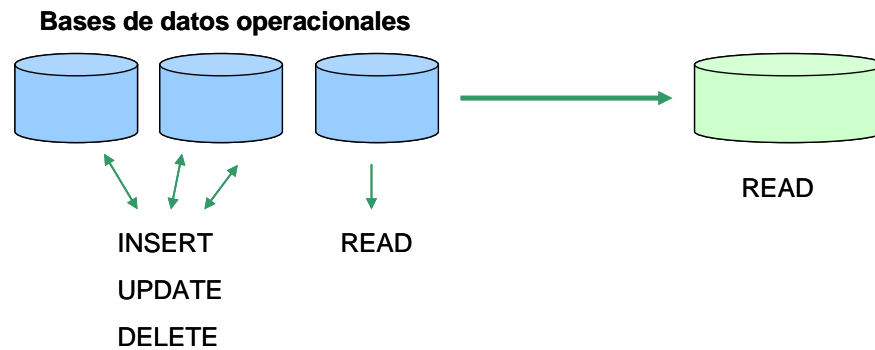


Figura 11. Almacenes no volátiles (Herrera, 2007)

- Integrado: Debido a que los datos del almacén deben ser integrados en estructuras consistentes, eliminando así las inconsistencias existentes en las fuentes de datos, como se muestra en la figura 12. Esta característica permite reunir todos los datos, provenientes de las diferentes fuentes, y agruparlos en un modelo generalizado, único y coherente. Adicionalmente, estos datos deben ser transformados antes de incorporarlos al almacén, asegurando así la consistencia de los mismos.

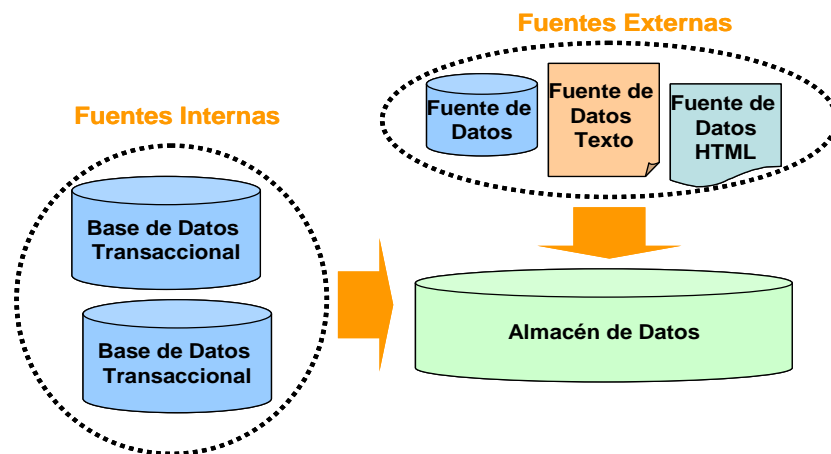


Figura 12. Integridad del almacén (Herrera, 2007)

2.3.2 Bodega de Datos (Data Mart)

Según Kimball (1997), los almacenes de datos son más que la unión de todas las bodegas de datos. Sin embargo, Inmon (1998) establece que a pesar de atrapar a todos los peces del océano y juntarlos, no harán una ballena.

Apreciando las definiciones de dos de los pioneros de los almacenes de datos, se puede observar una cierta contradicción entre ambas definiciones. Generalmente, se constituye una mala interpretación de su uso al momento del desarrollo e implementación de una bodega de datos. Es por esto, que Inmon (1998) estructuró una tabla de conceptos erróneos, que se puede apreciar en la tabla 2, donde explica la verdad de los mismos.

Tabla 2. Conceptos erróneos de Bodega de Datos (Inmon, 1998)

Conceptos Erróneos	Verdad
Son más simples que un Almacén de Datos.	La implementación de una bodega de datos es muy similar a la de un almacén dado que ambos poseen las mismas funcionalidades.
Son pequeños conjuntos de datos y, por ende, necesitan menor cantidad de recursos.	Las aplicaciones que corren sobre bodegas de datos necesitan la misma cantidad de recursos que si corrieran en un almacén de datos.
Las consultas son más rápidas debido a la menor cantidad de datos que manejan.	La menor cantidad de datos se debe a que estos poseen los datos de un sector de la organización, sin embargo las consultas sobre dicho sector tardan lo mismo que si se hicieran sobre el almacén de datos directamente.
En ciertos casos, añade tiempo a los procesos de actualización.	Actualizar la bodega desde el almacén cuesta menos que actualizar el almacén desde las fuentes de datos primarias, donde se necesitan aplicar operación de transformación.

Por lo tanto, se puede afirmar que una bodega de datos es un repositorio que almacena datos orientados a un tema, con una estructura similar a la de un almacén, con la diferencia que contiene datos de un sector específico de la organización, para dar soporte a la toma de decisiones. Además, para Inmon (1998) las bodegas de datos se construyen con un modelo dimensional.

2.3.3 Extracción, Transformación y Carga (ETL)

Según el Portal de ETL y Data Warehousing (ETLTools.org, 2009), la extracción, transformación y carga es un proceso usado en las áreas de base de datos, especialmente en los almacenes de

datos, con el fin de migrar de un punto a otro. Este proceso consta de tres fases, que se combinan en una sola herramienta, facilitando el transporte de los datos, desde la fuente hasta el almacén. Las tres fases de este proceso se describen a continuación:

- a) Extracción: Es la primera fase del proceso ETL, en esta fase se obtienen los datos provenientes de las diferentes fuentes, sean externas o internas, que serán utilizados por el almacén de datos. Esta fase busca estandarizar los datos, de tal forma que los mismos sean entendibles en la fase de transformación. Además, se incluye un filtrado de datos para eliminar la redundancia ó los datos de poco interés.
- b) Transformación: En esta fase se aplican una serie de reglas y funciones a los datos extraídos y que serán cargados posteriormente en el almacén. En algunos casos, se aplican a los una serie de transformaciones, como por ejemplo: ordenación de los datos, unión de los datos desde diferentes fuentes, validación, etc.
- c) Carga: Representa la última fase del proceso ETL, y ésta se encarga de guardar los datos transformados en el almacén. Debido a que en esta fase existe una interacción directa con la base de datos, al momento de la inserción de datos, generalmente se activan los disparadores, o los llamados *triggers*, y las restricciones existentes.

2.3.4 Modelo Dimensional

Según Loshin (2003), generalmente los almacenes de datos son diseñados y construidos basándose en un modelo dimensional. Dicho modelo dimensional es considerado una variación del modelo relacional, pero no necesariamente involucra bases de datos relacionales.

Por su parte Kimball (1997), considera que un modelo dimensional es una técnica de diseño lógico que busca presentar los datos de manera estándar e intuitiva, y que además, permita un acceso de alto rendimiento.

Así, podemos concluir que un modelo dimensional es una variación del modelo relacional, que proporciona al usuario final una forma intuitiva para construir consultas y agilizar tiempos de respuestas.

Cuando se diseñan bases de datos utilizando el modelo dimensional se definen:

a) Hecho

El hecho, o también llamado *Fact*, es una medida o indicador que representa un evento dentro de una organización. Generalmente, el hecho viene representado en un valor numérico o agregado (Rasmussen, N., Goldy, P., & Solli, P., 2002). Cuando se habla de un valor agregado, se refiere a aquellos hechos que derivan de funciones como la suma, el promedio, el máximo, el mínimo, etc. Por ejemplo, se puede tener un hecho de valor agregado que defina el sueldo máximo de los empleados de la organización.

b) Dimensiones

Una dimensión se puede definir como una entidad independiente de un modelo dimensional, que agrupa información por categorías y se utiliza como mecanismo de acceso a dicha información. Generalmente, las dimensiones representan el contexto de un evento o hecho y responde a preguntas como qué, quién, dónde y cuándo (Rasmussen, N., Goldy, P., & Solli, P., 2002).

Asimismo, se dice que las dimensiones son cada uno de los ejes en un espacio dimensional. Generalmente esta información es estática en el tiempo y representan las distintas formas de ver la información, en el modelado dimensional se conoce como tablas de dimensiones (Kimball, R. & Ross, M., 2002).

En los almacenes de datos, las dimensiones permiten ver y analizar la información desde diferentes formas y se caracterizan por proporcionar filtrado, agrupación y etiquetado de datos.

c) Atributos de Dimensiones

Cuando se habla de atributos de dimensiones, se refiere a los atributos de tipo cualitativo que suministra el contexto en el que se obtienen las medidas (Kimball, R. & Ross, M., 2002), como por ejemplo: ciudad, estado, país, continente, etc.

d) Jerarquías

El término jerarquía tiene vital importancia cuando se habla de cubos de información, ya que representan la unión de varios atributos de una dimensión como una serie de relaciones padre-hijo, es decir, de tipo jerárquico. Por lo tanto, se puede decir que cada elemento en una dimensión se puede resumir utilizando una jerarquía. Por ejemplo: ciudad -> estado -> país -> continente. (Kimball, R. & Ross, M., 2002)

e) Granularidad

La granularidad se refiere al nivel de detalle con el que se almacenan los datos en una tabla de hechos o dimensiones. Existe una estrecha relación entre jerarquías y granularidad, ya que cuando se habla de jerarquías se está definiendo la granularidad de cada dimensión. El nivel superior de la jerarquía tiene una granularidad alta (grano grueso) y el nivel más bajo de la jerarquía cuenta con una granularidad baja (grano fino), donde se encuentran los datos en detalle (Kimball, R. & Ross, M., 2002).

f) Cubo

Son estructuras que permiten analizar rápidamente los datos, en ellos se permite manipular y detallar los datos desde diferentes perspectivas que varían dependiendo de los criterios construidos para su consulta. Principalmente, los cubos están compuestos por hechos, que representan algún evento dentro de la organización, y las dimensiones (Kimball, R. & Ross, M., 2002).

En la figura 13, se presenta un cubo que representa un espacio multidimensional y cada una de las dimensiones está constituida por cada arista o eje del cubo, interceptándose entre ellas.

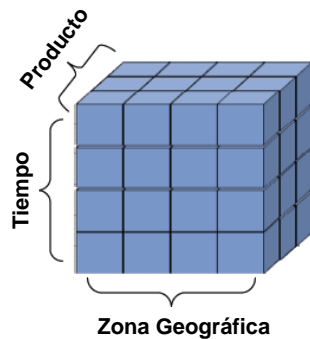


Figura 13. Cubo Representativo de Dimensiones (Wolff, 2002)

Como bien muestra la figura 9, se pueden apreciar tres dimensiones: tiempo, zona geográfica y producto. Cada dimensión presenta una jerarquía, es decir, la dimensión tiempo puede tener la jerarquía: día -> mes -> trimestre -> año. De este cubo se puede obtener información que permita responder a preguntas, dependiendo del punto de vista que se requiera, cruzando las tres dimensiones, como por ejemplo: ¿Qué productos fueron vendidos en Caracas para la fecha 1/1/2012?.

g) Tabla Dimensiones

Las tablas dimensiones son aquellas que se conectan a la tabla de hechos y permiten visualizar la información desde diferentes perspectivas. En ellas se almacenan un conjunto de valores o atributos que están relacionados con una dimensión en particular. Como bien se hablo anteriormente, esta relación viene expresada en jerarquía, debido a que los valores de la tabla están agrupados por niveles de detalle (Wolff, 2002).

Habitualmente, las tablas dimensiones contiene una clave primaria que la identifica, también conocida como clave subrogada, que es la clave que se usa en el almacén para enlazar con la tabla de hechos, además, contiene atributos descriptivos que se utilizan como filtros de las consultas o para agrupar datos bajo ciertos criterios.

h) Tabla de Hechos

Las tablas de hechos son consideradas las tablas centrales dentro de un modelos dimensional, y además, contienen los valores de las medidas del negocio que se desean

analizar. Comúnmente, las tablas de hechos contienen un conjunto de claves foráneas asociadas a las dimensiones, es decir, se expresa una cardinalidad de 'n' a 'n' entre la tabla de hechos y las dimensiones (Wolff, 2002).

Adicionalmente, las tablas de hechos contienen una clave primaria que lo identifica y un conjunto de indicadores, usados para representar los eventos a modelar, es decir, lo que se busca con la tabla de hechos es representar un proceso de negocio. Además, en ella se almacena información transaccional que varía en el tiempo (Kimball, R. & Ross, M., 2002).

Por lo general, cuando se habla de tabla de hechos, se suele hacer énfasis en los tipos de hechos que estas almacenan. Es por esto que el portal de de 1keydata.com (2012), establece que existen tres tipos de hechos:

- **Aditivos:** Son aquellos hechos que pueden ser resumidos a través de cualquiera de las dimensiones en la tabla de hechos. En la figura 14, se muestra un hecho aditivo, denominado "Monto", ya que el mismo se puede agregar usando cualquiera de las tres dimensiones y tomar un valor significativo, como por ejemplo: Fecha-Monto, para saber el monto de ventas en una fecha determinada, Tienda-Monto, para conocer el monto de ventas en una tienda; y Producto-Monto, para saber el monto de ventas de un producto específico.



Figura 14. Tabla de Hechos Aditivos (1keydata.com, 2012)

- **Semi-aditivos:** Son aquellos que pueden resumirse a través de alguna, pero no todas, las dimensiones en la tabla de hechos. En la figura 15, se observa una tabla de hechos, con un hecho semi-aditivo, ya que al unir el Saldo con la dimensión fecha no se produciría un valor significativo, en cambio al juntar Saldo con dimensión Cuenta, se podría conocer el saldo actual de una cuenta.

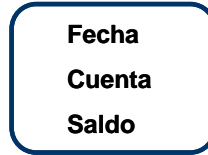


Figura 15. Tabla de Hechos Semi-aditivos (1keydata.com, 2012)

- No aditivos: Son aquellos que no pueden ser resumidos por ninguna de las dimensiones presentes en la tabla de hechos. En la figura 16, se observa el hecho Margen_Beneficio, se considera un hecho no aditivo ya que no tienen sentido al unirlos con alguna de las dos dimensiones.

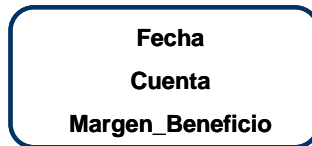


Figura 16. Tabla de Hechos No aditivos (1keydata.com, 2012)

La granularidad es uno de los aspectos más importante al momento de construir una tabla de hechos, ya que permite a los indicadores comparar valores específicos, o bien generalizarlos.

A continuación, en la figura 17, se presenta una tabla de hechos con sus respectivas dimensiones, en donde se observa que partiendo de los datos contenidos en estas dimensiones, se pueden construir indicadores o medidas exigentes en dicha tabla. Por ejemplo, se puede tomar el monto de la dimensión producto y crear un indicador que represente las ganancias de dicho producto. Así, se podría analizar las ganancias de dicho productos en las distintas tiendas.

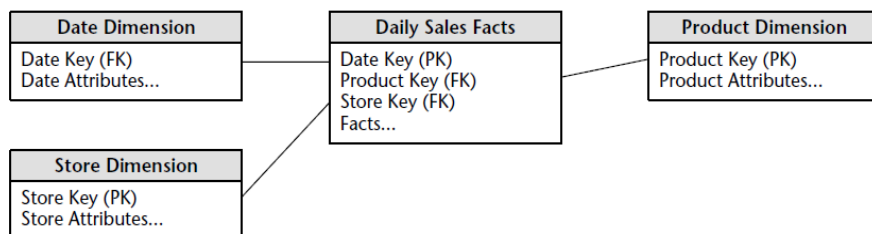


Figura 17. Tabla de Hechos y Dimensiones (Kimball, 1997)

Asimismo, las tablas de hechos, según el portal 1keydata.com (2012), se pueden clasificar en:

- Acumulativas: Son aquellas tablas de hechos que describen lo que ha pasado en un período de tiempo. Por ejemplo, las ganancias obtenidas en una tienda determinada por mes.
- Fotográficas: Son las tablas de hechos que describen el estado de las cosas en un período determinado del tiempo.

i) Formas de Modelado

- Esquema Estrella: Basicamente, el esquema estrella deriva su nombre del hecho que su diagrama forma una estrella, con puntos radiales desde el centro. El centro de la estrella consiste en una o más tablas de hechos y las puntas de la estrella están representadas por las tablas dimensiones (Wolff, 2002).

En el esquema estrella la tabla de hecho es enlazada por cada tabla dimensión por medio de referencia de un clave foránea. En la figura 18, se puede observar un esquema estrella con una tabla de hechos “Ventas” y las tablas dimensiones asociadas como Marcas, Productos, Cliente y Tiempo. Cada una de estas tablas tiene un identificador único, el cual corresponde a la clave foránea en la tabla de hechos “Ventas”.

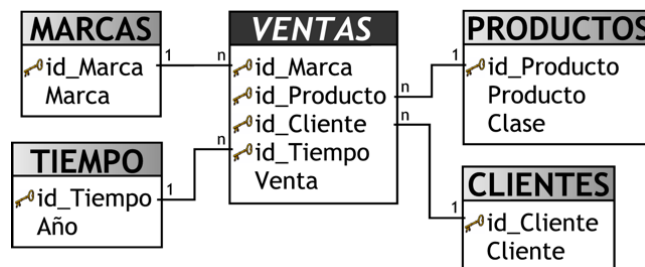


Figura 18. Esquema Estrella (dataprix.com, 2009)

- Esquema Copo de Nieve: La diferencia del esquema copo de nieve con el esquema estrella, se localiza en la estructura de las tablas dimensiones. Las mismas en este

esquema se encuentran normalizadas y contienen solo el nivel que es claver primaria en la tabla y la clave foránea de su padre del nivel más cercano (Wolff, 2002).

Es decir, en este esquema las tablas dimensiones, además de estar enlazadas a la tabla de hechos, tambien se enlazan con otras tablas dimensiones. Así se presenta un copo de nieve. En la figura 19, se muestra como se presenta este esquema, en donde se aprecia la normalización en algunas dimensiones.

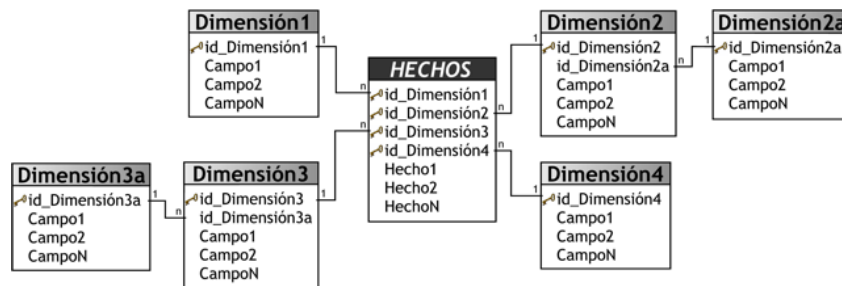


Figura 19. Esquema Copo de Nieve (dataprix.com, 2009)

2.4 Inteligencia de Negocios

2.4.1 Definición

La Inteligencia de Negocios también se conoce como *Business Intelligence* (BI, por sus siglas en inglés). Existe diversidad de definiciones que se le han otorgado a este término. Para el Instituto de Almacenamiento de Datos (2002), la Inteligencia de Negocios es un conjunto de procesos, tecnologías y herramientas necesarias para la transformación de datos en información, de información en conocimiento y el conocimiento en planes que gestionen las acciones rentables de negocio.

Por otra parte, la Inteligencia de Negocios es considerada una arquitectura y colección de operaciones integradas, así como la aplicación al apoyo de la toma de decisiones y bases de datos que proporcionan a la comunidad empresarial, la facilidad de acceso a los datos (Most & Atre, 2003).

Asimismo, Larson (2006) generaliza dicho término, definiéndolo como la entrega de información exacta, útil y necesaria para apoyar en la toma de decisiones de una organización.

No obstante, cuando se habla del término de Inteligencia de Negocios, se debe aclarar que para Williams & Williams (2007), este concepto no es un método ya que se tiene que combinar con tecnologías adecuadas para encontrar soluciones más eficientes y adaptables a los cambios de una organización, no es considerada una tecnología porque las herramientas que usa sólo definen una parte y no todo el concepto y tampoco es un producto.

Por lo tanto, se puede afirmar que la Inteligencia de Negocios, conocida también como Inteligencia Corporativa, Empresarial o Institucional, es un conjunto de partes (métodos, aplicaciones, tecnologías) que se combinan para reunir, procesar y transformar datos, provenientes de sistemas transaccionales, en información organizada y estructurada, permitiendo obtener una potencial ventaja competitiva que proporcione información para la toma de decisiones.

2.4.2 Características

En general, se tiene que una solución de Inteligencia de Negocios, permite (Sánchez, 2003):

- Accesibilidad de los usuarios a la información.
- Compartir información entre los distintos departamentos de una organización.
- Generar y procesar datos para su posterior análisis.
- Análisis multidimensionales.
- Apoyo a la toma de decisiones, seleccionando y manipulando sólo aquellos datos que interesen.
- Generar reportes departamentales o globalizados.
- Realizar pronósticos, basándose en información generada.
- Generar orientación al usuario final.

2.4.3 Arquitectura

Una solución de Inteligencia de Negocios se caracteriza por presentar 3 bloques importantes: las fuentes de datos, los componentes y las herramientas para presentar los datos resumidos o en detalle (NASE, 2010). En la figura 20, se diferencian los tres bloques que contiene una arquitectura de una solución de Inteligencia de Negocios:

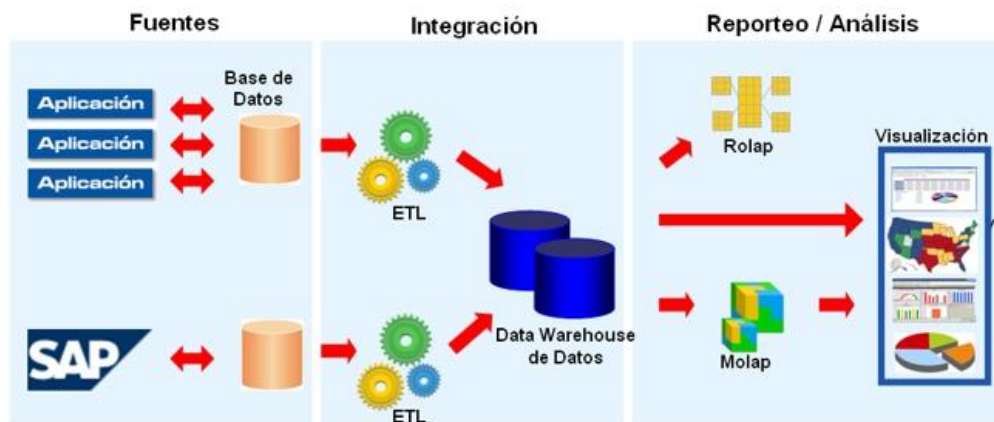


Figura 20. Arquitectura de una solución de Inteligencia de Negocios (NASE, 2010)

En el primer bloque se tienen las fuentes de datos, que están representadas por aquellos sistemas que contienen datos que pueden ser explotados potencialmente por los componentes de una solución de Inteligencia de Negocios. Generalmente, estos datos suelen presentarse en sistemas transaccionales, archivos planos, hojas de cálculos o bodegas de datos de diferentes departamentos de una organización.

El segundo bloque contiene los componentes que integran los datos obtenidos de las fuentes, ya sean internas o externas. Estos datos pasan por un proceso en el cual se filtran, se estandarizan y se cargan en almacenes o bodegas. A este proceso se le denomina Extracción, Transformación y Carga (ETL, por sus siglas en inglés). Estos almacenes son los componentes de una solución BI, ya que se encargan de garantizar la integridad de los datos contenidos en ellos.

Por último, se tiene el tercer bloque, donde se encuentran las herramientas que se utilizan para representar los datos contenidos en el almacén o bodega de datos. Estos datos pueden representarse en forma de consultas, reportes, análisis, cuadros de mando, tablas dinámicas, etc.

Estas herramientas son indispensables para mostrar los datos relevantes, resumidos o en detalle, de una organización o empresa. Asimismo, se garantiza facilidad al momento de realizar los análisis y reportes, con el fin de agilizar el proceso de toma de decisiones.

Por lo tanto, se puede decir que toda solución de Inteligencia de Negocios permite agrupar datos de diferentes fuentes en un repositorio centralizado, con el fin de representar los datos más relevantes de manera resumida o detallada.

2.4.4 Pirámide de Información

Cuando se habla de pirámide información, se refiere a la representación, mediante un diagrama, de los niveles de información de una organización o empresa, en donde la información contenida en cada nivel se va haciendo más elaborada a medida que se asciende. En la figura 21, se puede observar la pirámide de información con los dos bloques que contiene, uno relacionado a los sistemas de procesamiento transaccional y otro relacionado a los sistemas de procesamiento analítico.

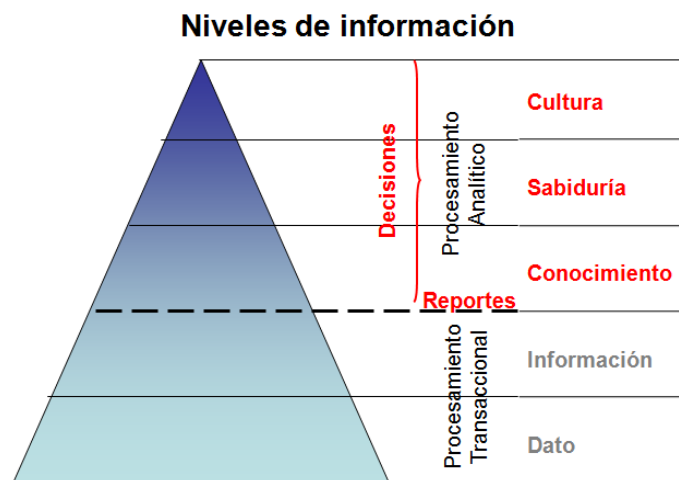


Figura 21. Pirámide de la Información (Adaptación Tian Consultores)

El primer bloque contiene los sistemas de procesamiento transaccional, éstos se refieren a todos los sistemas que representan operaciones del día a día de una organización, en donde se almacena una gran cantidad de datos, se procesan y se transforman en información. Luego con la información obtenida se generan reportes o análisis que representan conocimiento. No obstante,

estos reportes no generan solución a preguntas complejas, ya que los sistemas transaccionales no están diseñados para responder complejidades (De Pablos, Albarrán & Castilla, 2001).

En el segundo bloque se encuentran los sistemas analíticos que obtienen conocimiento básico de los sistemas transaccionales. Estos sistemas se encargan de extraer información de los transaccionales, que ya constituye un conocimiento, y representarla tanto en reportes como en indicadores, gráficos, tablas dinámicas, etc. (De Pablos, Albarrán & Castilla, 2001).

Adicionalmente, a este bloque se le incorpora un nivel de sabiduría, donde se integran todos los conocimientos obtenidos. Poco a poco, estos conocimientos se van almacenando y con el uso se convierten en sabiduría, es decir, un tipo de información que apoya a la gerencia media en la toma de decisiones determinantes. Asimismo, este tipo de información también permite realizar predicciones.

La cultura es otro nivel que se suma a este bloque, ya que representa la integración de todos los niveles anteriores, es decir, la cultura es la manera en la que se hacen las cosas. Por lo tanto, en una empresa, la cultura permite tener una visión global del negocio, llevando a tener un amplio criterio y capacidad de juicio para tomar decisiones acertadas.

2.4.5 Componentes

Al hablar de Inteligencia de Negocios, se debe mencionar las herramientas que serán utilizadas para facilitar la visualización de la información obtenida por sus componentes, es decir, representar la información más relevante, que ya ha sido previamente procesada por los almacenes o bodegas de datos (Cortes, 2007). Para presentar dichos datos, existen varias herramientas, tales como:

a) Reportes:

Son herramientas utilizadas para la elaboración de informes, listados y tablas de información, resumida o en detalle, partiendo de la información de los almacenes o bodegas de datos (Cortes, 2007). Adicionalmente, presentan un formato predefinido y tienen la finalidad de dar a conocer, a la alta gerencia de una organización o empresa, el

estado de la misma, proporcionándoles distintas opciones a la hora de tomar decisiones (Hill, 2008).

La principal ventaja que ofrece el uso de reportes, es la forma dinámica en que éstos pueden ser contruidos, utilizando mayor interés en casos específicos y evitando mostrar todos los datos en un solo reporte. Por lo tanto, los datos se pueden mostrar resumidos o en detalle, facilitando a la alta gerencia a tomar decisiones rápidas y acertadas (Hill, 2008).

b) Diagramas o Gráficos

Los diagramas son representaciones gráficas de los datos y se pueden simbolizar mediante barras en gráficos de barra, líneas en gráficos lineales o pedazos en gráficos de torta.

Generalmente, los diagramas representan valores numéricos, funciones o algún tipo de estructura cualitativa, que facilitan el entendimiento y análisis de grandes cantidades de datos y sus relaciones entre las partes que componen dichos datos (Jensen & Anderson, 1992).

c) Indicadores

Un indicador es considerado una variable o una magnitud que mide de forma cuantitativa o cualitativa un suceso u objetivo relacionado a un evento en particular, ayudando a determinar los cambios ocurridos (Ateico, 2007).

Por otra parte, un indicador es la medida de la condición de un proceso o evento en un momento determinado, proporcionando un panorama de la situación de un proceso, de un negocio, etc. Además, son una forma clave de retroalimentar un proceso, de monitorear el avance o la ejecución de un proyecto y de los planes estratégicos (Pérez, 2006).

d) Cuadros de Mando (Dashboards)

Los Cuadros de Mando son herramientas utilizadas por la alta gerencia de una organización, para representar indicadores numéricos con la finalidad de monitorear y analizar los resultados obtenidos a lo largo de un período. Principalmente, los cuadros de mando se enfocan en el seguimiento y monitoreo de indicadores (Ateico, 2007).

Además, son elaborados a partir de bodegas de datos, reportes, resúmenes e indicadores claves para la gestión, permitiendo a la alta gerencia de la organización, analizar los resultados de la misma de forma rápida y eficaz (Cortes, 2007).

Con el uso de los cuadros de mandos se logra:

- Facilitar la planificación y control de resultados.
- Disponer eficientemente de información indispensable y significativa.

e) Cuadro de Mando Integral

Este tipo de cuadro de mando parte de que la estrategia de la empresa es el punto de referencia para todo el proceso de gestión interno. Con él, los diferentes niveles de dirección y gestión de la organización disponen de una visión de la estrategia de la empresa traducida en un conjunto de objetivos, iniciativas e indicadores de evolución (Cortes, 2007).

Es un tipo de Cuadro de Mando más sofisticado, ya que enlaza distintos procesos y están orientados, principalmente, a la alta gerencia. Los más utilizados se basan en los métodos de Kaplan y Norton, como se observa en la figura 22, y permiten a una organización trasladar su visión y estrategia a la implementación (Kaplan & Norton, 2009).

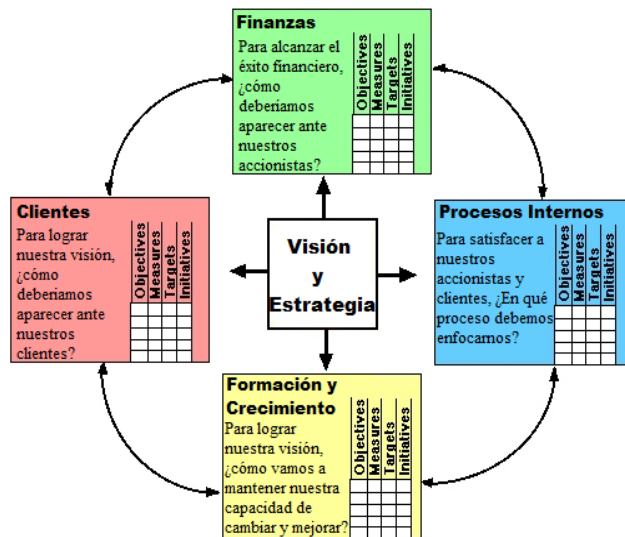


Figura 22. Esquema de Cuadros de Mando Integrales (Kaplan & Norton, 2009)

Los objetivos estratégicos se asocian mediante relaciones causa-efecto y se organizan en cuatro áreas o perspectivas (Kaplan & Norton, 2009):

- Financiera: Esta perspectiva se enfoca en los objetivos más importantes de las organizaciones que buscan ganancias. Además, incorpora la visión y las necesidades de los accionistas.
- Cliente: Esta perspectiva se relaciona con las necesidades de los consumidores, generando una relación implícita entre esta perspectiva y la financiera, ya que a más clientes, más ganancia.
- Proceso Interno: Esta perspectiva mide los indicadores basados en procesos internos de la organización que se relacionan con la satisfacción del cliente y el rendimiento financiero.
- Formación o Desarrollo: Esta perspectiva es difícil medir ya que se desarrollan pocos indicadores en esta área. Sin embargo, lo que se obtiene de estos índices es importante ya que se establece un camino de enfoque.

2.4.6 Ventajas

Los sistemas de Inteligencia de Negocios traen consigo gran cantidad de beneficios para el desarrollo de una organización o empresa. Partiendo de distintas fuentes, tales como NASE (2010),

Sánchez (2003), Williams & Williams (2007), Larson (2006) y Most & Atre (2003) se puede decir que las principales ventajas de los sistemas BI son:

- Capacidad de extraer, depurar, consolidar, sintetizar y presentar los datos en forma automatizada.
- Disponer de información necesaria para el análisis y la toma de decisiones.
- Apoyo a la alta gerencia a planear y obtener pronósticos.
- Mejora la colaboración y calidad de las decisiones, dado que la información se comparte en los distintos niveles de la organización.
- Proporciona una visión más profunda del negocio.
- Optimiza el rendimiento de los sistemas.
- Realiza seguimiento de los planes estratégicos.
- Obtiene el valor real de las aplicaciones de gestión.

Sin embargo, no sólo ofrece beneficios a la organización, sino que también permite presentar diferentes tipos de información (Rasmussen, Goldy & Solli, 2002), tales como:

- Comparativa, que permite realizar comparaciones en distintas fechas.
- De tendencias, permitiendo conocer lo más demandado.
- Veraz, conociendo si se cumplen los objetivos propuestos.
- Rentable, que permite distinguir en la continuación de la inversión.

2.5 Indicadores de Gestión

Una vez conocida la definición de un indicador, Beltran (2000) establece que los indicadores de gestión son relaciones entre variables cuantitativas o cualitativas, que permiten observar situaciones y tendencias generadas en el objetivo o fenómeno observado, respecto a objetivos y metas previstos; son factores para establecer el logro y el cumplimiento de objetivos y metas de un determinado proceso; además representan información importante no son sólo datos.

2.5.1 Atributos

En la tabla 3 se presentan los atributos que debe tener un indicador de gestión, según Beltran (2000).

Tabla 3. Atributos de un Indicador de Gestión (Beltran, 2000)

Atributos	Especificación
Exactitud	La información debe presentar la situación o el estado como realmente es.
Forma	La información puede ser cuantitativa o cualitativa, numérica o gráfica, impresa o visualizada, resumida o detallada. La forma debe ser elegida según la situación y necesidades de quien la recibe y procesa.
Frecuencia	Es la medida de cuán a menudo se requiere, se recaba, se produce o se analiza.
Extensión	Se refiere al alcance en términos de cobertura del área de interés. Además tiene que ver con la brevedad requerida, según el tópico..
Origen	Puede originarse dentro o fuera de la organización. Lo fundamental es que la fuente que la genera sea la fuente correcta.
Temporalidad	La información puede hablar del pasado, de los sucesos actuales o sucesos futuros.
Relevancia	La información es relevante si es necesaria para una situación en particular.
Integridad	Una información completa proporciona al usuario el panorama integral de lo que necesita saber acerca de una situación determinada.
Oportunidad	Debe estar disponible y actualizada cuando se le necesite.

2.5.2 Componentes

Para Beltran (2000) un indicador correcto y válido debe estar compuesto por:

- **Nombre:** La identificación es vital, su nombre debe ser concreto y definir claramente su objetivo y utilidad. Ejemplo: Eficacia Mensual en Ventas

- Forma de Cálculo: Cuando son cuantitativos se debe tener muy claro la fórmula matemática para su cálculo, lo cual implica la identificación exacta de todos los factores que lo afectan y su relación. Ejemplo: Eficacia Mensual en Ventas = (Total de clientes que compraron / Total de clientes que visitaron)*100
- Unidad: La manera como se expresa el valor de determinado indicador está dada por las unidades, las cuales varían de acuerdo con los factores que se relacionan. Ejemplo: Porcentaje (%)
- Glosario: Es fundamental que el indicador se encuentre documentado en términos de especificar de manera precisa los factores que se relacionan en su cálculo.

2.5.3 Clasificación

Beltran (2000) establece distintos enfoques para clasificar los indicadores de gestión. En cuanto a su naturaleza se refiere, los indicadores se clasifican de acuerdo a los factores claves de éxito. Así, se encuentra los indicadores de eficacia que miden el grado de cumplimiento de las metas (resultados, calidad y satisfacción del cliente), de eficiencia que miden el nivel de ejecución de un proceso (tiempos de procesos, costos operativos y desperdicios), de efectividad que relacionan el cumplimiento de las metas con los costos y de productividad. En la figura 23 se muestra la relación que existe entre los factores claves mencionados, y así se establece la clasificación de los indicadores.

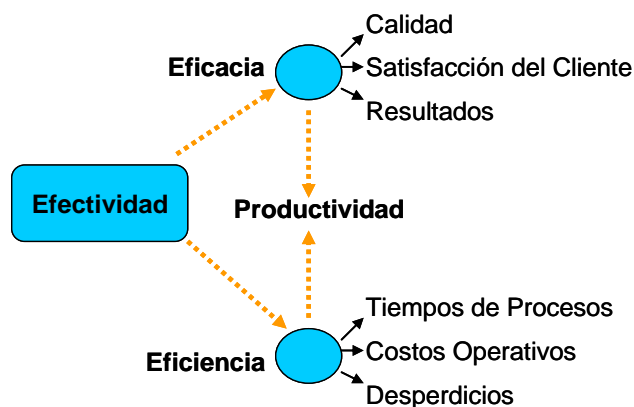


Figura 23. Mapa de Factores Claves de éxito de la Gestión (Beltran, 2000)

Por otra parte, depende de su vigencia, los indicadores de gestión se clasifican en (Beltran, 2000):

- **Temporales:** Cuando su validez tiene un lapso finito. Por lo general, se asocian al logro de un objetivo a la ejecución de un proyecto, al lograrse el objetivo o cuando éste pierde interés para la organización, los indicadores asociados deberán desaparecer.
- **Permanentes:** Son indicadores que se asocian a variables o factores que están presentes siempre en la organización y se asocian por lo regular a los procesos.

Además, también se clasifican dependiendo del nivel de la organización en estratégico, táctico u operativo, donde se recoge la información y se consolida el indicador; y dependiendo del nivel de la organización donde se utiliza el indicador como insumo para la toma de decisiones (Beltran, 2000).

2.5.4 Ventajas

Para Beltran (2000), la ventaja principal que trae consigo el uso de indicadores se resume en la reducción drástica de la incertidumbre, de la angustia y de la subjetividad, con el consecuente incremento de la efectividad de la organización y el bienestar de todos los trabajadores.

Sin embargo Pérez (2006) establece un listado de beneficios que puede proporcionar a la organización la implementación de un sistema de indicadores de gestión:

- **Satisfacción del Cliente:** En la medida en que la satisfacción del cliente sea una prioridad para la empresa, así lo comunicará a su personal y enlazará las estrategias con los indicadores de gestión, de manera que el personal se dirija en dicho sentido y sean logrados los resultados deseados.
- **Monitoreo del Proceso:** El mejoramiento continuo sólo es posible si se hace un seguimiento exhaustivo a cada eslabón de la cadena que conforma el proceso. Las mediciones son las herramientas básicas no sólo para detectar las oportunidades de mejora, sino además para implementar las acciones.

- **Evaluación Comparativa:** Si una organización pretende mejorar sus procesos, una buena alternativa es traspasar sus fronteras y conocer el entorno para aprender e implementar lo aprendido. Una forma de lograrlo es a través del benchmarking para evaluar productos, procesos y actividades y compararlos con los de otra empresa.
- **Gerencia del Cambio:** Un adecuado sistema de medición le permite a las personas conocer su aporte en las metas organizacionales y cuáles son los resultados que soportan la afirmación de que lo está realizando bien.

2.5.5 Metodología General

Según Beltran (2000), para establecer indicadores de gestión a cualquier nivel es de vital importancia tener claro qué es lo correcto y cómo hacerlo correctamente. Lo correcto se asimila a los resultados específicos que se esperan, las características específicas de esos resultados, y esto representaría el conjunto de factores clave para lograr la eficacia. En cuanto a cómo sería hacer las cosas correctamente significa tener presente la capacidad de gestión actual, la secuencia de pasos que nos llevarán a lograr los resultados y los recursos disponibles teniendo en cuenta factores óptimos de aprovechamiento. Esto constituiría el conjunto de factores claves para lograr la eficiencia.

No obstante, en la figura 24 se muestran los nueve pasos genéricos de la metodología general para el establecimiento de indicadores de gestión:



Figura 24. Metodología general para establecimiento de Indicadores de Gestión (Beltran, 2000)

- **PASO 1. Contar con objetivos y estrategias:** Es fundamental contar con objetivos claros, precisos, cuantificados y tener establecidas las estrategias que se utilizarán para lograr los objetivos, ya que ellos dan el punto de llegada y las características del resultado que se espera.
- **PASO 2. Identificar factores claves de éxito:** Se entiende por factor de éxito como el aspecto que es necesario mantener bajo control para lograr el éxito de la gestión. Éstos permiten realizar un control integral de la gestión tanto en la medida de orientación de resultados, como en la de lograr los mismos.
- **PASO 3. Definir indicadores para factores claves:** Una vez identificados los factores claves de éxito, es necesario establecer unos indicadores que permitan hacer el monitoreo antes del proyecto, durante éste y después de la ejecución del proceso respectivo.
- **PASO 4. Determinar estado, umbral y rango de gestión:** El término estado, se refiere al valor inicial o actual del indicador. El término umbral corresponde al valor del indicador que se desee lograr o mantener. Y por último, el rango de gestión es el espacio comprendido entre los valores mínimo y máximo que el indicador puede tomar.

En esta fase juega un papel importante la alerta estratégica. Se entiende por alerta estratégica, el mecanismo que permite identificar niveles de prevención o alarma, dependiendo del valor en el que se encuentre un indicador. Generalmente, la alerta estratégica cuenta con tres colores: el verde, para cuando el valor se encuentre en lo esperado, el amarillo, para cuando el valor esté acercándose al umbral y el rojo, para cuando el valor sobrepase el umbral e indique un nivel de alarma.

- **PASO 5. Diseñar la medición:** Consiste en determinar fuentes de información, frecuencia de medición, presentación de la información, asignación de responsables de la recolección, tabulación, análisis y presentación de la información.
- **PASO 6. Determinar y asignar recursos:** Basado en las características del punto anterior, para la medición se establecen las necesidades de recursos que demanda la realización de las mediciones.

- **PASO 7. Medir y ajustar:** Generalmente, la primera vez que se realiza la medición de un sistema de indicadores de gestión, surgen una serie de factores que es necesario ajustar o cambiar en los siguientes sentidos:
 - Pertinencia del indicador.
 - Valores y rangos establecidos.
 - Fuentes de información seleccionadas.
 - Proceso de toma y presentación de la información.
 - Frecuencia en la toma de información.
 - Destinatario de la información.
- **PASO 8. Estandarizar y formalizar:** Consiste en el proceso de especificación completa, documentación, divulgación e inclusión entre los sistemas de operación del negocio de los indicadores de gestión.
- **PASO 9. Mantener y mejorar continuamente:** Realizar mantenimiento general al sistema es darle continuidad operativa y efectuar los ajustes que se deriven del permanente monitoreo del sistema empresarial y de su entorno. Asimismo, es necesario mejorar continuamente, incrementando el valor que el sistema de indicadores de gestión agrega a las personas usuarias.

2.6 Plataformas de Inteligencia de Negocios Software Libre

Para realizar una solución de Inteligencia de Negocios, existen diferentes plataformas: las de software propietario y las de software libre. Las primeras son aquellas que tienen limitaciones para usarlo, modificarlo o redistribuirlo y por lo tanto pueden incluir algunos costos para su uso. Por el contrario, las de software libre son aquellas que ofrecen una libertad a los usuarios para su uso, su estudio, su modificación y su redistribución. En el presente T.E.G. se utilizará una plataforma de software libre.

A continuación, en la tabla 4, se presenta un cuadro comparativo entre tres plataformas software libre, que permiten desarrollar un sistema de Inteligencia de Negocios.

Tabla 4. Cuadro Comparativo entre Plataformas BI Software Libre analizadas y estudiadas (Stratebi, 2010)

Características	Pentaho	Jasper	Actuate - BIRT
ETL	Buena	Buena	No tiene
Aplicación Web	Buena	Buena	No tiene
Informes	Buena	Buena	Buena
Informes Predefinidos	Buena	Buena	Buena
Motor OLAP	Buena	Buena	Regular
Visor OLAP	Buena	Regular	Buena
Cuadros de Mando	Buena	No tiene	Regular

Como bien se puede observar, la plataforma con las características mejores adaptadas al desarrollo de este Trabajo Especial de Grado es *Pentaho Business Intelligence*, ya que cuenta con una gama de soluciones para desarrollar un sistema de Inteligencia de Negocios con facilidades de uso, intuitivas, flexibles, de fácil acceso, entre otras.

Por otra parte, se tiene *Jasper* que presenta debilidad en cuanto a la realización de cuadros de mando, que no se encuentra disponible en la versión Software Libre. Sin embargo, la versión Enterprise si dispone de una solución para desarrollar Cuadros de Mandos.

Por último, se observa *Actuate-BIRT*, quien en su versión software libre no cuenta con una herramienta para realizar el proceso de ETL, ni con una aplicación Web. Y presenta debilidad en la realización de cuadros de mandos, ya que aunque el servidor no contempla cuadros de mandos, se pueden hacer 100% funcionales con los reportes BIRT.

2.6.1 Plataforma Pentaho Business Intelligence (Pentaho, 2005)

Esta plataforma de Inteligencia de Negocios ofrece una infraestructura de herramientas de análisis e informes integrados, con un motor de *workflow* de procesos de negocio. La plataforma será capaz de ejecutar las reglas de negocio necesarias, expresadas en forma de procesos y actividades, así como también de presentar y entregar la información específica en el momento

adecuado. Su modelo de ingresos parece estar orientado a los servicios de soporte, formación y consultoría.

En la Web se presenta una organización por productos: Reportes, Análisis, Cuadros de Mando, Minería de datos y se hace mención específica al *workflow* como una de las capacidades BI claves de la plataforma.

2.6.1.1 Características

- Es multiplataforma, tanto a nivel de cliente como a nivel del servidor no hay dependencia, utiliza lenguaje interpretado.
- Está construido en torno al servidor de aplicaciones *J2EE JBoss* y *Jboss Portal*, *Tomcat* y es portable en cualquier servidor de aplicaciones como *glassfish*, habilitando que toda la información sea accesible mediante un browser en la intranet de la empresa.
- *Pentaho* presenta informes en los formatos habituales (*Html*, Excel, PDF, entre otros) mediante *JfreeReport* u otras plataformas como *BIRT* o *JasperReports*. Para la generación de PDFs utilizan el conocido Apache FOP. Asimismo incorpora la librería *JPivot*, la cual permite visualizar tablas OLAP a través de un browser y realizar las operaciones típicas de análisis OLAP.
- Se cuenta con áreas de reporte llamado *Pentaho Reporting* en el cual existe tres productos con diferentes enfoques y dirigidos a diferentes tipos de usuarios:
 - *Pentaho Report Designer*: Editor basado en eclipse con capacidad de personalización de informes a las necesidades de negocio destinado a desarrolladores, está incluye asistente para facilitar las configuración de propiedades.
 - *Pentaho Report Design Wizard*: Herramienta de diseño de informe que facilita el trabajo y permite a los usuarios obtener resultados de forma inmediata. Está destinada a usuarios con pocos conocimientos técnicos.
 - *Web ad-hoc reporting*: Es similar a la herramienta anterior pero vía Web. Extiende la capacidad

- Para generar gráficos se de los usuarios finales para la creación de informes a partir de plantillas pre configuradas y siguiendo un asistente de creación. Apoyan en *JFreeChart*, una librería para generar los gráficos más comunes (2D, 3D, barras, líneas series temporales, entre otros), interfaces para acceder a diferentes fuentes de datos, exportación a PNG, JPEG y PDF, soporte para *servlets*, *JSPs*, *applets* y aplicaciones clientes.
- Los Cuadros de Mandos (*Dashboards*) son un desarrollo propio de *Pentaho*, recogen información de todos los componentes de la plataforma incluyendo aplicaciones externas, *feeds RSS* y páginas Web. Incluyen gestión y filtrado del contenido, seguridad basada en roles y *drill down*. Pueden ser integrados en terceras aplicaciones, en portales o dentro de la plataforma *Pentaho*. En la figura 25 se puede observar el portal de la plataforma BI *Pentaho* usando *dashboards*.



Figura 25. Plataforma Business Intelligence utilizando Dashboard (Pentaho, 2005)

- Los servicios Web constituye una característica esencial de *Pentaho*. Las acciones, que son las tareas más sencillas que constituyen una solución de *Pentaho*, pueden publicarse como *WebServices*. Se utiliza como motor de *WebServices Apache Axis*, quedando los servicios descritos en el lenguaje de definición de servicios Web WSDL.
- Proporciona amplia capacidad de preparación de informes de clases y muy flexibles, desde simples informes en un sitio Web de gran formato a los informes para aplicaciones, como la presentación de informes financieros.
- Capacidad de conectarse directamente a las fuentes de datos o para conectar a través de una capa de metadatos centralizados. Se cuenta con una amplia gama de formatos

incluyendo *Adobe PDF, HTML, Microsoft Excel*, RTF o texto plano, informes con gran contenido que incluyen tablas, gráficos, plantillas, galerías, formato para cabeceras, pies de página, entre otros.

- Es compatible con amplia fuente de datos, compatible con conectividad de base de datos de Java (JDBC) 2.0 o posterior, soporta bases de datos propietarias incluyendo Oracle, IBM, DB2, *Microsoft SQL Server*, NCR entre otras. Soporta populares bases de datos de código abierto incluyendo *MySQL, PostgreSQL, EnterpriseDB*, entre otras.
- Se cuenta con el *Pentaho Data Integration* que permite una poderosa extracción, transformación y carga (ETL), ya que muchas organizaciones tienen información disponible en aplicaciones y base de datos separados. *Pentaho Data Integration* abre, limpia e integra esta información y la pone en manos del usuario, provee una sola versión de todos los recursos de información.

2.6.1.2 Arquitectura

La plataforma de *Business Intelligence Pentaho* cuenta con la siguiente arquitectura:

- Plataforma cien por ciento *J2EE* asegurando la escalabilidad, integración y portabilidad.
- Puede correr en servidores compatibles con *J2EE* como *JBoss AS, WebSphere, Tomcat, WebLogic* y *Oracle AS*.
- Base de datos, vía *JDBC, IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR Teradata, Firebird*, etc.
- No hay dependencia del sistema operativo, utiliza lenguaje interpretado Java.
- Lenguajes de programación *Java, Javascript, JSP, XSL (XSLT/XPath/XSL-FO)*.
- Interfaz de desarrollo basado en *Java SWT, Eclipse, Web-based*.
- Repositorio de datos basado en XML.
- Todos los componentes están expuestos vía *Web Services* para facilitar la integración con arquitecturas orientadas a servicios (SOA).

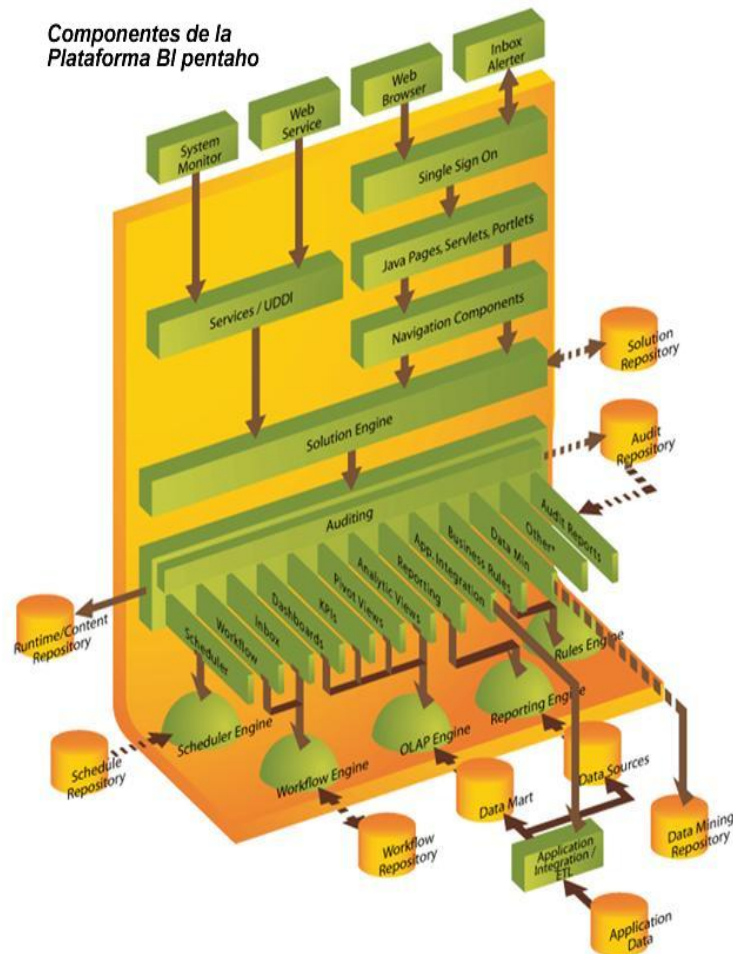


Figura 26. Arquitectura de Pentaho Business Intelligence (Pentaho, 2005)

En la figura 26, se muestra la interacción de los componentes y la arquitectura de *Pentaho Business Intelligence*. El servidor de *Pentaho* contiene los motores y componentes para la presentación de informes, análisis, reglas de negocio, correo electrónico, notificaciones de escritorio y flujo de trabajo. Estos componentes se integran en conjunto a fin de que puedan ser utilizados para resolver un problema de Inteligencia de Negocio. Dentro de una solución de la conducta, la interoperabilidad, y la interacción del usuario de cada subsistema se define por una colección de documentos de definición de soluciones y estos son documentos XML que contienen definiciones de las actividades que se ejecutan como parte de los procesos, en la demanda o llamado por los servicios Web. Estas actividades incluyen las definiciones de: fuentes de datos, consultas, plantillas de informes, las normas de notificación, las normas de negocios, tableros de control, análisis y puntos de vista.

El servidor de *Pentaho* incluye la infraestructura que proporciona la administración avanzada del sistema. Esto incluye la supervisión del sistema SMNP (*Simple Network Management Protocol*), los servicios, informes de uso, soporte de servicios Web, herramientas de validación de configuración y herramientas de diagnóstico, también incluye una herramienta para la Integración de Aplicaciones Empresariales (EAI, por sus siglas en inglés) y ETL.

2.6.1.3 Componentes

La plataforma *Pentaho Business Intelligence Open Source* incluye todos aquellos componentes, que se pueden encontrar en las soluciones de Inteligencia de Negocios propietarias más avanzadas, dichos componentes son: reportes, análisis, cuadros de mandos, flujo de trabajo, minería de dato, integración de datos, sistema de autenticación para acceder a varios sistemas, auditoría de uso y rendimiento, planificador, notificador, seguridad y perfiles.

A continuación, en la figura 27, se muestra la estructura de los diferentes módulos que forman la plataforma de *Pentaho Business Intelligence Open Souce*.

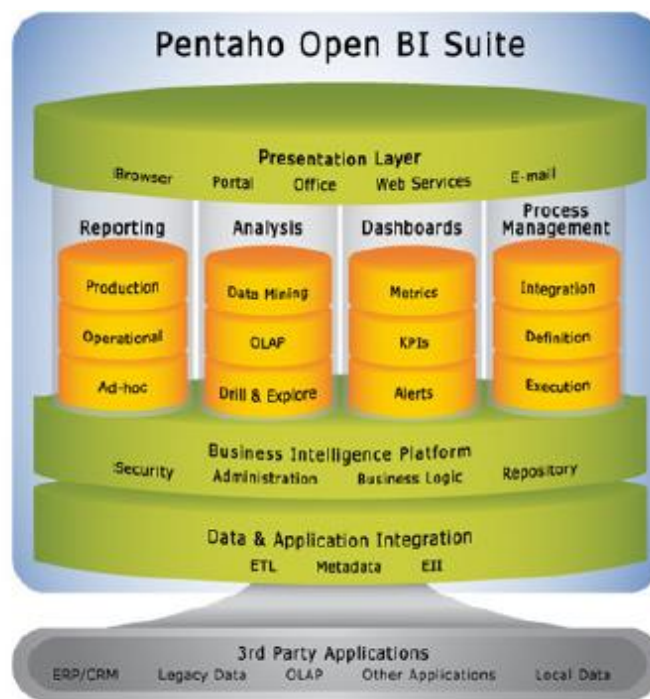


Figura 27. Componentes de Pentaho Business Intelligence (Pentaho, 2005)

a) Reporte (Reporting)

Pentaho Reporting es una solución basada en el proyecto *JFreeReport* que permite generar informes ágiles y de gran capacidad. Éste permite la distribución de los resultados del análisis en múltiples formatos, todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. *Pentaho Reporting* permite también la programación de tareas y ejecución automática de informes con una determinada periodicidad. Soporta amplia fuente de datos relacionales, OLAP y XML basados en fuentes de datos.

b) Análisis (Analysis)

Pentaho Analysis suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (*pivot tables, crosstabs*) generadas por *Mondrian* y *JPivot*, el usuario puede navegar por los datos ajustando la visión de los datos y los filtros de visualización añadiendo o quitando los campos de agregación. Los datos pueden ser representados en una forma de SVG o Flash, cuadros de mando o también integrados con los sistemas de minería de datos y los portales Web (*portlets*). Además, con el Microsoft Excel Analysis Services se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP server *Mondrian*).

c) Cuadros de Mandos (Dashboards)

Todos los componentes del módulo *Pentaho Reporting* y *Pentaho Analysis* pueden formar parte de un cuadro de mando. En *Pentaho Dashboards* es muy fácil incorporar una gran variedad en tipos de gráficos, tablas y velocímetros e integrarlos con los *Portlets JSP*, en donde podrá visualizar informes, gráficos y análisis OLAP.

d) Minería de Datos (Data Mining)

Pentaho incorpora la tecnología *Weka*, es un conjunto integral de herramientas para aprendizaje automático y minería de datos. Su amplio conjunto de clasificación, regresión, reglas de asociación y algoritmos de agrupación se puede utilizar para ayudarle a

comprender mejor la empresa y también ser aprovechado para mejorar el rendimiento futuro a través de análisis predictivo. Como principal características tenemos:

- Motor de *Data Mining* provee de un set de algoritmos de aprendizaje del proyecto *Weka* incluyendo *clustering*, segmentación, árboles de decisión, redes neuronales y análisis de componentes principales, entre otros.
- Integración de *Pentaho Data Integration*, automatización de procesos de transformación de datos al formato que requiere el motor de *Data Mining*.
- Los algoritmos pueden ser aplicados directamente a un set de datos o invocados desde código Java
- Provee filtros para normalización, re-muestreo, selección de atributos, transformación y combinación de atributos
- Modelos para predicción de variables nominales o numéricas
- Herramientas gráficas de diseño para pre-procesamiento de datos, regresión, *clustering*, reglas de asociación

e) Integración de Datos

Pentaho Data Integration (PDI, también llamada *Kettle*) es el componente de *Pentaho* responsable de la extracción, transformación y carga (ETL) de procesos. Aunque las herramientas ETL son los más usados en entornos de almacenes de datos, *Kettle* también se puede utilizar para otros fines:

- Migración de datos entre las aplicaciones o bases de datos.
- Exportación de datos desde bases de datos para archivos planos.
- Carga de datos de forma masiva en bases de datos.
- Limpieza de datos.

Como principal características se tiene:

- Cada proceso es creado con una herramienta gráfica donde se especifica qué se va hacer sin necesidad de escribir código que indique cómo hacerlo.

- Admite una amplia gama de formatos de entrada y salida, incluyendo archivos de texto, hojas de datos, archivos XML, propiedades de *Java* y los motores de base de datos software libre y propietarios.
- Arquitectura extensible permite desarrollar conectores y *plugins* fácilmente
- Basado en repositorio facilita la reutilización de componentes de transformación, colaboración y administración de modelos, conexiones, etc.
- Depurador integrado
- Todos los pasos de base de datos existe un botón SQL que generará el código requerido para las tablas y/o índices. También está disponible a nivel de transformación y trabajo.

En la figura 24, se observa la arquitectura de este componente, en donde se toman los datos desde las diferentes fuentes, sean externas o internas, son transformadas y estandarizadas para ser cargadas en el almacén de datos. Es esta es la tarea principal del *Pentaho Data Integration*.

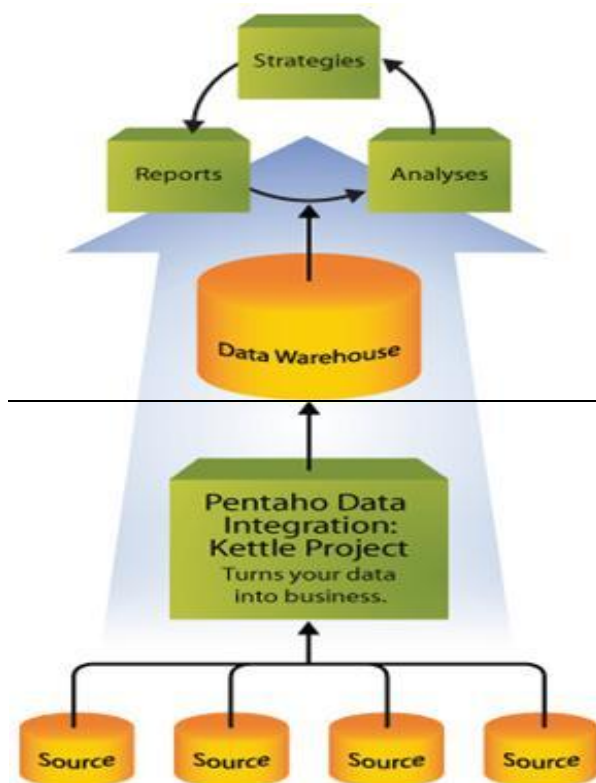


Figura 28. Arquitectura de Pentaho Data Integration (Pentaho, 2005)

f) El *workflow* de procesos de negocio

Los fundamentos del *workflow* de procesos de negocio son: el motor de workflow Enhydra Shark y el estándar WPDL, ayudado por la *Workflow Management Coalition* (WFMC, por sus siglas), organismo que declara tener más de 300 empresas asociadas incluyendo a IBM, Oracle, BEA, Adobe, SAP, TIBCO o SUN, entre otros.

2.7 Método de Desarrollo de Aplicaciones (Fuente: Tian Consultores C. A.)

Para el desarrollo de este Trabajo Especial de Grado, bajo la solución de un sistema de Inteligencia de Negocios, se tomará como base un método que pretende ser de rápida implementación, y además, provee gran integración de datos provenientes de las diferentes fuentes. Este método consta de una serie de fases cíclicas, lo que conlleva a que se puedan hacer iteraciones entre las fases, sin necesidad de regresar a la primera. En la figura 29, se observan cada una de las fases contempladas en este método.

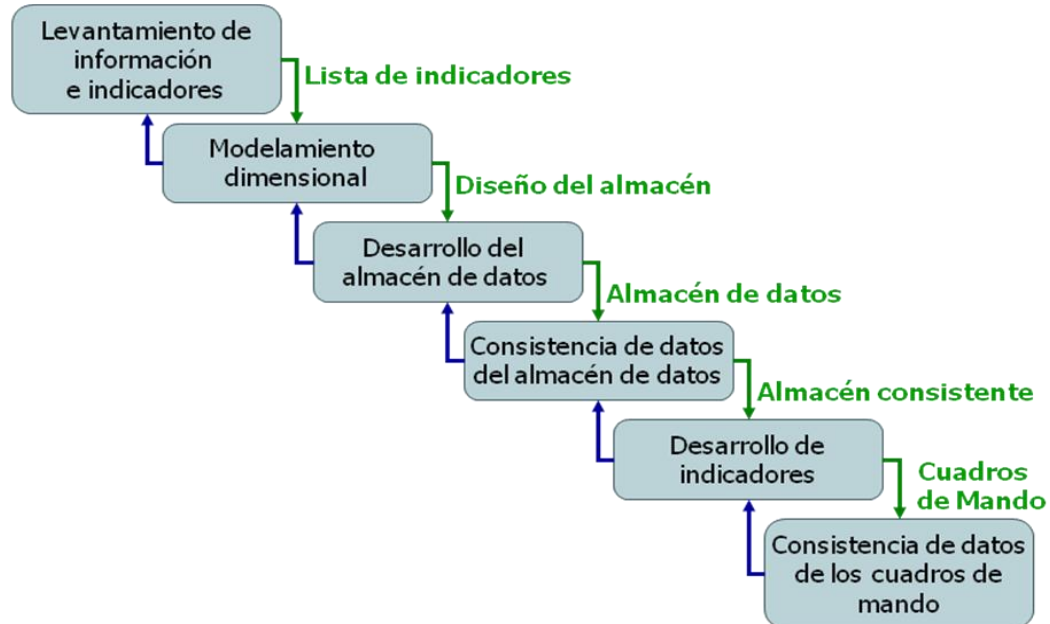


Figura 29. Fases del Método de Desarrollo de la Aplicación (Fuente: Tian Consultores)

Se debe mencionar que en la primera parte se identifica la implementación que tendrá el desarrollo de la solución. Existen dos posibles implementaciones que puede tener el desarrollo de

una solución, donde la primera va orientada a departamentos dentro de una organización, lo cual implica que se parte de los requerimientos del negocio para construir la solución requerida o puede que la solución vaya orientada a toda la organización en general, donde primero se construyen el almacén y las posibles bodegas de datos existente y luego se modelan los requerimientos definidos por los usuarios finales.

De esta manera, el desarrollo de una solución de Inteligencia de Negocios empleando este método garantiza que la implementación que se identifique vaya acorde con las necesidades del negocio. Por lo tanto, la facilidad y rapidez del desarrollo vendrá determinado por la complejidad del sistema a modelar.

2.7.1 Levantamiento de la Información e Indicadores

Esta es la primera fase del método y en ella se realiza un levantamiento de la información, donde además de describir y priorizar los procesos de la organización, se definen los reportes e indicadores que deben ser desarrollados. Finalmente, se identifican las fuentes de datos en donde reside la información a extraer.

2.7.2 Modelamiento Dimensional

En esta segunda fase se lleva a cabo la toma de los requerimientos levantados en la fase anterior, se identifica la jerarquía de los datos involucrados, además, se identifican los hechos a medir y se identifican las dimensiones o criterios de clasificación de los hechos.

2.7.3 Desarrollo del Almacén de Datos

La tercera fase consiste en la creación del Almacén de Datos, de acuerdo con el modelo dimensional obtenido. Luego, realizar la carga de los datos involucrados desde las fuentes de datos hacia el Almacén de Datos.

2.7.4 Análisis de Consistencia de Datos del Almacén de Datos

Esta es la cuarta fase y consiste en verificar que los datos sean correctos, consistentes, íntegros y durables. Esto se hace realizando consultas en las tablas transaccionales y comparándolas con los resultados de las consultas al Almacén de Datos. Si existe inconsistencia entre los datos, se verifica la fase anterior con el fin de encontrar posibles errores en la carga de datos.

2.7.5 Desarrollo de Indicadores

La quinta fase consiste en la creación de las consultas y cuadros de mando necesarios para mostrar los indicadores y reportes que se definieron en la primera fase, utilizando los datos contenidos en el Almacén de Datos.

2.7.6 Análisis de Consistencia de los Cuadros de Mando

La sexta y última fase consiste en verificar que los datos que se muestran en los cuadros de mando, tengan una lógica y puedan brindar la información necesaria para el análisis. También se verifica si los datos son correctos con respecto a las distintas fuentes de datos. Si se encuentra inconsistencia en los datos que se encuentran en los cuadros de mando, se regresa a la fase anterior en busca de posibles errores en la creación de los cuadros de mando.

2.8 Proceso de Revisión Editorial

La Revisión Editorial es un proceso que se lleva a cabo, generalmente en sistemas de publicación especializados, ya sean impresos o digitales, donde los artículos enviados por los autores son pasados por una evaluación antes de ser publicados (Barruecos, 2000). Principalmente, esta actividad es llevada a cabo por un comité editorial o por unos jurados, quienes son los encargados de analizar, revisar y evaluar los artículos que le sean asignados, asegurándose de que los mismos cumplan con ciertos criterios de calidad (Campanario, 2002).

Sin embargo, a pesar de que la producción de publicaciones especializadas ha ido incrementando y, a su vez, han ido modificando su proceso de elaboración y edición; para

Barruecos (2000), el sistema de publicación de los artículos o documentos ha permanecido inalterable en el tiempo (ver figura 30).



Figura 30. Gráfico del Sistema de Publicación Científica (Barruecos, 2000)

Si bien se puede notar, en la figura 30, ambos extremos del gráfico (destacados en color amarillo) están ocupados por científicos o investigadores, ya sea como autor o lector de artículos en el ámbito de la ciencia. Para Barruecos (2000), esta característica es la primordial para distinguir este sistema de publicación de otros sistemas.

En primer lugar, el denominado autor (primer eslabón del gráfico) realiza una investigación acerca de un tema específico y se dispone a crear un documento, bajo ciertas normas establecidas sobre los resultados y conclusiones de su investigación. Una vez finalizado, lo remite a un editor (segundo eslabón del gráfico).

Seguidamente, el editor del sistema de publicación, quien debe tener conocimiento en el tema, realiza una primera valoración del documento, donde evalúa si dicho artículo es pertinente para los objetivos de publicación. Si así es, el editor envía la publicación a un comité evaluador (tercer eslabón del gráfico), que debe contar con científicos expertos en el tema.

El comité evaluador, llevará a cabo un análisis sobre la calidad del trabajo, tomando en cuenta la adecuación del mismo a los objetivos planteados en el sistema de publicación. Una vez finalizada la evaluación del artículo, el comité propondrá los cambios que consideren oportunos tanto de estilo como de contenido, si así lo requiere el documento. Si el trabajo pasa la evaluación del comité, está listo para su propagación en un próximo número del sistema de publicación.

En caso contrario, o se puede remitir al autor con las modificaciones pertinentes o se da una respuesta negativa. Una vez realizada las correcciones, el trabajo debe pasar por el mismo sistema de publicación.

Para el caso de estudio, implementado en este Trabajo Especial de Grado, se ha desarrollado un diagrama de actividades, que se puede observar en la figura 31, donde se muestra el flujo de trabajo que lleva a cabo sistema de publicación digital para la divulgación de un artículo.

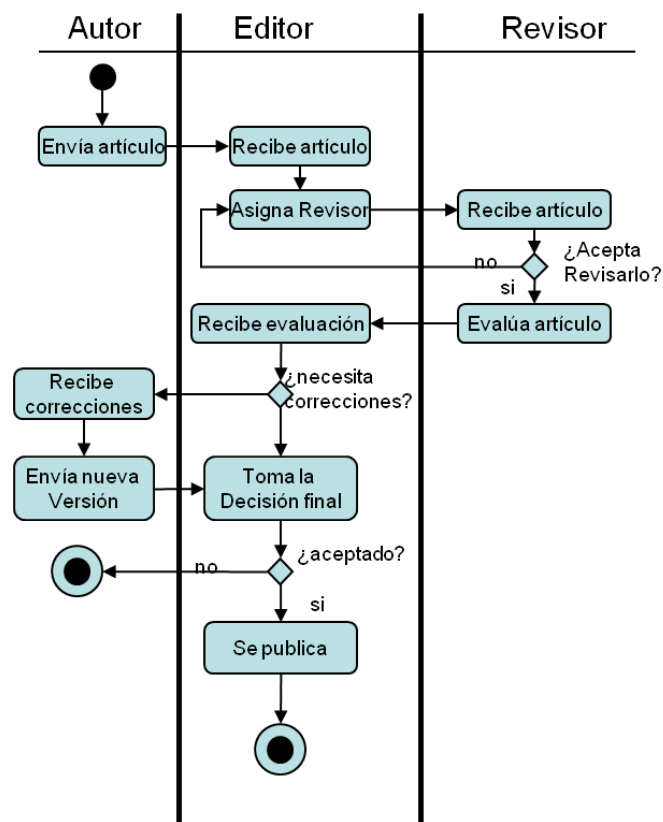


Figura 31. Diagrama de Actividades de un Sistema de Publicación (Fuente: el Autor)

En primer lugar, el Autor envía el artículo. Seguidamente el Editor recibe el artículo y le asigna los revisores que evaluarán dicho artículo. El Revisor recibe la invitación a revisar el artículo y toma una decisión si lo revisa o no. En caso que un Revisor decida no tomar la evaluación, el Editor debe asignar otro revisor. En caso que el Revisor decida tomar la evaluación, revisa el artículo y lo evalúa. Luego envía el resultado de la evaluación al Editor: si es aceptado, si necesita revisiones o si es rechazado.

Una vez evaluado el artículo el Editor envía las revisiones al Autor, si son necesarias. Si el artículo necesita correcciones, el Autor envía una nueva versión del artículo al Editor para que éste tome una decisión final. Si no necesita revisiones, el Editor toma la decisión final: si el artículo es aceptado o rechazado.

Luego de que un artículo es aceptado, se publica.

Capítulo III. Marco Aplicativo

En el siguiente capítulo se presenta un método de desarrollo, de la empresa Tian Consultores, adaptado a los métodos: ascendente (Ralph Kimball) y descendente (Bill Inmon), aprovechando sus mejores prácticas, por parte de la empresa Tian Consultores. De esta manera, se presenta el contexto de desarrollo y cada una de las fases establecidas en dicho método utilizado para el desarrollo del objetivo principal de este Trabajo Especial de Grado. Asimismo, dentro de la primera fase de este método se describe la metodología general presentada por Beltran (2000) para la definición y documentación de un conjunto de indicadores.

3.1 Contexto de Desarrollo

El presente proyecto ha sido desarrollado en el contexto de T. E. G., presentado ante la ilustre Universidad Central de Venezuela, Facultad de Ciencias, Escuela de Computación, bajo la tutoría del profesor Franky Uzcátegui, para optar al título de Licenciado en Computación, por parte del bachiller Fabián E. Moreno C. como único integrante del equipo desarrollador del mismo.

3.2 Proyecto

Dicho T. E. G., como bien se ha señalado, se basa en la implementación de indicadores de gestión, que contribuyan al monitoreo del rendimiento del proceso de Revisión Editorial de un sistema de publicación digital, bajo una plataforma de Inteligencia de Negocios de Software Libre, denominada *Pentaho Business Intelligence Open Source*. Así, el proyecto permite, a la gerencia, elaborar y obtener cuadros de mando, facilitando el análisis de la situación para la toma de decisiones.

Pentaho Business Intelligence Open Source presenta una serie de componentes y herramientas que facilitan las etapas de desarrollo, implementación y pruebas.

Es por esto que para desarrollar la solución propuesta en el presente documento, se utilizó el método de desarrollo, de Tian Consultores, adaptado a los métodos propuestos por Ralph Kimball y Bill Inmon, aprovechando las mejores prácticas de cada uno. Dicho método de desarrollo cuenta con seis fases que serán descritas y detalladas a continuación.

3.3 Fases del Método de Desarrollo

De esta manera, se presentan cada una de las fases aplicadas del método de desarrollo para una solución de Inteligencia de Negocios.

3.3.1 Fase1: Levantamiento de la Información e Indicadores

Primeramente se realizó un estudio entre cuatro sistemas manejadores de bases de datos, como se muestra en la tabla 1, ubicada en la segunda sección del Marco Conceptual, con la finalidad de escoger aquel sistema que aportara mejores características y beneficios para contribuir con el desarrollo de la solución planteada. De esta forma, *PostgreSQL* fue el sistema manejador de base de datos elegido, ya que además de ser licencia libre, maneja modelo relacional y objeto-relacional.

De igual forma, tomando el estudio comparativo de Stratebi (2010) sobre tres plataformas de Inteligencia de Negocios, *Pentaho Business Intelligence* fue la elegida, ya que aportaba mejores características y componentes para la realización de la solución planteada. Además, permitía la integración con *PostgreSQL*.

Por otra parte, la implementación de indicadores de gestión es de gran utilidad para las organizaciones y empresas que deseen mantener un control y seguimiento de un determinado proceso, estableciendo su logro, meta, objetivo o punto de éxito. Por esta razón, se hizo contacto con un experto en sistemas de publicación (Reina Taylhardat, ex directora de la revista *Complot*), con el fin de identificar los principales problemas que trae consigo el proceso de revisión de artículos y los puntos de éxitos que se pueden lograr.

En un principio se estudió y se analizó el proceso de Revisión Editorial que lleva a cabo un sistema de publicación digital, antes de realizar la propagación de artículos. Asimismo, se pudo detallar los datos involucrados en dicho proceso y se pudo conocer la intervención de varios datos de tipo fecha como: la fecha en que se envía un artículo, la fecha en que se inicia la revisión de un artículo, etc.

De esta misma forma, y con ayuda del experto, se realizó la construcción de un sistema de indicadores de gestión que establecieran los factores claves de éxito del proceso de Revisión Editorial. Siguiendo con la metodología planteada por Beltran (2000), presentada en la quinta sección del Marco Conceptual y constituida por nueve pasos para la definición y documentación de indicadores de gestión, se tiene lo siguiente:

a) Paso 1 - Contar con objetivos y estrategias:

Con ayuda del experto se pudo distinguir que el objetivo estratégico del proceso de Revisión Editorial estaba basado en la cantidad de tiempo en que se publica un artículo, desde el envío del mismo por parte del autor y la distribución del mismo en los subprocesos que lo componen.

Por esto, se estableció como objetivo estratégico mantener disminuidos los tiempos de respuestas al autor, en cuanto a días, para la divulgación de su artículo en uno de los ejemplares del sistema de publicación digital.

b) Paso 2 - Identificar factores claves de éxitos:

Se pudo identificar: el tiempo total de publicación, el tiempo de respuesta cuando se asigna un revisor, el tiempo de revisión, el tiempo en dar una decisión final y el tiempo en publicar el artículo.

c) Paso 3 - Definir indicadores para factores claves:

- Tiempo total de publicación: Es la cantidad de días, desde que el autor envía el artículo hasta que el mismo se publica.

$$\text{Total_Publicación} = \text{Fecha_Publicación} - \text{Fecha_Envío}$$

- Tiempo de asignación de revisión: Es la cantidad de días, desde que se envía el artículo hasta que el editor asigna un revisor, y dicho revisor acepta evaluarlo.

$$\text{Asignación} = \text{Fecha_Inicio_Revisión} - \text{Fecha_Envío}$$

- Tiempo de revisión de artículo: Es la cantidad de días que se tarda un revisor en evaluar el artículo completo y enviar dicha evaluación.

$$\text{Revisión} = \text{Fecha_Fin_Revisión} - \text{Fecha_Inicio_Revisión}$$

- Tiempo de decisión final: Se refiere a la cantidad de días que se tarda el editor, desde que se finalizó la revisión hasta dar una respuesta definitiva de publicación del artículo al autor.

$$\text{Decisión} = \text{Fecha_Decisión} - \text{Fecha_Fin_Revisión}$$

- Tiempo de publicación: Representa la cantidad de días, desde que se toma la decisión final de publicación hasta la publicación del artículo.

$$\text{Publicación} = \text{Fecha_Publicación} - \text{Fecha_Decisión}$$

d) Paso 4 - Determinar estado, umbral y rango de gestión:

Una vez definidos los indicadores, en la tabla 5, se determina el estado, umbral y rango de gestión de los mismos.

Tabla 5. Indicadores construidos con su estado, umbral y rango

Índice	Estado	Umbral	Rango
Tiempo de publicación total	Desconocido	30 días	(0-50)
Tiempo de asignación de revisión	Desconocido	3 días	(0-10)
Tiempo de revisión de artículo	Desconocido	3 días	(0-10)
Tiempo de decisión final	Desconocido	3 días	(0-20)
Tiempo de publicación	Desconocido	10 días	(0-20)

En esta fase del método para la construcción de un sistema de indicadores de gestión, influye la alerta estratégica implementada bajo tres colores fundamentales: el verde para cuando el valor del indicador se encuentre por debajo del umbral, el amarillo para cuando el indicador ya se esté acercando al umbral y el rojo para cuando el indicador sobrepasó el valor del umbral. En este último caso el color rojo advierte alerta.

e) Paso 5 - Diseñar la medición:

En la tabla 6 se muestran los indicadores de gestión diseñando la medición de los mismos, estableciendo la fuente de donde serán extraídos, la frecuencia de medición y el responsable de la misma.

Tabla 6. Indicadores de Gestión

Índice	Fuente	Frecuencia	Responsable
Tiempo de publicación total	Base de datos transaccional	Mensual	Director o Editor
Tiempo de asignación de revisión	Base de datos transaccional	Mensual	Director o Editor
Tiempo de revisión de artículo	Base de datos transaccional	Mensual	Director o Editor
Tiempo de decisión final	Base de datos transaccional	Mensual	Director o Editor
Tiempo de publicación	Base de datos transaccional	Mensual	Director o Editor

f) Paso 6 - Determinar y asignar recursos:

En este paso se establecen las necesidades del recurso que demanda la realización de las mediciones. Lo ideal es que la medición integre el desarrollo del trabajo y los recursos que se utilicen para la medición sean parte de los recursos que se emplean para el desarrollo del proceso.

g) Paso 7 - Medir y ajustar:

Este paso se realiza una vez implementados los indicadores de gestión en la plataforma de Inteligencia de Negocios.

h) Paso 8 - Estandarizar y formalizar:

Este paso se realiza una vez implementados los indicadores de gestión en la plataforma de Inteligencia de Negocios.

i) Paso 9 - Mantener y mejorar continuamente:

Este paso se realiza una vez implementados los indicadores de gestión en la plataforma de Inteligencia de Negocios.

Una vez identificados y establecidos los indicadores que se van a modelar, se diseña el área o ambiente transaccional, desde donde se van a extraer los datos necesarios para representarlos en los cuadros de mandos y los reportes. Asimismo, se establece un sistema transaccional del proceso de Revisión Editorial, diseñando el modelo entidad-relación que permitirá la creación de la base de datos intermedia.

De este modo, en la figura 32 se observa el modelo entidad-relación diseñado para la base de datos transaccional:

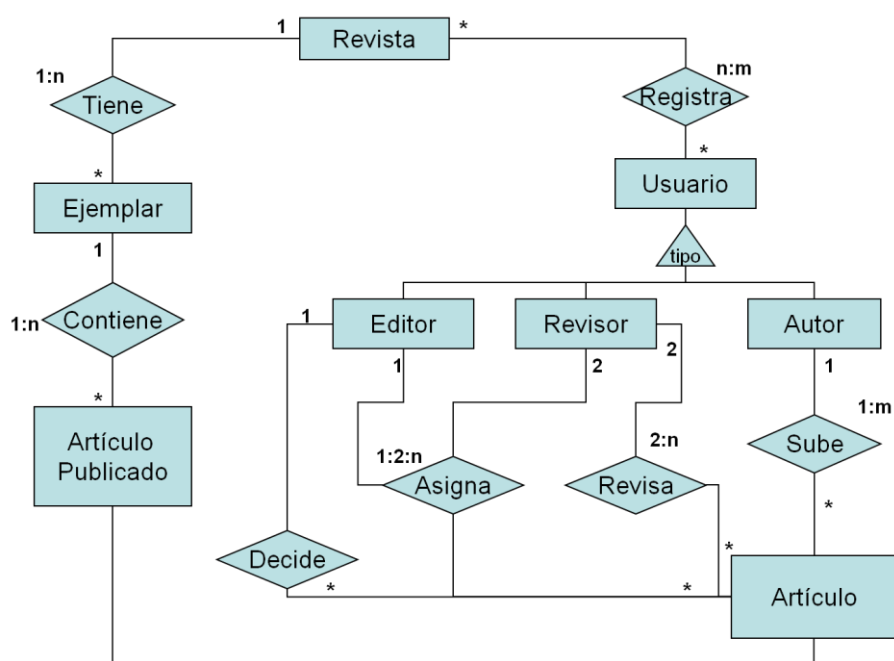


Figura 32. Modelo Entidad-Relación del ambiente intermedio

Este modelo representa un ambiente donde una revista está compuesta por ejemplares y éstos a su vez, contienen una cantidad de artículos publicados. Dentro de dicha revista están registrados un grupo de usuarios, destacados como: editor, revisor y autor. El usuario autor es el que sube artículos al sistema, el editor asigna un revisor a cada artículo y éste último revisa el mismo.

También se puede apreciar que el editor es el encargado de decidir sobre la publicación, o no del artículo, en uno de los ejemplares que componen a dicho sistema de publicación. Los artículos que sean aprobados son los llamados artículos publicados.

Por lo tanto, se puede decir que en esta primera fase del método de desarrollo lo que se buscó, principalmente, es realizar el levantamiento de información que permita establecer y construir los indicadores de gestión que van a monitorear el proceso de revisión de artículos de la revista digital. Asimismo, se diseñó el modelo de la base de datos transaccional intermedia, que simule un sistema operacional de donde se extraerán los datos que se van a cargar en el almacén de datos.

3.3.2 Fase 2: Modelamiento Dimensional

Una vez identificados los requerimientos principales y establecidos los indicadores, se define una jerarquía de información transaccional con el objetivo de proveer una lógica a los datos del proceso con respecto al ambiente de Inteligencia de Negocios. En la figura 33, se puede observar dicha jerarquía, donde se identifican los atributos por los cuales se navegará dentro de la aplicación.



Figura 33. Jerarquía de la Información

Por cada fecha de envío se puede detallar los actores “autores” que han enviado artículo en la misma, para una revista específica. A su vez, dicha revista está compuesta por una variedad de ejemplares donde serán publicados los artículos enviados por los autores.

Ya definida la jerarquía de la información transaccional, se dispuso a realiza el modelo dimensional. En dicho modelo se definen cada una de las dimensiones que conformarán los

criterios de consulta sobre el cubo. Adicionalmente, se define el cubo y cada uno de los hechos o eventos a medir. Cabe destacar, que la jerarquía de la información transaccional es la que determina la granularidad de los hechos a medir en dicho cubo, así como también la granularidad de cada una de las dimensiones definidas.

Es por esto que luego de haber definido las dimensiones del almacén de datos, especificamos los diferentes grados de detalle (atributos) dentro de cada dimensión, así como también la granularidad de cada indicador y las diferentes jerarquías que dan forma al modelo dimensional.

El modelo dimensional que se presenta en la figura 34, determina el modo de almacenamiento de la información transaccional dentro del almacén de datos para que pueda ser explotable por la plataforma de Inteligencia de Negocios. Logrando de esta forma, optimizar el proceso de consulta sobre los datos contenidos en dicho almacén.



Figura 34. Modelo Dimensional

El modelo dimensional, representado en la figura 34, está compuesto de 4 dimensiones y un cubo:

- Dimensión Tiempo: Esta dimensión representa la fecha en que los autores envían los artículos a la revista digital.
- Dimensión Usuario: La dimensión usuario representa todos los usuarios asociados a la revista digital, ya sean autores, revisores y editor.

- Dimensión Revista: En esta dimensión se representa una revista digital y los ejemplares que la misma emite.
- Dimensión Artículo: Esta dimensión constituye a los artículos que han sido evaluados y aprobados, para ser publicados en un ejemplar de una revista digital.
- Cubo Publicación: En el cubo publicación se representan todos los eventos o hechos que se desean medir, así como también las relaciones entre cada una de las dimensiones.

Se puede afirmar que el modelo dimensional, principalmente, determina el modo de almacenamiento de la información transaccional dentro del almacén de datos, con la finalidad de explotar los mismos para la implementación de los indicadores de gestión, bajo la plataforma de Inteligencia de Negocios. Así, se consigue optimizar el proceso de consultas sobre los datos contenidos en el almacén.

A continuación, en la figura 35, se representa el modelo dimensional, especificado anteriormente, en forma de árbol jerárquico, detallando los niveles y atributos, críticos y no críticos, de cada una de las dimensiones. Son llamados atributos críticos a aquellos atributos por medio de los cuales se puede navegar dentro de la aplicación y los no críticos, representan características adicionales de las dimensiones. Vale destacar, que los artículos que se cargarán en el almacén son aquellos que hayan pasado por todos los subprocessos de la Revisión Editorial, es decir aquellos que ya hayan sido publicados.



Figura 35. Niveles jerárquicos y medidas

Luego, a manera de detalle, se describe la estructura de cada una de las dimensiones del modelo dimensional y del cubo.

En la tabla 7, se muestra la estructura de la dimensión tiempo, con su definición, atributos y tipos de datos.

Tabla 7. Dimensión Tiempo

TIEMPO		
Descripción:	Tabla que almacena las fechas en que se envía un producto por parte del autor.	
Tipo de Tabla:	Tabla de dimensión	
Atributos		
id_technical_tiempo	INT (11) NOT NULL AUTO_INCREMENT	Identificador único y técnico del registro en la tabla dimensión tiempo.
Versión	INT (11) NOT NULL	Campo para la versión de datos.
date_from	DATE NOT NULL	Fecha de creación
date_to	DATE NOT NULL	Fecha de expiración
Fecha	DATE NOT NULL	Fecha de envió de artículo
dia_mes	INT (11) NOT NULL	Día del mes de envió de artículo
Mes	INT (11) NOT NULL	Mes en número de envió de artículo
Año	INT (11) NOT NULL	Año de envió de artículo
dia_ano	INT (11) NOT NULL	Día del año en que se envió el artículo
dia_semana	VARCHAR (250) NOT NULL	Nombre del día de la semana en que se envió el artículo
mes_ano	VARCHAR (250) NOT NULL	Nombre del mes en que se envió el artículo

En la tabla 8, se muestra la estructura de la dimensión usuario, con su definición, atributos y tipos de datos.

Tabla 8. Dimensión Usuario

USUARIO		
Descripción:	Tabla que almacena la información de los usuarios registrados en la revista.	
Tipo de Tabla:	Tabla de dimensión	
Atributos		
id_technical_tiempo	INT (11) NOT NULL AUTO_INCREMENT	Identificador único y técnico del registro en la tabla dimensión usuario.
Versión	INT (11) NOT NULL	Campo para la versión de datos.
date_from	DATE NOT NULL	Fecha de creación
date_to	DATE NOT NULL	Fecha de expiración
id_usuario	INT (11) NOT NULL	identificador único del usuario en la revista digital
nombre_usuario	VARCHAR (250) NOT NULL	Nombre y apellido del usuario
Nacionalidad	INT (11) NOT NULL	Nacionalidad del usuario
Rol	INT (11) NOT NULL	Rol del usuario

En la tabla 9, se muestra la estructura de la dimensión revista, con su definición, atributos y tipos de datos.

Tabla 9. Dimensión Revista

USUARIO		
Descripción:	Tabla que almacena la información de una revista digital y sus ejemplares.	
Tipo de Tabla:	Tabla de dimensión	
Atributos		
id_technical_tiempo	INT (11) NOT NULL AUTO_INCREMENT	Identificador único y técnico del registro en la tabla dimensión revista.
Versión	INT (11) NOT NULL	Campo para la versión de datos.
date_from	DATE NOT NULL	Fecha de creación
date_to	DATE NOT NULL	Fecha de expiración
id_revista	INT (11) NOT NULL	identificador único de la revista
id_ejemplar	INT (11) NOT NULL	Identificador único del ejemplar concatenado con el identificador de la revista
Name_revista	VARCHAR (250) NOT NULL	Nombre de la Revista
Name_ejemplar	VARCHAR (250) NOT NULL	Nombre del ejemplar
Ano	INT (11) NOT NULL	Año de publicación del ejemplar
Volumen	INT (11) NOT NULL	Volumen del ejemplar
Numero	INT (11) NOT NULL	Número del ejemplar

En la tabla 10, se muestra la estructura de la dimensión artículo, con su definición, atributos y tipos de datos.

Tabla 10. Dimensión Artículo

ARTÍCULO		
Descripción:	Tabla que almacena la información de los artículos que son aprobados y publicados	
Tipo de Tabla:	Tabla de dimensión	
Atributos		
id_technical_tiempo	INT (11) NOT NULL AUTO_INCREMENT	Identificador único y técnico del registro en la tabla dimensión artículo.
Versión	INT (11) NOT NULL	Campo para la versión de datos.
date_from	DATE NOT NULL	Fecha de creación
date_to	DATE NOT NULL	Fecha de expiración
id_articulo	INT (11) NOT NULL	identificador único del artículo
Titulo	VARCHAR (250) NOT NULL	Título del artículo
Estado	VARCHAR (250) NOT NULL	Estado del artículo
Fecha_comienzo	DATE NOT NULL	Fecha en que se inició la revisión
Fecha_fin	DATE NOT NULL	Fecha en que se finalizó la revisión
Fecha_publicacion	DATE NOT NULL	Fecha en que se publicó
Fecha_decision	DATE NOT NULL	Fecha en que se tomó la decisión de publicarlo
Revisor	VARCHAR (250) NOT NULL	Nombre y Apellido del revisor que evaluó el artículo

En la tabla 11, se muestra la estructura del cubo publicación, con su definición, atributos y tipos de datos.

Tabla 11. Cubo Publicación

PUBLICACIÓN		
Descripción:	Tabla que almacena la información de los artículos que son aprobados y publicados	
Tipo de Tabla:	Tabla de dimensión	
Atributos		
Id_technical_publicación	INT (11) NOT NULL AUTO_INCREMENT	Identificador único y técnico del registro en la tabla publicación.
Id_usuario	INT (11) NOT NULL	Identificador del usuario que envió el artículo.
Id_rev_eje	INT (11) NOT NULL	Identificador del ejemplar concatenado con el identificador de la revista donde se publicó.
Id_artículo	INT (11) NOT NULL	Identificador del artículo que se publicó.
Fecha	DATE NOT NULL	Fecha de envío del artículo que se publicó
Días_asignación	INT (11) NOT NULL	Cantidad de días en asignar un revisor.
Días_revisión	INT (11) NOT NULL	Cantidad de días que tardó su revisión.
Días_decision	INT (11) NOT NULL	Cantidad de días que tardó la decisión final.
Días_publicación	INT (11) NOT NULL	Cantidad de días que tardó su publicación

En resumen, se puede afirmar que en esta fase es donde se diseña cómo se va a agrupar la información transaccional, de tal forma que pueda ser representada y almacenada dentro del ambiente de Inteligencia de Negocios. Asimismo, se detallan cada una de las dimensiones y los hechos a medir, modelando la lógica del proceso de Revisión Editorial.

3.3.3 Fase 3: Desarrollo del Almacén de Datos

En esta tercera fase del método de desarrollo, se realiza la creación del almacén de datos, de acuerdo con el modelo dimensional diseñado anteriormente.

La creación del almacén de datos consta de una serie de subfases necesarias para lograr exitosamente la extracción, estandarización y carga de los datos desde las diferentes fuentes hacia el almacén. Así, se describen a continuación cada una de las subfases desarrolladas:

a) Creación de las tablas del modelo de entidad-relación e importación de los datos de la base de datos transaccional:

En esta primera subfase se deben crear las tablas de la base de datos intermedia que se utilizará para almacenar los datos los datos que cumplan con los requerimientos mínimos para la construcción de un modelo de Inteligencia de Negocios. Para realizar esta tarea se utilizó el componente de integración de datos de la suite de *Pentaho*, denominado *Data Integration*.

Dicho componente, no sólo permite construir las tablas del modelo entidad - relación, sino que además permite migrar los datos desde las diferentes fuentes hacia las tablas destinos. Asimismo, este integrador de datos contiene una gran variedad de operadores que permiten realizar los procesos de extracción, transformación y carga (ETL, por sus siglas en inglés). En la figura 36, se detallan algunos operadores incluidos en este componente para realizar las distintas transformaciones.














	Paso de Entrada <i>Excel Input:</i> Lee los datos de un archivo de Excel.
	Paso de Entrada <i>Get System Info:</i> Obtiene información de la fecha del sistema.
	Paso de Salida <i>Table output:</i> Escribe información en una base de datos usando una conexión a la misma.
	Paso de Transformación <i>Calculator:</i> Proporciona operaciones para crear nuevos campos mediante la realización de cálculos simples.
	Paso de Transformación <i>Filter rows:</i> Filtra filas mediante ecuaciones simples.
	Paso de Transformación <i>Append streams:</i> Une dos bloques de datos con la misma estructura.
	Paso de Transformación <i>Row denormaliser:</i> Nos permite transformas filas en columnas dado un valor clave. Dado que queremos operar entre filas, para ello lo llevamos a columnas.
	Paso de Transformación <i>Row Normaliser:</i> Nos permite transformas columnas en filas.
	Paso de Transformación <i>Group by:</i> Agrupar valores por una característica determinada.
	Paso de Transformación <i>Select values:</i> Escoge los valores que interesan. Se utiliza para seleccionar y modificar el orden exacto y el nombre en que los campos tienen que ser colocados en las filas de salida, también se utiliza para eliminar los campos que se quiere retirar.
	Paso de Transformación <i>Sort rows:</i> Ordena ascendente y descendente una serie de campos que se especifican.
	Paso de Datawarehouse <i>Dimension Lookup/update</i> Actualización de una dimensión que cambian lentamente en el almacén de datos. Y también es utilizado para buscar información en una dimensión.
	Paso de Transformación <i>Value Mapper</i> Busca valores en específico y lo transforma en uno solo.

Figura 36. Algunos operadores de Pentaho Data Integration

En este proyecto, los datos operacionales se encuentran en archivos Excel. En la figura 37, se muestra la creación de las tablas del modelo entidad relación y la migración automática de los datos hacia las mismas, por medio del *Data Integration*.

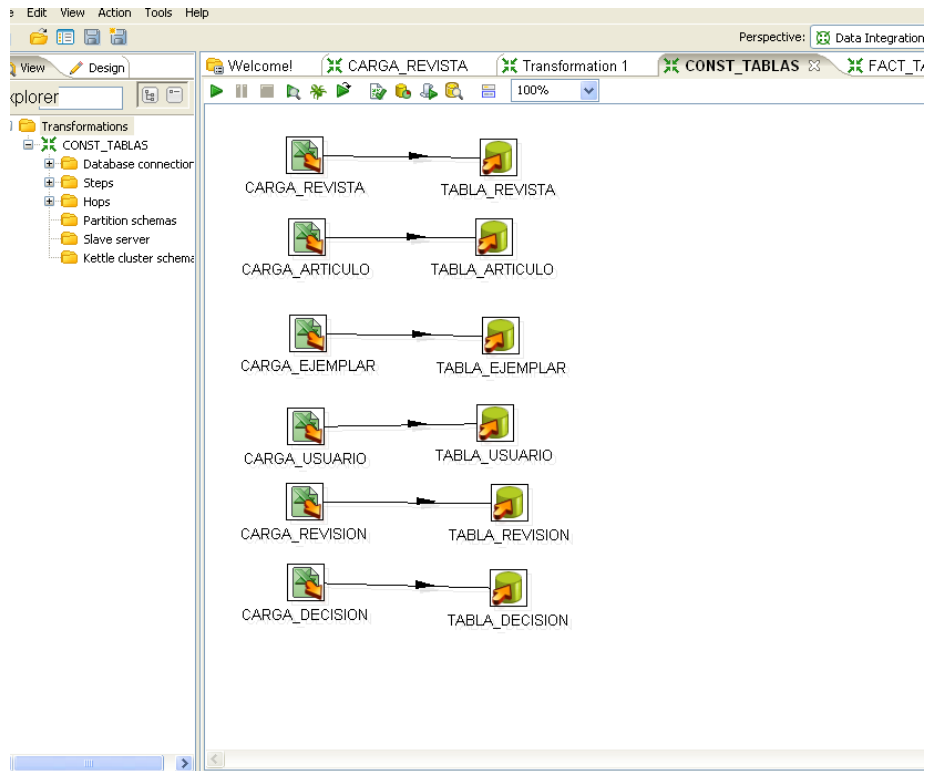


Figura 37. Creación de las tablas del Modelo Entidad-Relación

b) Creación de las tablas dimensiones y el mapa para cargar los datos:

Una vez que los datos del ambiente transaccional se encuentran en las tablas creadas en la subfase anterior, se deben crear las dimensiones y el cubo, basados en el modelo dimensional diseñado en la fase anterior.

Siguiendo con el modelo dimensional, primero se deben crear las tablas de cada una de las dimensiones, junto con sus atributos. El componente *Data Integration*, nos permite con un solo mapa realizar la creación de la tabla y la migración de los datos, desde las tablas construidas en la subfase anterior.

En la figura 38, se presenta el mapa utilizado para realizar la creación y migración de datos de la dimensión artículo.

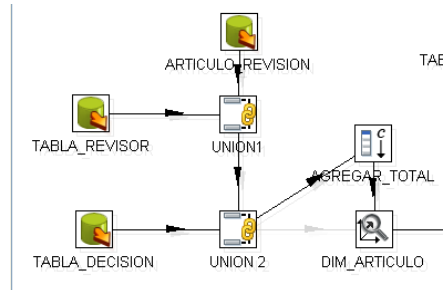


Figura 38. Mapa para la creación de la dimensión artículo

En el mapa de la figura 38, se puede apreciar como partiendo de las tablas artículo, revisión y decisión se construye la dimensión artículo junto con sus atributos.

En la figura 39, se muestra el mapa utilizado para crear y migrar los datos de la dimensión revista.

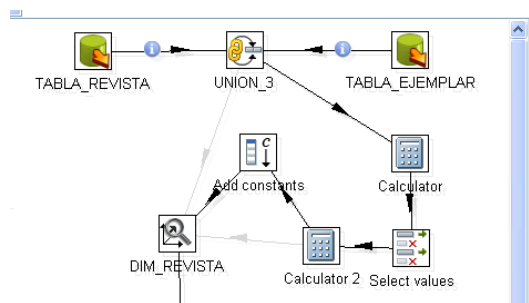


Figura 39. Mapa para la creación de la dimensión revista

En este mapa se muestra como se parte de las tablas ejemplar y revista para construir la tabla dimensión artículo junto con sus atributos. De igual forma, se utilizaron el operador de "Calculator" y "Select Values" para construir un identificador concatenando los identificadores de revista y ejemplar.

En la figura 40, se muestra el mapa utilizado para crear y migrar los datos de la dimensión usuario.

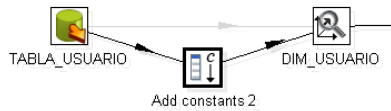


Figura 40. Mapa para la creación de la dimensión usuario

En este mapa se muestra como se parte de la tabla usuario, para construir la tabla de dimensión usuario, junto con sus atributos.

En la figura 41, se muestra el mapa utilizado para crear y migrar los datos de la dimensión tiempo.

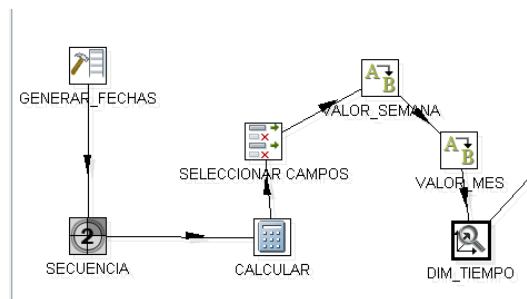


Figura 41. Mapa para construir la dimensión Tiempo

En este mapa se muestra como partiendo del operador “Generate Row” se generan diez mil filas con un atributo fecha, que tiene como valor por defecto ‘01/01/2011’. Luego con el operador de “Add Sequence”, con el cual se agrega una secuencia que comience en cero e incremente en uno. Seguidamente con el operador “Calculator” se agrega un atributo que sume el campo fecha por defecto más el primer valor de la secuencia, luego con el segundo valor de la secuencia y así sucesivamente. De esta forma, se construyen diez mil filas de fechas, comenzando por el valor ‘01/01/2011’ e incrementando en un día. De igual manera, con el mismo operador se calculan los demás atributos de la dimensión, ya sea día del mes, número del mes, año, día de la semana, etc. Así, con el operador “Select Values”, se seleccionan los valores de número de mes y día de la semana, para construir con el operador “Value Mapper” un atributo *string*, con el nombre del mes y el nombre del día de la semana. Finalmente, se construye la dimensión tiempo.

c) Creación de la tabla del cubo y mapa para cargar los datos:

Ya construidas todas las dimensiones, se debe crear el cubo y la tabla de hechos, junto con sus atributos. En el mismo se representan los hechos o eventos que se van a medir y las dimensiones con las cuales se relaciona, así como también, por cuales atributos de esas dimensiones se relacionan.

Por lo tanto, se construyó un mapa que involucrara los mapas de la creación de todas las tablas dimensiones, y a su vez, realizara la creación de la tabla del cubo y sus atributos. De tal forma que cuando se hiciera migración de datos, se almacenaran datos en las dimensiones, si no existían en estas. En la figura 42, se presenta el mapa que construye y migra los datos hacia el cubo.

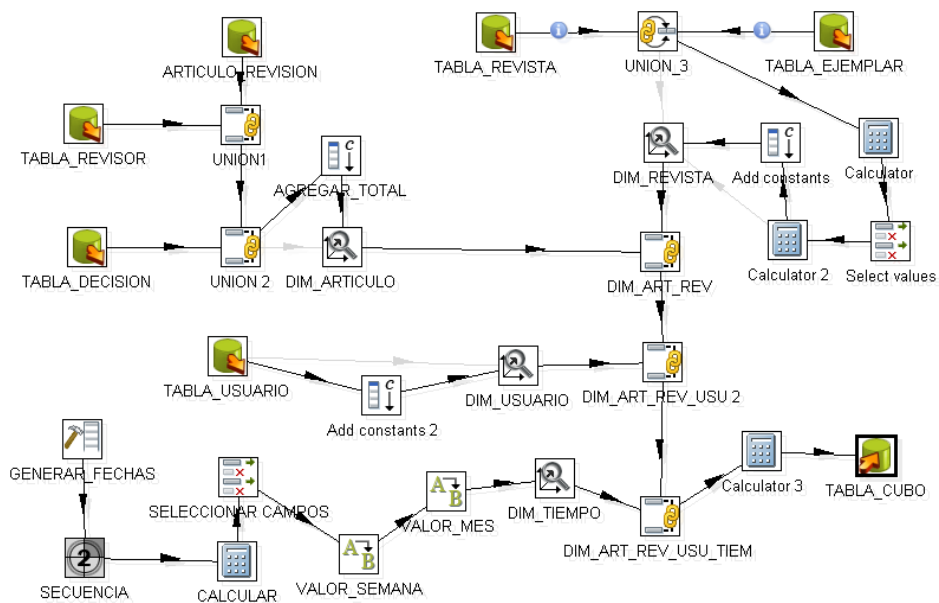


Figura 42. Creación de la tabla del cubo publicación

En este mapa se observa cómo se involucra la creación de todas las tablas dimensiones y la tabla del cubo. Se observa cómo se utiliza el operador "Join Rows" para ir uniando aquellos registros que nos interesan guardar en el cubo de cada una de las dimensiones. Finalmente se usa el operador "Calculator" para realizar operaciones y generar los atributos calculados.

d) Diseño del Cubo:

Ya construidas las tablas dimensiones y la tabla del cubo, se debe diseñar el esquema lógico del cubo, identificando las dimensiones que relaciona, junto con sus atributos y jerarquía. Adicionalmente, se deben establecer los hechos a medir. Para llevar a cabo esta tarea, se utilizó el componente de Pentaho que permite diseñar esquemas, denominado *Schema Workbench*.

Dicho componente permite especificar el tipo de agregación que pueden tener los hechos a medir, como pudiera ser suma, promedio, máximo, mínimo, entre otros. Esto permite tener una regla que define la manera en que se van a agrupar los atributos pertenecientes a los niveles más detallados, para formar los niveles resumidos.

Por lo tanto, la creación del esquema del cubo es de suma importancia ya que por medio del mismo se definen los hechos o eventos que se quieren representar, basados en los requerimientos del proceso.

En la figura 43, se puede apreciar la construcción de un esquema llamado “Revisión”, dentro de dicho esquema se construyó un cubo, denominado “Revisión Editorial”. Una vez creado el cubo, se le asigna la tabla de hechos relacionada. Luego se definen las dimensiones que se relacionan en dicho cubo, por medio de los atributos establecidos.

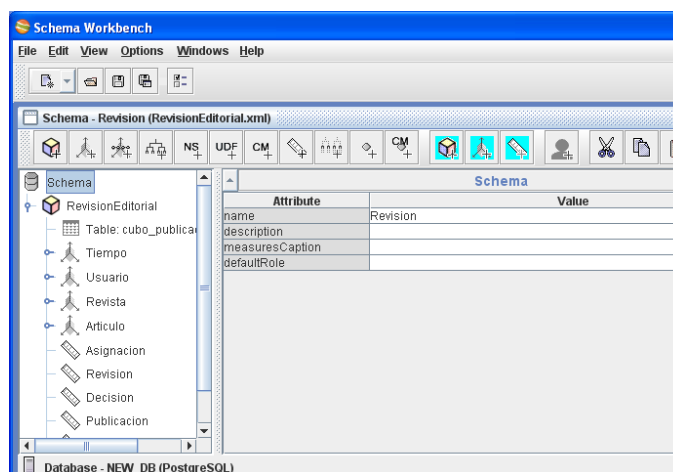


Figura 43. Creación del esquema del cubo

En la figura 44, se presenta la definición de la dimensión tiempo, especificando la jerarquía de la misma.

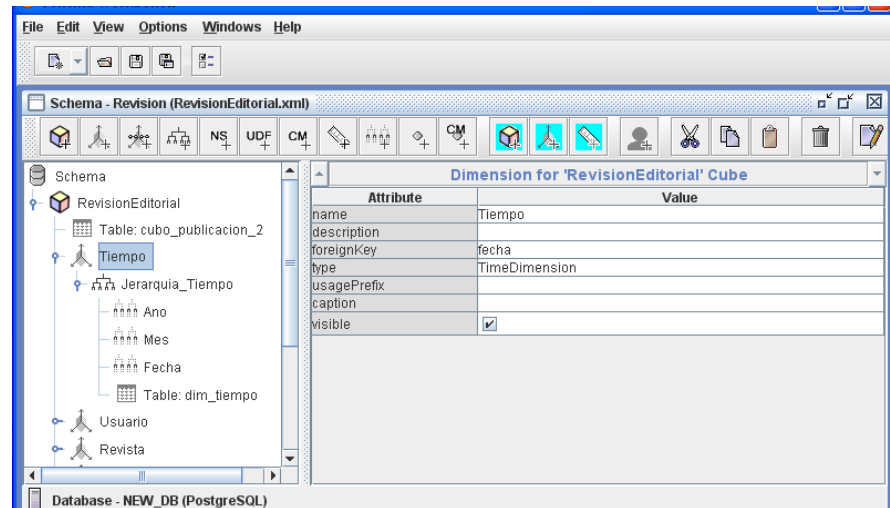


Figura 44. Creación de la Dimensión Tiempo

Asimismo, en la figura 45, se muestra la declaración de la dimensión usuario y su jerarquía establecida.

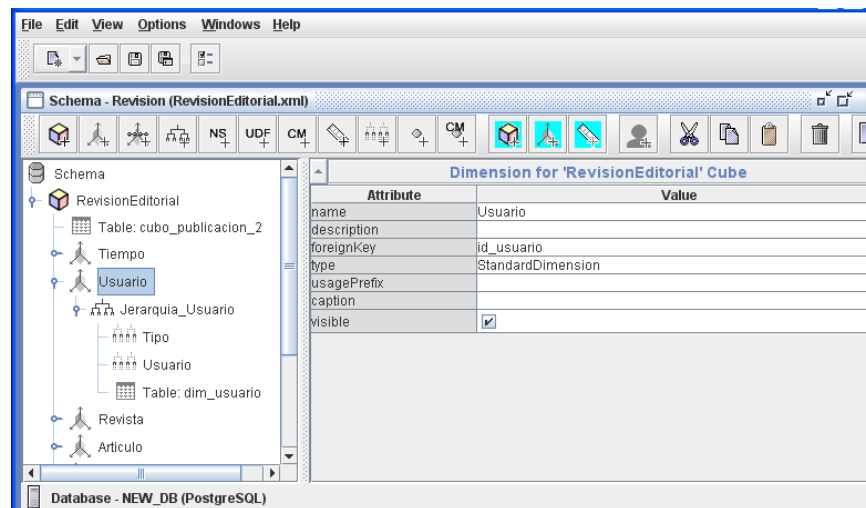


Figura 45. Creación de la dimensión Usuario

En la figura 46, se visualiza la definición de la dimensión revista y su jerarquía respectiva.

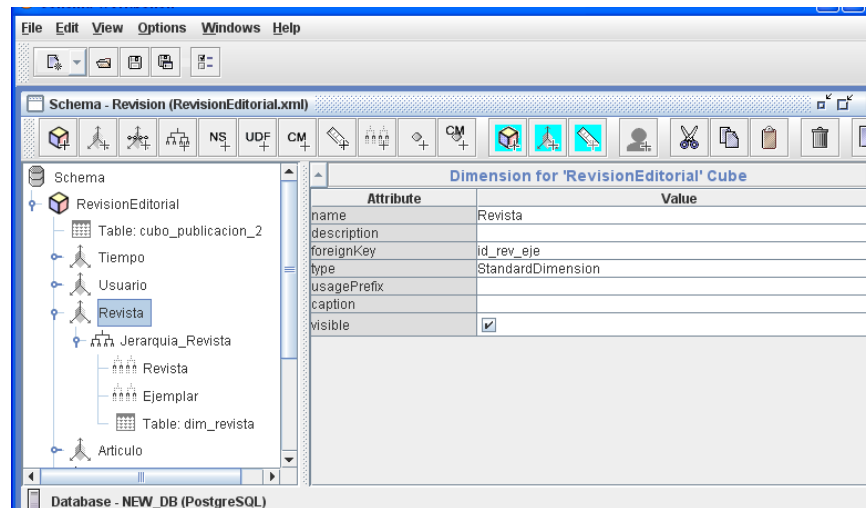


Figura 46. Creación de la dimensión Revista

Luego, en la figura 47, se presenta la definición de la dimensión artículo, detallando la jerarquía de la misma.

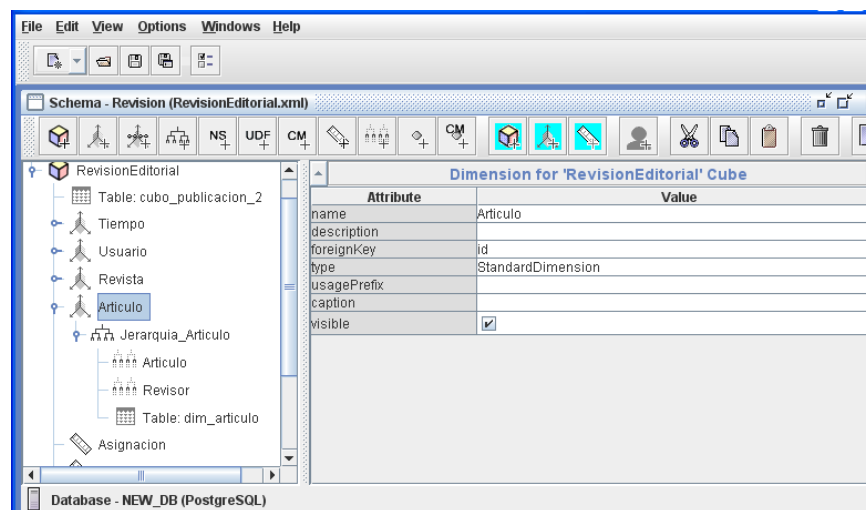


Figura 47. Creación de la dimensión Artículo

Por último, se declaran los identificadores del cubo, que fueron especificados en la primera fase del método de desarrollo. Como bien se puede observar, en la figura 48, se muestran los seis indicadores principales, que fueron establecidos como valores importantes o puntos clave de éxito para el monitoreo del proceso de Revisión Editorial.

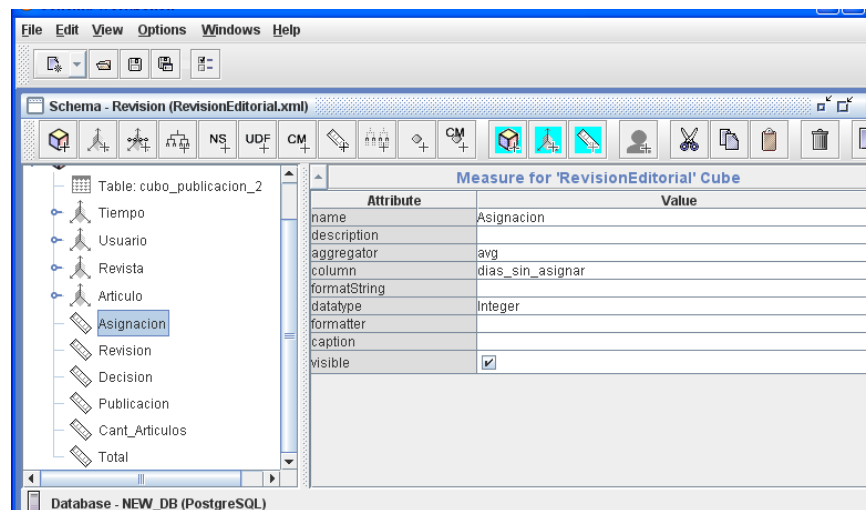


Figura 48. Creación de los hechos a medir

- El indicador “Asignación” calcula un promedio de cantidad de días, desde que el autor envió el artículo hasta que un revisor acepta evaluarlo.
- El indicador “Revisión” calcula un promedio de cantidad de días, desde que el revisor comienza la evaluación del artículo hasta que finaliza la misma.
- El indicador “Decisión” calcula el promedio de la cantidad de días, desde que el revisor termina la evaluación y el editor toma la decisión de aprobarlo.
- El indicador “Publicación” calcula un promedio de cantidad de días, desde que el editor acepta la publicación del artículo hasta la publicación del mismo.
- El indicador “Total” calcula un promedio de días, desde la fecha en que el autor envía el artículo hasta que el mismo se publica.

Por último, el cubo se publica en la plataforma *Pentaho BI*, para poder visualizar los datos del mismo en dicho servidor.

3.3.4 Fase 4: Análisis de Consistencia de Datos del Almacén de Datos

Una vez culminada la construcción del almacén de datos, se realiza el proceso de verificación de los datos, con el objetivo de asegurar su consistencia e integridad, así como también verificar si los datos transaccionales son correctos.

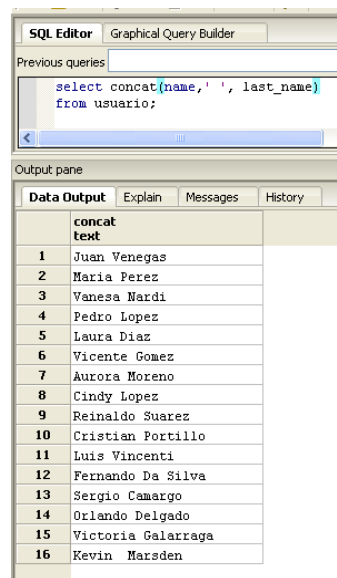
Este proceso de verificación de datos se lleva a cabo realizando consultas a los sistemas transaccionales y al almacén de datos. Luego, se verifica que los datos obtenidos en las consultas de los sistemas transaccionales correspondan con las consultas realizadas al almacén de datos. Así, se garantiza la carga exitosa de los datos.

A continuación, se realiza la verificación a cada una de las dimensiones y finalmente al cubo.

- **Dimensión Usuario**

Para realizar la comparación entre las tablas, se seleccionaron el nombre y apellido de los usuarios registrados, tanto de la tabla dimensión como de la tabla del sistema transaccional. Adicionalmente, se realizó la selección de los diferentes tipos de roles, en ambas tablas. En las figuras 49, 50, 51 y 52, se pueden observar los resultados arrojados en las consultas.

De esta manera, si tanto el primer par de consultas, como el segundo, retornan el mismo valor, entonces se asegura que los datos cargados en la tabla dimensión usuario, son consistentes, íntegros y correctos, respecto a los datos del sistema transaccional.



The screenshot shows an SQL Editor window with the following content:

```
SQL Editor Graphical Query Builder
Previous queries
select concat(name, ' ', last_name)
from usuario;
Output pane
Data Output Explain Messages History
concat
text
1 Juan Venegas
2 Maria Perez
3 Vanesa Nardi
4 Pedro Lopez
5 Laura Diaz
6 Vicente Gomez
7 Aurora Moreno
8 Cindy Lopez
9 Reinaldo Suarez
10 Cristian Portillo
11 Luis Vincenti
12 Fernando Da Silva
13 Sergio Camargo
14 Orlando Delgado
15 Victoria Galarraga
16 Kevin Marsden
```

Figura 49. Consulta 1 en el sistema transaccional de usuario

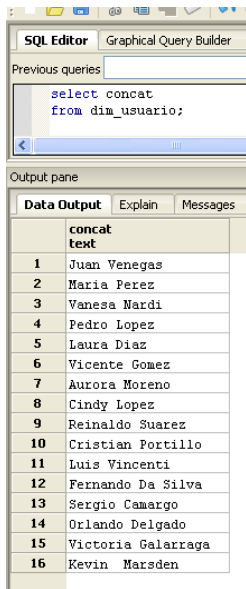


Figura 50. Consulta 1 de la dimensión Usuario

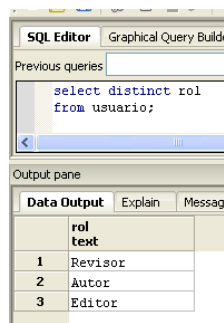


Figura 51. Consulta 2 en el sistema transaccional de usuario

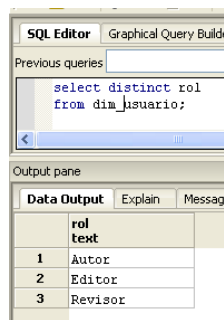


Figura 52. Consulta 2 en la dimensión Usuario

- **Dimensión Revista**

Para realizar la comparación entre las tablas, se seleccionó el nombre de la revista y la cantidad de ejemplares que esta tenía publicados, tanto de la tabla dimensión como de la tabla del sistema transaccional. Adicionalmente, se realizó la selección de los nombres de los ejemplares, en ambas tablas. En las figuras 53, 54, 55 y 56, se pueden observar los resultados arrojados en las consultas.

De esta manera, si tanto el primer par de consultas, como el segundo, retornan el mismo valor, entonces se asegura que los datos cargados en la tabla dimensión revista, son consistentes, íntegros y correctos, respecto a los datos del sistema transaccional.

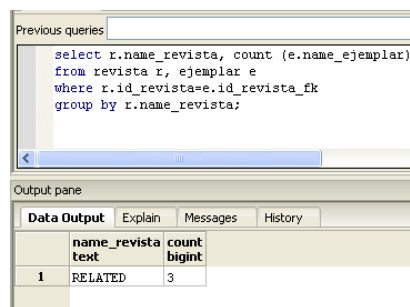


Figura 53. Consulta 1 en el sistema transaccional de revista

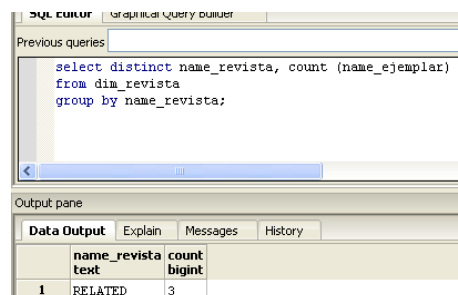


Figura 54. Consulta 1 en la dimensión Revista

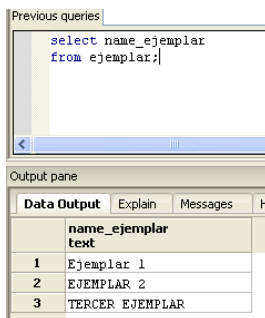


Figura 55. Consulta 2 en el sistema transaccional de ejemplar

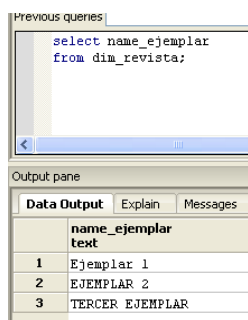


Figura 56. Consulta 2 en la dimensión Revista

- **Dimensión Artículo**

Para realizar la comparación entre las tablas, se realizó un conteo de la cantidad de artículos que habían sido publicados, en ambas tablas. En la figura 57 y 58, se puede observar los resultados arrojados en las consultas.

De esta manera, si el par de consultas retorna el mismo valor, entonces se asegura que los datos cargados en la tabla dimensión artículo, son consistentes, íntegros y correctos, respecto a los datos del sistema transaccional.

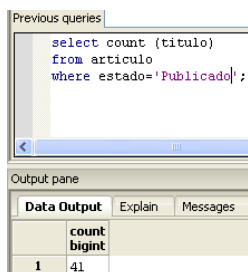


Figura 57. Consulta 1 del sistema transaccional en artículo

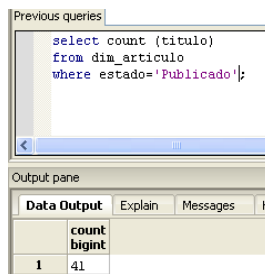


Figura 58. Consulta 1 en la dimensión Artículo

- **Cubo Publicación**

Para realizar la comparación entre la tabla del cubo y la tabla artículo de la base de datos transaccional. Tomando en cuenta que los datos almacenados en el cubo, registran los artículos que son publicados en un ejemplar de una revista digital; entonces, se cuenta el número de registros que existen en el cubo y la cantidad de artículos con el estado de “publicado” en la base de datos transaccional (ver figuras 59 y 60).

De esta manera, si ambas consultas retornan el mismo valor, entonces se asegura que los datos cargados en la tabla del cubo publicación, son consistentes, íntegros y correctos, respecto a los datos del sistema transaccional.

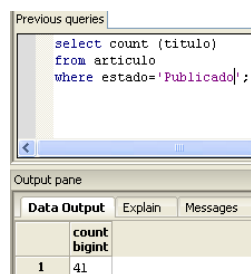


Figura 59. Consulta 1 en el sistema transaccional en artículos publicados

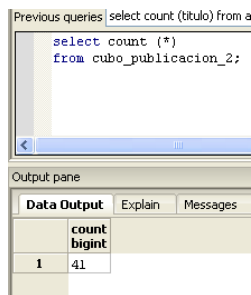


Figura 60. Consulta 1 en el cubo Publicación

Por lo tanto, se puede decir que el proceso de verificación de datos contenidos en las tablas dimensiones y la tabla de hechos, del almacén de datos, comprende una fase importante durante el desarrollo de la solución, ya que permite identificar posibles inconsistencias entre los datos. De esta manera, se procede a realizar correcciones necesarias, antes de seguir avanzando en el desarrollo.

3.3.5 Fase 5: Desarrollo de Indicadores

Ahora bien, ya certificada la consistencia e integridad de los datos contenidos en el almacén y habiendo publicado el esquema del cubo en el servidor Web de *Pentaho*, éste permite visualizar los datos del cubo, a través de la herramienta de *Pentaho Analyzer*. Adicionalmente, permite realizar cuadros de mandos, a través de la herramienta *Community Dashboard Editor* (CDE, por sus siglas).

Para la creación de análisis con el cubo, se deben realizar consultas, especificando filtros, seleccionando las dimensiones con los atributos y las métricas que se desean visualizar, es decir, *Pentaho Analyzer* permite modelar las representaciones en base a consultas, con el fin de crear un cuadro de mando integral que sirva de base para la generación de cuadros de mando. En la figura 61, se muestra el modo de realizar análisis con un cubo específico.

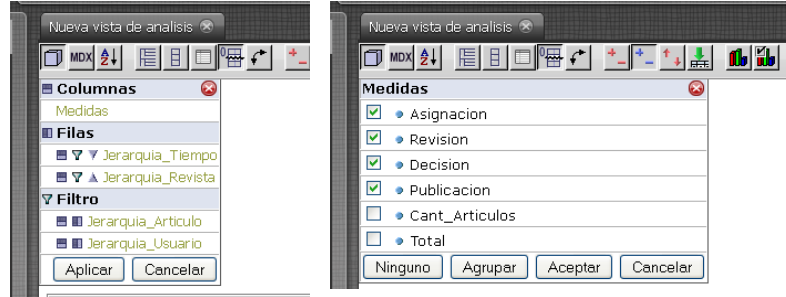


Figura 61. Visualización de un cubo en Pentaho Analyzer

Pentaho Analyzer utiliza expresiones multidimensionales, denominadas MDX (por el acrónimo de *MultiDimensional eXpressions*), como lenguaje de consulta para las bases de datos multidimensionales sobre cubos OLAP. Las bases de datos multidimensionales se pueden ver como bases de datos de una sola tabla, en donde cada dimensión y cada métrica tienen una columna.

Una vez seleccionadas las dimensiones con sus atributos, las métricas y habiendo especificado los filtros, la herramienta nos generará un código MDX, que servirá de base para la creación de los cuadros de mandos. En la figura 62, se muestra el código generado por dicha herramienta.

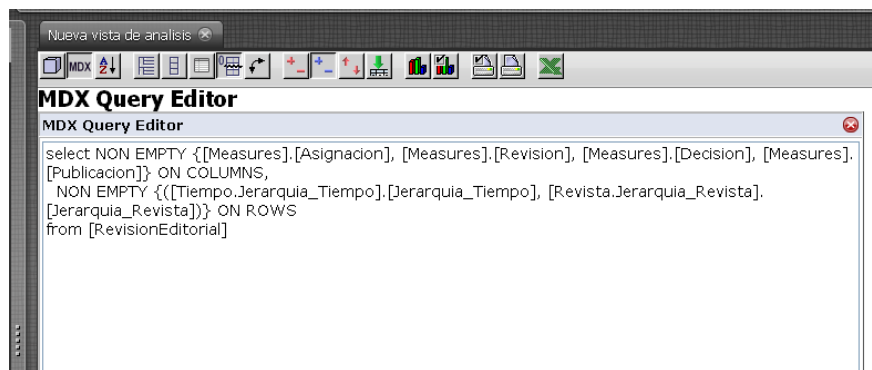


Figura 62. Consulta generada por el Pentaho Analyzer

Por otra parte, la herramienta *Community Dashboard Editor*, como ya se dijo, fue la herramienta utilizada para la creación del cuadro de mando. Dicha herramienta cuenta con tres pasos fundamentales para la creación de los mismos:

- a) Creación de la plantilla: En este primer paso se debe crear o seleccionar la plantilla, adaptada a las necesidades, que se desea utilizar para los cuadros de mando (ver figura 63).

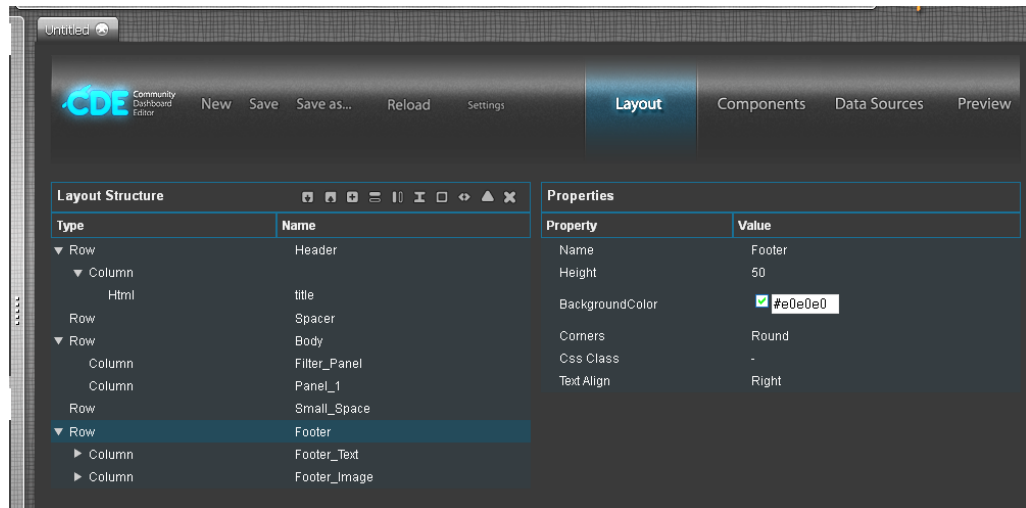


Figura 63. Creación de la plantilla en CDE

- b) Creación de los componentes: Es en este paso donde se selecciona el contenido que tendrá el cuadro de mando y se le da la ubicación a cada componente dentro de la plantilla (ver figura 64). Entre los componentes más importantes, se tienen: parámetros, funciones *javascript*, tablas, textos, comentarios, *xactions*, diferentes tipos de gráficos, elementos de selección, entre otros.

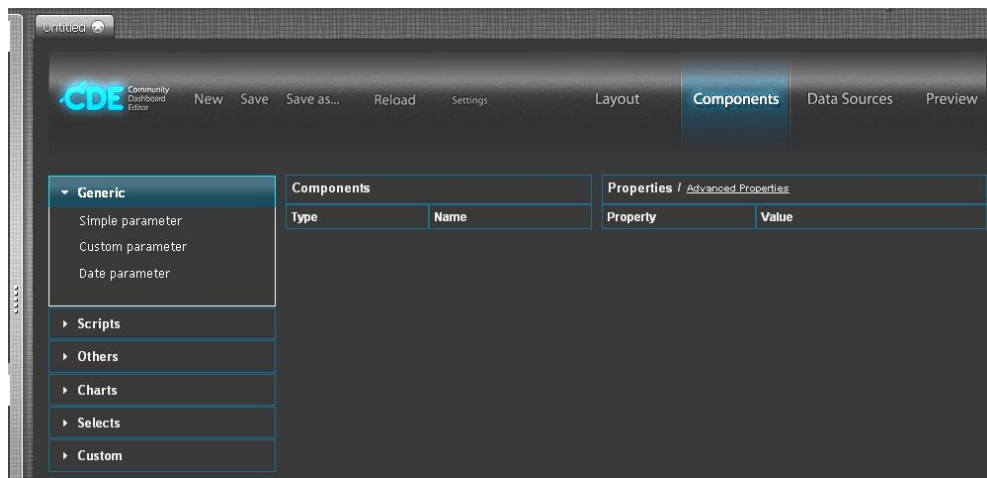


Figura 64. Creación de los componentes en CDE

- c) Creación de las consultas: En este paso es donde se establecen los orígenes de datos que contendrá cada componente seleccionado en el paso anterior (ver figura 65). Existen variedad de soluciones para realizar consultas y extraer la información, tales como: SQL, MDX, *Kettle*, etc.

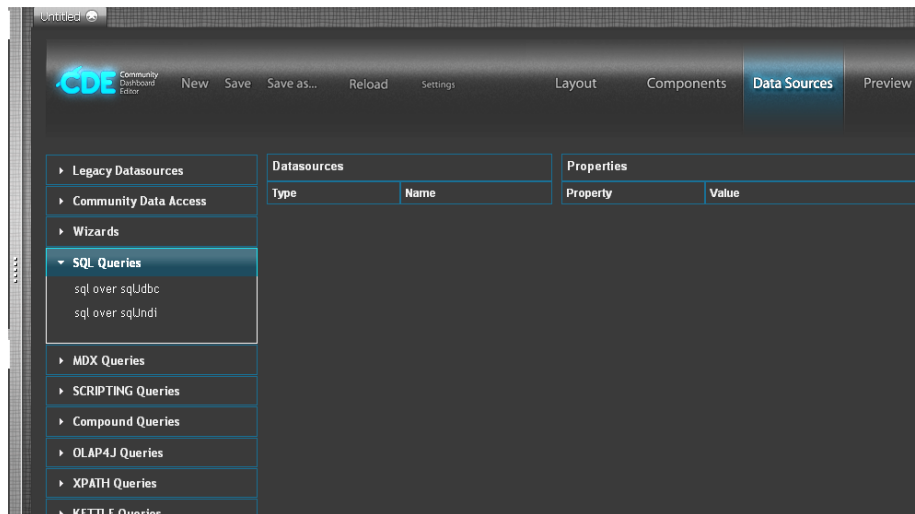


Figura 65. Creación de los orígenes de datos en CDE

Por lo tanto, para la creación del cuadro de mando, primeramente se diseñó la plantilla que se quería visualizar, de tal forma que todas las pestañas del cuadro de mando mantuvieran la misma. En la figura 66, se puede observar el boceto de la plantilla construida.

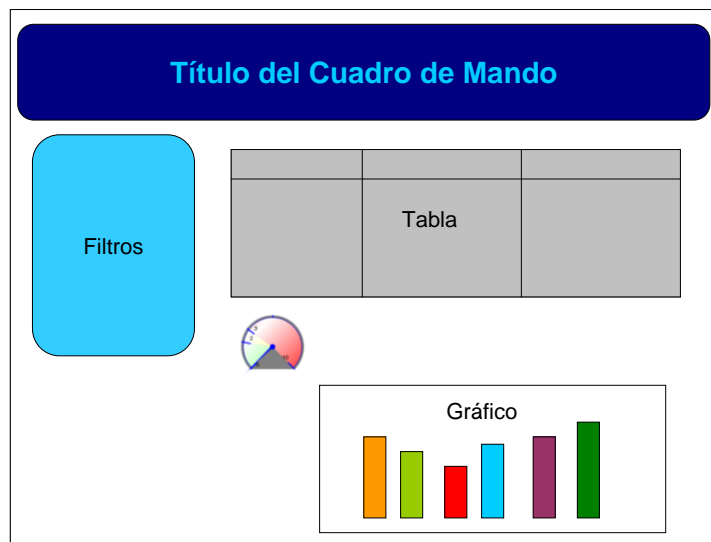


Figura 66. Plantilla de los Cuadros de Mando

En la figura 66, se muestra la composición de las pestañas del cuadro de mando, donde se aprecia que en la parte superior se ubica el título y en la parte lateral izquierda los filtros utilizados para hacer el cuadro de mando dinámico. Asimismo, se presenta una tabla donde se desplegarán los datos y debajo de ésta se ubica un componente en forma de tacómetro. Se entiende por tacómetro, el dispositivo que se usa para medir la velocidad de giro de un eje, pero en este caso medirá el valor que tiene el indicador, respecto a la alerta estratégica. Cuando el indicador se ubique en la zona verde del tacómetro, implicará que se encuentra en un valor esperado, en la zona amarilla, indicará que se encuentra en el margen del punto de éxito y en la zona roja reflejará que está por encima del valor que se espera. Por último se desplegará un gráfico que resuma la información de la tabla.

Cabe destacar que la herramienta *Community Dashboard Editor* no traía el tacómetro en su variedad de componentes. Sin embargo, el mismo fue desarrollado manualmente por medio de un archivo *xactions*, y luego pudo integrarse con la herramienta para la visualización del mismo.

Cada una de las pestañas creadas en el cuadro de mando para representar los indicadores, se diseñaron siguiendo la plantilla de la figura 66. Además, se realizaron consultas dependiendo del indicador que se quería mostrar. Los filtros implementados en estas pestañas se basan en los criterios de clasificación expuestos en las jerarquías de las dimensiones, como por ejemplo: año de envío, mes de envío, revista y ejemplar.

Por otra parte se diseñó una pestaña comparativa donde se muestran los valores de todos los indicadores construidos para un sistema de publicación específico. En esta pestaña se evalúan todos los artículos que ya han sido publicados por un sistema de publicación, de tal forma que se puedan observar las desviaciones de cada indicador. A diferencia de las otras pestañas, el único filtro implementado en ésta se relaciona con la revista donde los artículos están publicados. De esta forma, se puede comparar el indicador de asignación, revisión, decisión y publicación para detectar alguna variación en el tiempo.

Se describe a continuación la creación de las pestañas que componen cuadro de mando, para la visualización de información fundamental:

- **Tiempo de asignación de revisión**

Primeramente, se debe aclarar que tanto este indicador, como el resto, están filtrados por año de envío de artículo, mes de envío, revista y ejemplar donde se publicó. Las consultas utilizadas para dichos filtros se visualizan en las figuras 67, 68, 69 y 70.



Figura 67. Consulta para filtro del año



Figura 68. Consulta para filtro de mes



Figura 69. Consulta para filtro de revista



Figura 70. Consulta para filtro de ejemplar

Dado que este indicador se representa con la diferencia de días, entre la fecha que el autor envió el artículo y la fecha cuando se inicia la revisión, se debe seleccionar el nombre del artículo, la fecha de envío, la fecha de inicio de revisión y la diferencia de ambas para construir la tabla que proveerá dicha información. En la figura 71 se muestra dicha consulta.

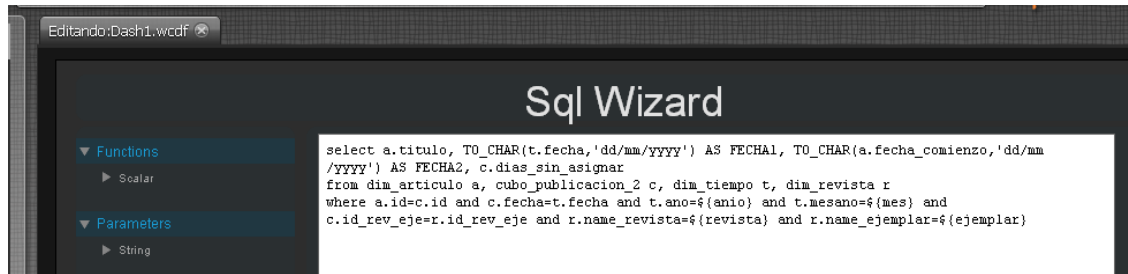


Figura 71. Consulta SQL para la tabla de Tiempo de asignación de revisión

Además, se necesita conocer el promedio total de los días. Para esto, se construyó un tacómetro que evaluara dicho indicador, si el valor se encuentra entre 0 y 2 está en lo esperado o por debajo, entre 2 y 3, ya es una señal de advertencia y por arriba de 3 es una señal de alerta, ya que se encuentra por encima de lo esperado. La consulta realizada para dicho tacómetro se puede observar en la figura 72.

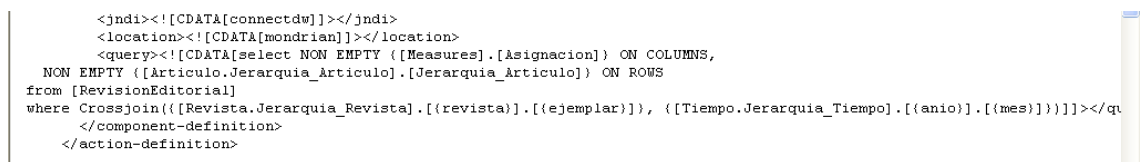


Figura 72. Consulta MDX para el tacómetro de Tiempo de asignación de revisión

Por último, es necesario construir un gráfico de barras que muestre de manera ordenada, de mayor a menor, la diferencia de días por artículo. La consulta utilizada para dicho gráfico se presenta en la figura 73.



Figura 73. Consulta MDX para el gráfico de Tiempo de asignación de revisión

- **Tiempo de revisión de un artículo**

Como bien se mencionó, las consultas utilizadas para los filtros de este indicador se pueden visualizar en las figuras 67, 68, 69 y 70.

Dado que este indicador se representa con la diferencia de días, entre la fecha que se inicia la revisión y la fecha cuando se finaliza la misma, se debe seleccionar el nombre del artículo, la fecha de inicio de revisión, la fecha fin de revisión y la diferencia de ambas para construir la tabla que proveerá dicha información. En la figura 74 se muestra dicha consulta.



Figura 74. Consulta SQL para la tabla de Tiempo de revisión de un artículo

Además, se necesita conocer el promedio total de los días. Para esto, se construyó un tacómetro que evaluara dicho indicador, si el valor se encuentra entre 0 y 2 está en lo esperado o por debajo, entre 2 y 3, ya es una señal de advertencia y por arriba de 3 es una señal de alerta, ya que se encuentra por encima de lo esperado. La consulta realizada para dicho tacómetro se puede observar en la figura 75.

```

~JMI~: [CDATA[mondrion]]~/JMI~
<location><![CDATA[mondrion]]></location>
<query><![CDATA[select NON EMPTY {[Measures].[Revision]} ON COLUMNS,
NON EMPTY {[Articulo.Jerarquia_Articulo].[Jerarquia_Articulo]} ON ROWS
from [RevisionEditorial]
where Crossjoin([Revista.Jerarquia_Revista].[{revista}].[{ejemplar}], {[Tiempo.Jerarquia_Tiempo].[{anio}].[{mes}]])]></q
</component-definition>
</action-definition>

```

Figura 75. Consulta MDX para el tacómetro de Tiempo de revisión de artículo

Por último, es necesario construir un gráfico de barras que muestre de manera ordenada, de mayor a menor, la diferencia de días por artículo. La consulta utilizada para dicho gráfico se presenta en la figura 76.

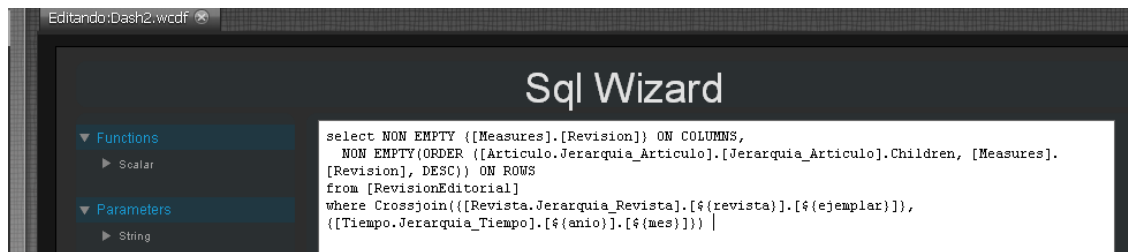


Figura 76. Consulta MDX para el gráfico de Tiempo de revisión de artículo

- **Tiempo de decisión final**

Como se sabe, las consultas utilizadas para los filtros de este indicador se pueden visualizar en las figuras 67, 68, 69 y 70.

Dado que este indicador se representa con la diferencia de días, entre la fecha fin de la revisión y la fecha cuando se toma una decisión final, se debe seleccionar el nombre del artículo, la fecha de fin de revisión, la fecha de decisión final y la diferencia de ambas para construir la tabla que proveerá dicha información. En la figura 77 se muestra dicha consulta.

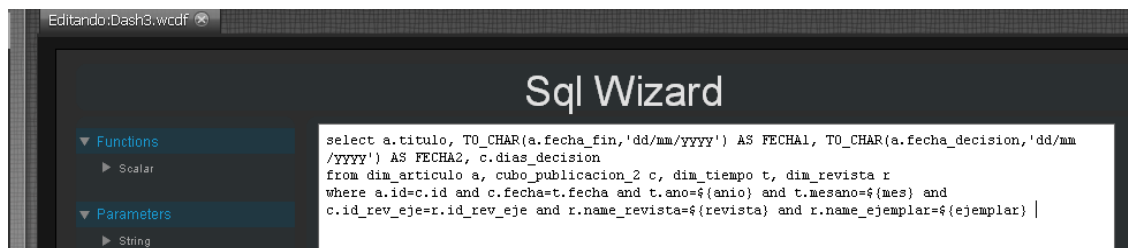


Figura 77. Consulta SQL para la tabla de Tiempo de decisión final

Además, se necesita conocer el promedio total de los días. Para esto, se construyó un tacómetro que evaluara dicho indicador, si el valor se encuentra entre 0 y 2 está en lo esperado o por debajo, entre 2 y 3, ya es una señal de advertencia y por arriba de 3 es una señal de alerta, ya que se encuentra por encima de lo esperado. La consulta realizada para dicho tacómetro se puede observar en la figura 78.

```

<jndi><![CDATA[connectdw]]></jndi>
<location><![CDATA[mondrian]]></location>
<query><![CDATA[select NON EMPTY {[Measures].[Decision]} ON COLUMNS,
NON EMPTY {[Articulo.Jerarquia_Articulo].[Jerarquia_Articulo]} ON ROWS
from [RevisionEditorial]
where Crossjoin({[Revista.Jerarquia_Revista].[revista].[ejemplar]}), {[Tiempo.Jerarquia_Tiempo].[(anio)].[(mes)]}]]></que
</component-definition>
</action-definition>

```

Figura 78. Consulta MDX para el tacómetro de Tiempo de decisión final

Por último, es necesario construir un gráfico de barras que muestre de manera ordenada, de mayor a menor, la diferencia de días por artículo. La consulta utilizada para dicho gráfico se presenta en la figura 79.

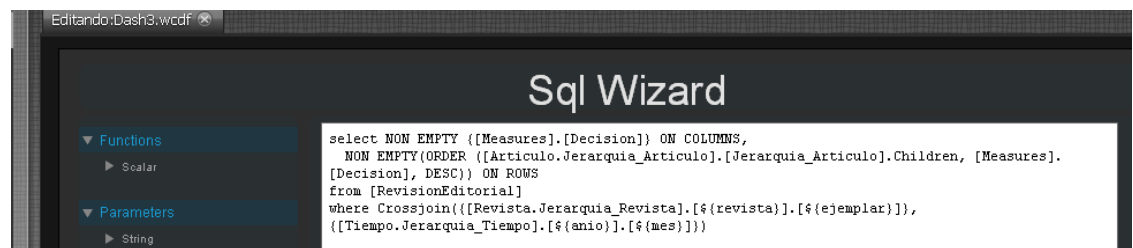


Figura 79. Consulta MDX para el gráfico de Tiempo de decisión final

- **Tiempo de publicación**

Las consultas utilizadas para los filtros de este indicador se pueden visualizar en las figuras 67, 68, 69 y 70.

Dado que este indicador se representa con la diferencia de días, entre la fecha de decisión final y la fecha cuando se publica el artículo, se debe seleccionar el nombre del artículo, la fecha de decisión final, la fecha de publicación del artículo y la diferencia de ambas para construir la tabla que proveerá dicha información. En la figura 80 se muestra dicha consulta.



Figura 80. Consulta SQL para la tabla de Tiempo de publicación

Además, se necesita conocer el promedio total de los días. Para esto, se construyó un tacómetro que evaluara dicho indicador, si el valor se encuentra entre 0 y 7 está en lo esperado o por debajo, entre 7 y 10, ya es una señal de advertencia y por arriba de 10 es una señal de alerta, ya que se encuentra por encima de lo esperado. La consulta realizada para dicho tacómetro se puede observar en la figura 81.

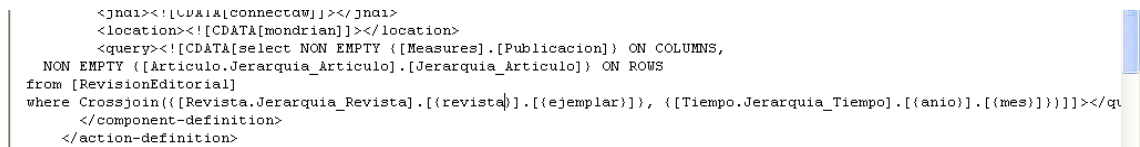


Figura 81. Consulta MDX para el tacómetro de Tiempo de publicación

Por último, es necesario construir un gráfico de barras que muestre de manera ordenada, de mayor a menor, la diferencia de días por artículo. La consulta utilizada para dicho gráfico se presenta en la figura 82.

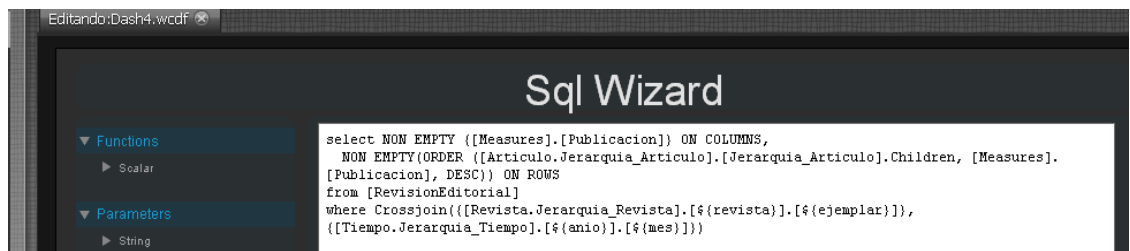


Figura 82. Consulta MDX para el gráfico de Tiempo de publicación

- **Tiempo final de publicación**

Las consultas utilizadas para los filtros de este indicador se pueden visualizar en las figuras 67, 68, 69 y 70.

Dado que este indicador se representa con la diferencia de días, entre la fecha de envío del artículo, por parte del autor, y la fecha cuando se publica el artículo, se debe seleccionar el nombre del artículo, la fecha de envío del mismo, la fecha de publicación y la diferencia de ambas para construir la tabla que proveerá dicha información. En la figura 83 se muestra dicha consulta.



Figura 83. Consulta SQL para la tabla de Tiempo de publicación total

Además, se necesita conocer el promedio total de los días. Para esto, se construyó un tacómetro que evaluara dicho indicador, si el valor se encuentra entre 0 y 20 está en lo esperado o por debajo, entre 20 y 30, ya es una señal de advertencia y por arriba de 50 es una señal de alerta, ya que se encuentra por encima de lo esperado. La consulta realizada para dicho tacómetro se puede observar en la figura 84.

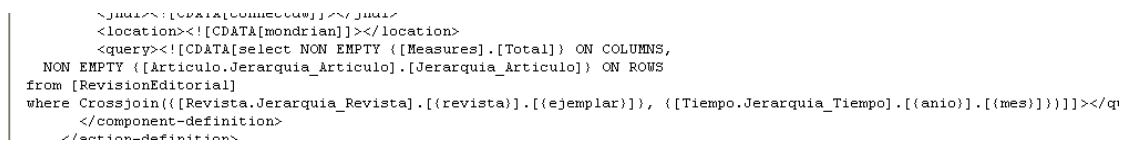


Figura 84. Consulta MDX para el tacómetro de Tiempo de publicación total

Por último, es necesario construir un gráfico de barras que muestre de manera ordenada, de mayor a menor, la diferencia de días por artículo. La consulta utilizada para dicho gráfico se presenta en la figura 85.

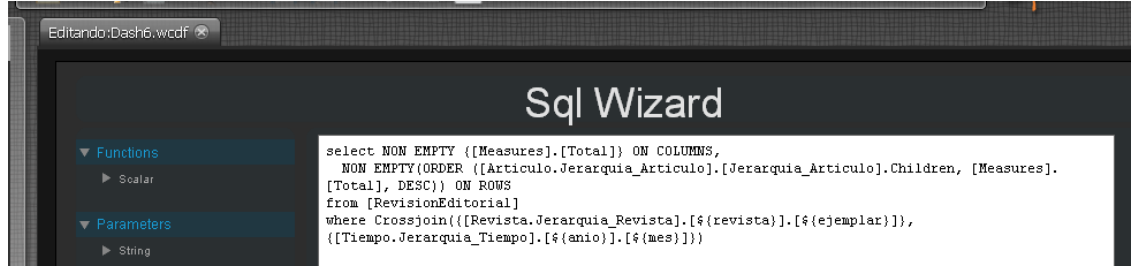


Figura 85. Consulta MDX para el gráfico de Tiempo de publicación Total

- **Comparación de indicadores**

En esta pestaña se muestran los indicadores de asignación, revisión, decisión y publicación de todos los artículos publicados en un sistema de publicación o revista, de tal forma que se puedan comparar y detectar desviaciones. Es por esto que para la construcción de la tabla es necesario seleccionar el nombre del artículo, la cantidad de días en asignación, en revisión, en decisión y en publicación. En la figura 86 se muestra dicha consulta.

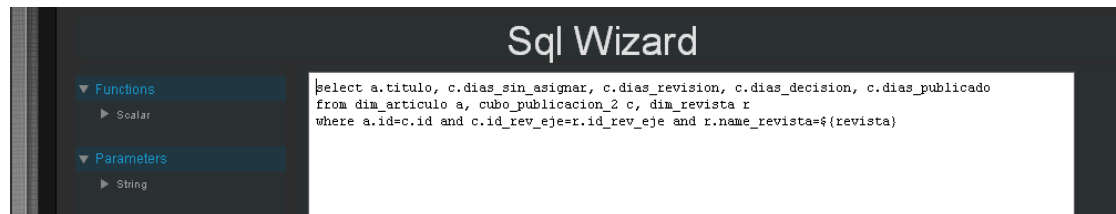


Figura 86. Consulta SQL para la tabla de comparación de indicadores

Además, se necesita conocer el promedio total de cada uno de los indicadores de la tabla. Para el promedio de cada indicador se construyó un tacómetro, en las figuras 87, 88, 89 y 90 se muestran las consultas utilizadas para mostrar la información.

```

<component-definition>
  <jndi>![CDATA[connectdw]]</jndi>
  <location>![CDATA[mondrian]]</location>
  <query>![CDATA[select NON EMPTY ([Measures].[Asignacion]) ON COLUMNS,
NON EMPTY Hierarchize([Articulo.Jerarquia_Articulo].[Jerarquia_Articulo]) ON ROWS
from [RevisionEditorial]
where ([Revista.Jerarquia_Revista].[Revista].[ejemplar])]]</query>
</component-definition>
</action-definition>

```

Figura 87. Consulta MDX para indicador de asignación

```

<jndi><![CDATA[connectdw]]></jndi>
<location><![CDATA[mondrian]]></location>
<query><![CDATA[select NON EMPTY ([Measures].[Revision]) ON COLUMNS,
NON EMPTY Hierarchize({[Articulo.Jerarquia_Articulo].[Jerarquia_Articulo]}) ON ROWS
from [RevisionEditorial]
where {[Revista.Jerarquia_Revista].[(revista)]}]]></query>
</component-definition>
</action-definition>

```

Figura 88. Consulta MDX para indicador de revisión

```

<jndi><![CDATA[connectdw]]></jndi>
<location><![CDATA[mondrian]]></location>
<query><![CDATA[select NON EMPTY ([Measures].[Decision]) ON COLUMNS,
NON EMPTY Hierarchize({[Articulo.Jerarquia_Articulo].[Jerarquia_Articulo]}) ON ROWS
from [RevisionEditorial]
where {[Revista.Jerarquia_Revista].[(revista)]}]]></query>
</component-definition>
</action-definition>

```

Figura 89. Consulta MDX para indicador de decisión final

```

<jndi><![CDATA[connectdw]]></jndi>
<location><![CDATA[mondrian]]></location>
<query><![CDATA[select NON EMPTY ([Measures].[Publicacion]) ON COLUMNS,
NON EMPTY Hierarchize({[Articulo.Jerarquia_Articulo].[Jerarquia_Articulo]}) ON ROWS
from [RevisionEditorial]
where {[Revista.Jerarquia_Revista].[(revista)]}]]></query>
</component-definition>
</action-definition>

```

Figura 90. Consulta MDX para indicador de publicación

3.3.6 Fase 6: Análisis de Consistencia de los Cuadros de Mando

Finalmente, se crean los cuadros de mando necesarios para representar la información fundamental de los indicadores a modelar, desarrollados en la primera fase. Ahora bien, para realizar esta tarea se utilizan las consultas creadas en la fase anterior, publicando su información tanto en componentes de selección donde se encuentran los filtros, como en la tabla, tacómetro y gráfico.

De esta manera, se describe a continuación el proceso necesario para crear los cuadros de mandos utilizados para representar los indicadores de gestión desarrollados y poder suministrar información fundamental que permita monitorear el rendimiento del proceso de Revisión Editorial.

- **Tiempo de asignación de revisión**

Para la creación del cuadro de mando relacionado a este indicador, se establecieron cuatro filtros, detallados anteriormente (año de envío, mes de envío, revista y ejemplar donde se publicó), y tres consultas que muestran una tabla con información, un tacómetro evaluador de puntos clave del proceso y un gráfico (ver figura 91).

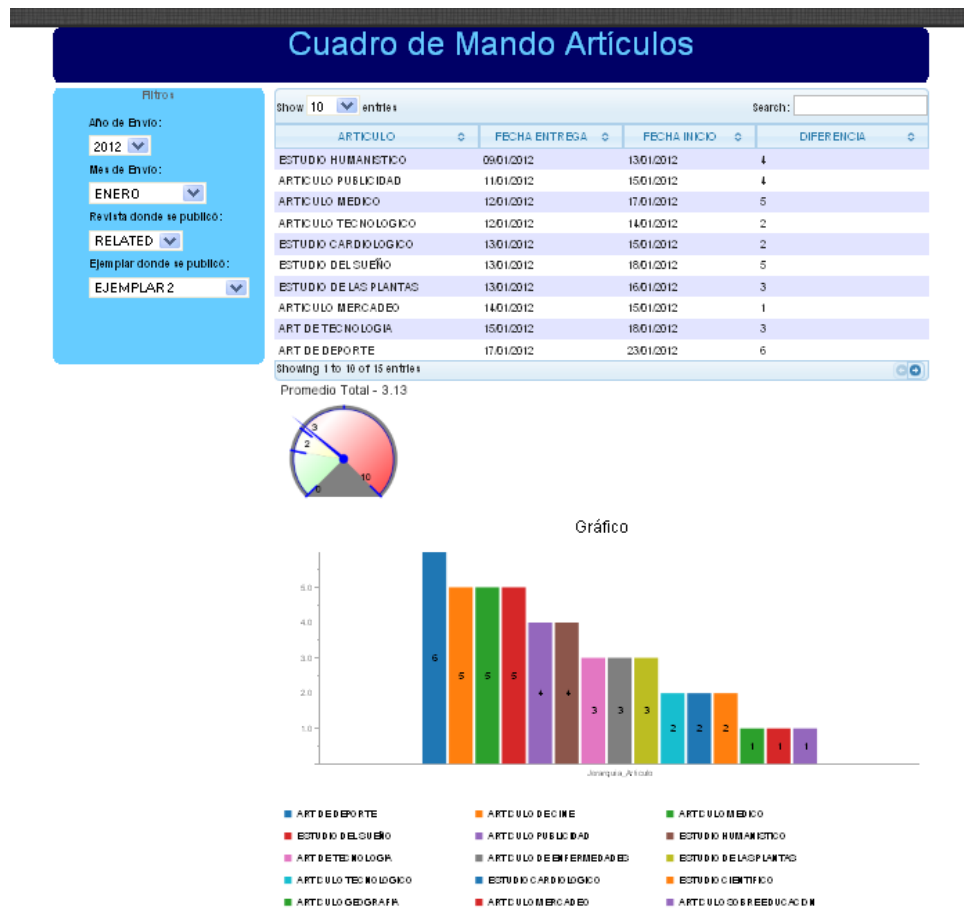


Figura 91. Vista de la pestaña de Tiempo de asignación de revisión

Análogo a la creación de la pestaña anterior, las siguientes pestañas del cuadro de mando se construyen utilizando el mismo procedimiento. Por lo tanto, el resultado de la construcción de las distintas pestañas que monitorean este proceso, se presentan a continuación:

- Tiempo de revisión de artículo (Ver figura 92)

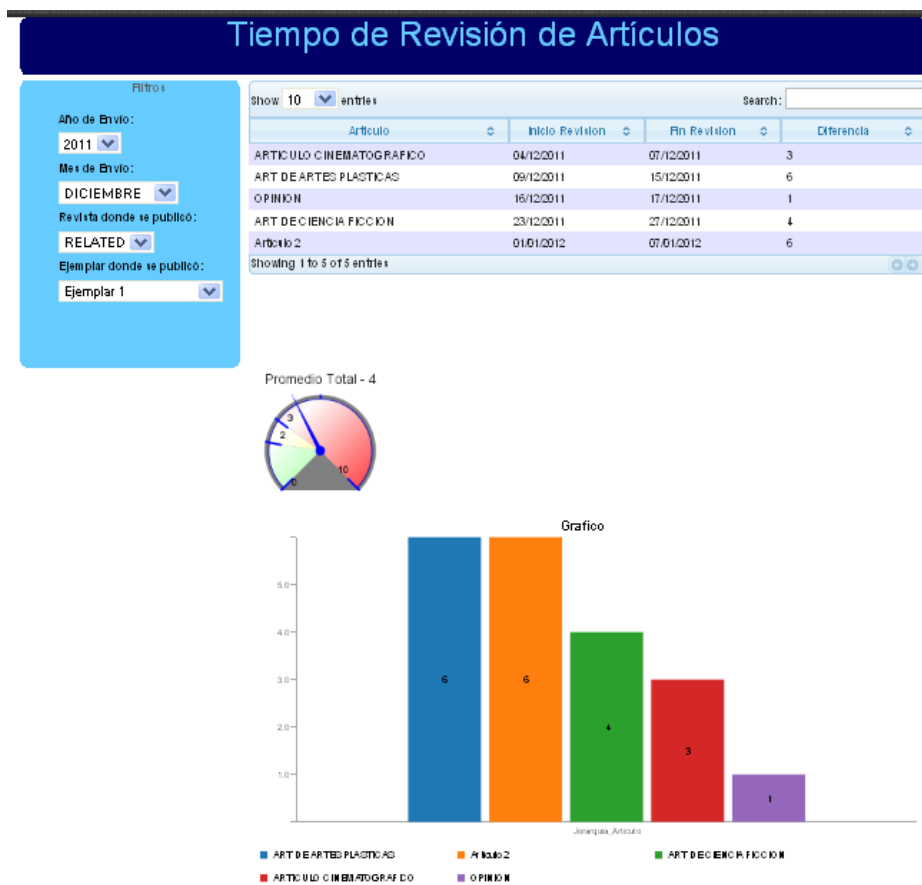


Figura 92. Vista de la pestaña de Tiempo de revisión de artículo

- Tiempo de decisión final (Ver figura 93)

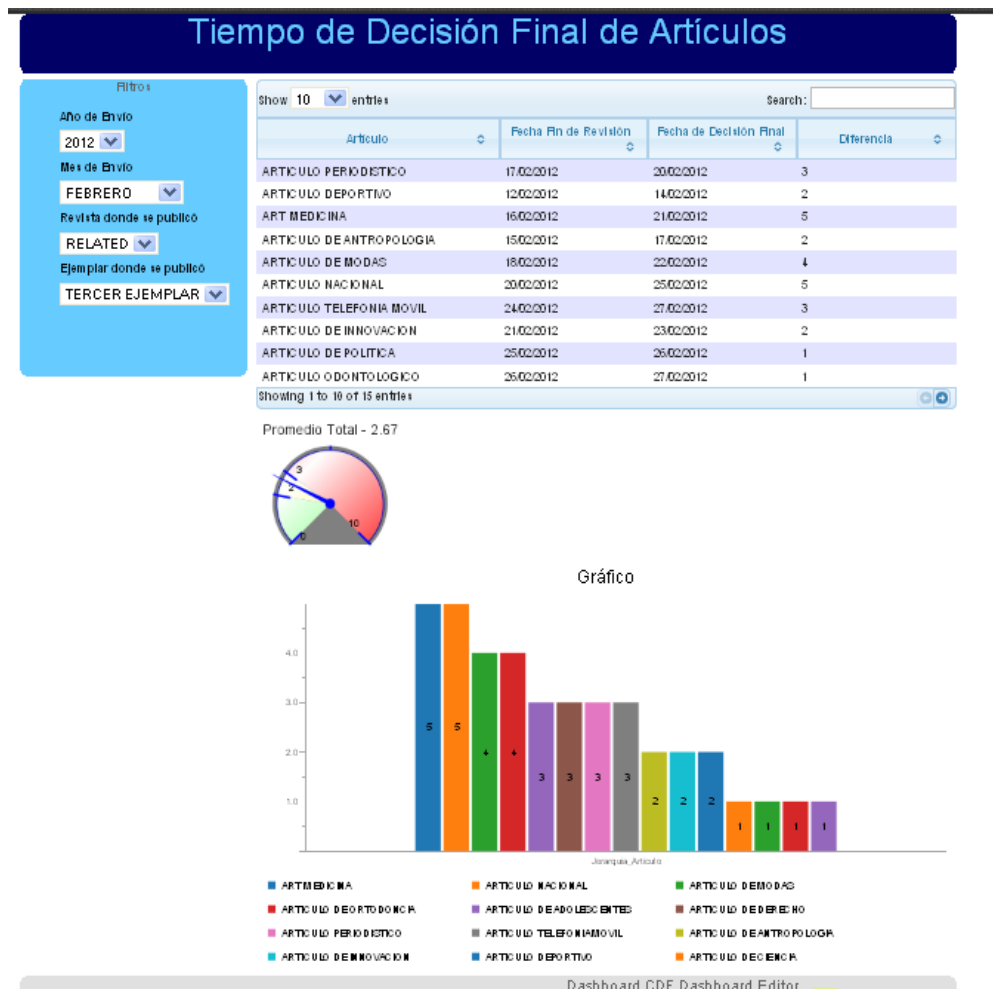


Figura 93. Vista de la pestaña de Tiempo de decisión final

- Tiempo de publicación (Ver figura 94)

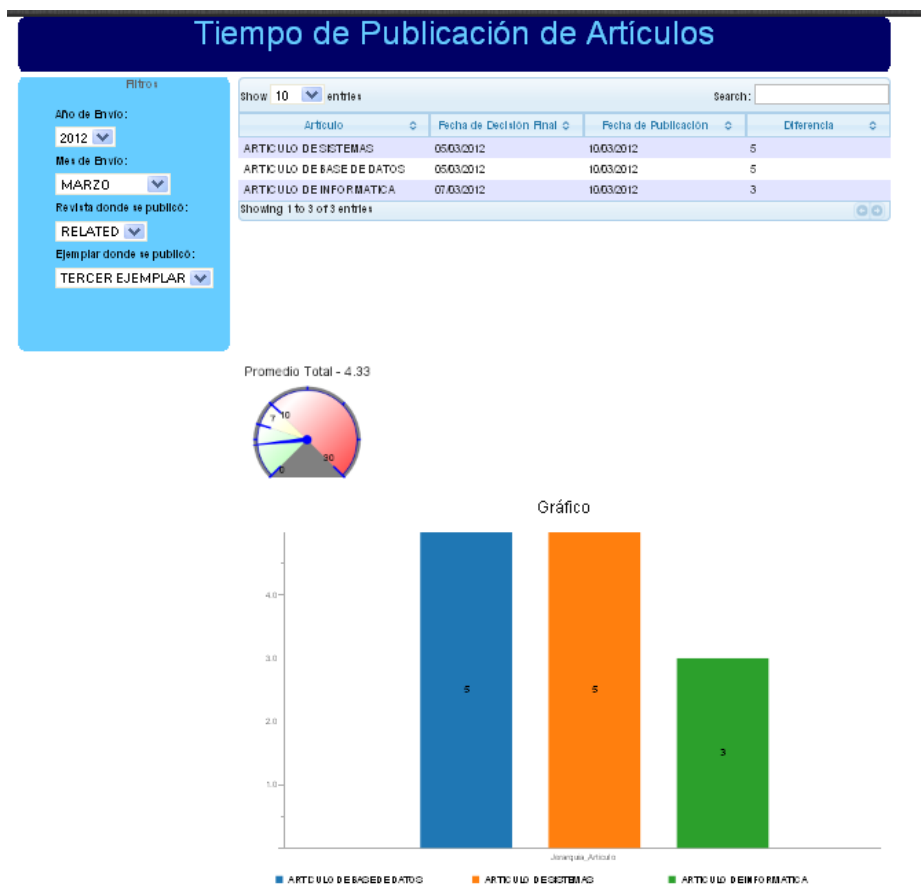


Figura 94. Vista de la pestaña de Tiempo de publicación

- Tiempo de publicación total (Ver figura 95)

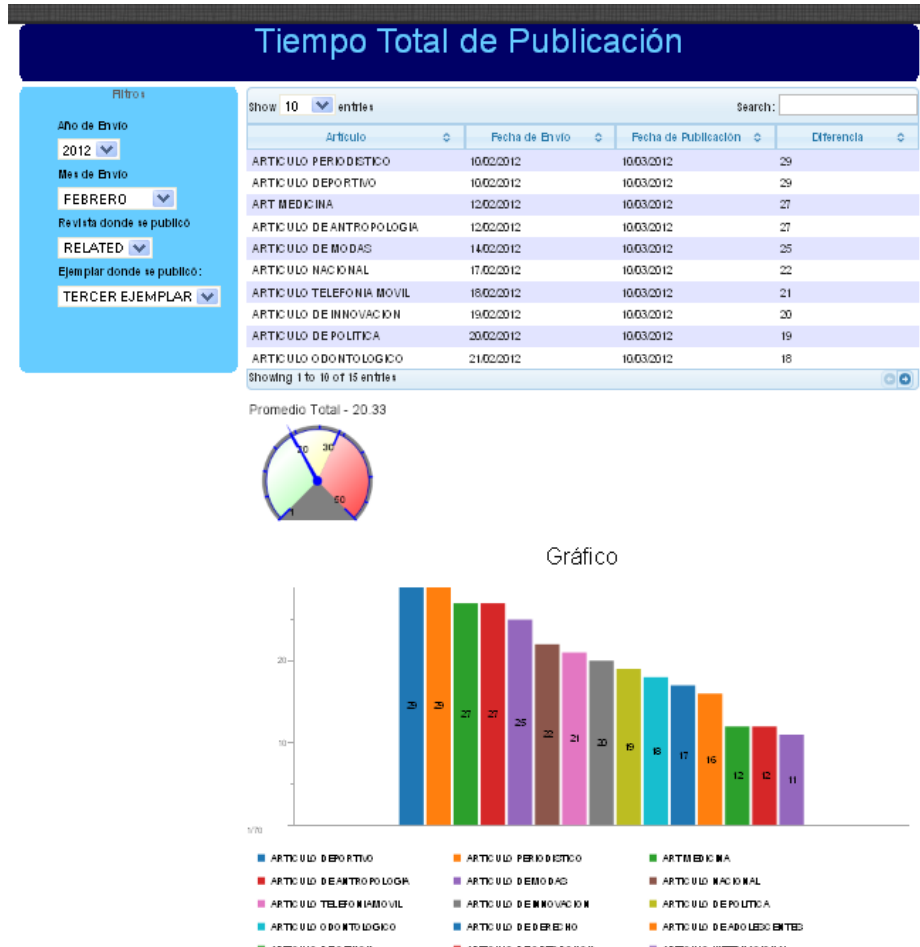


Figura 95. Vista de la pestaña de Tiempo total de publicación

- Comparación de Indicador por revista (Ver figura 96)

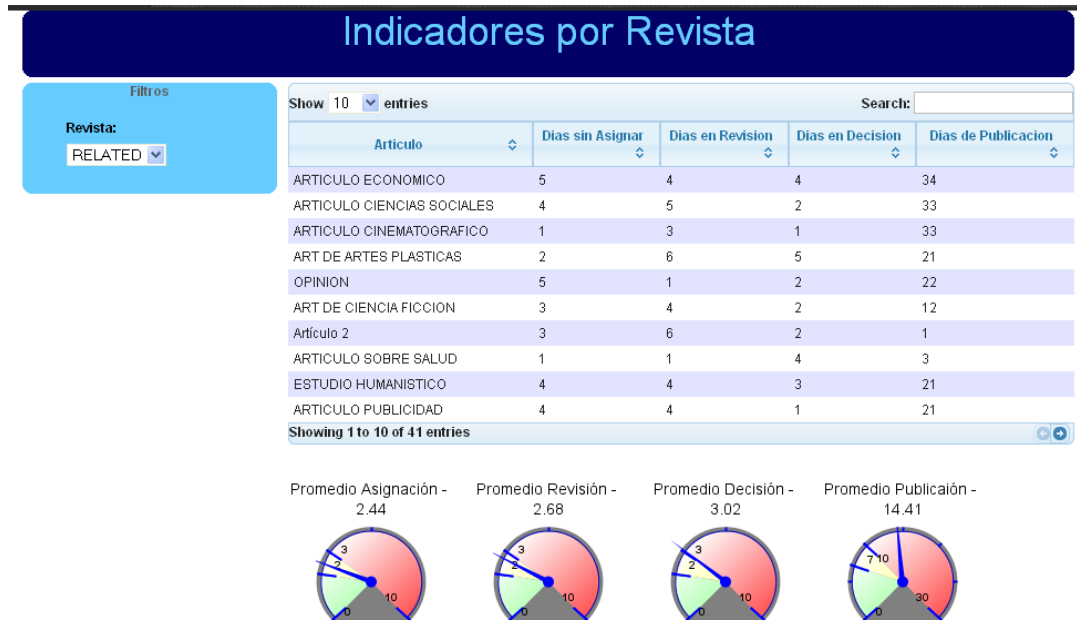


Figura 96. Vista de la pestaña de Comparación de indicadores por revista

Conclusiones

Una vez completado el objetivo principal del presente Trabajo Especial de Grado, se puede afirmar que se logró exitosamente el desarrollo y construcción de un sistema de indicadores de gestión que permitieran medir la eficiencia y efectividad del tiempo de duración del proceso de Revisión Editorial y determinar la distribución del mismo en los subprocesos.

En un principio, se estudió de manera exitosa el proceso de Revisión Editorial, llevado a cabo por revistas digitales, comprendiendo la forma en que trabajan y los datos que manejan. De esta forma, se pudo realizar de manera satisfactoria un proceso de levantamiento y análisis de requerimientos, que permitió establecer la información importante de dicho proceso. En conjunto con este levantamiento y análisis de información se establecieron un conjunto de indicadores de gestión orientados a medir la eficiencia del proceso.

Adicionalmente, se puede concluir que la plataforma Pentaho Business Intelligence Open Source, contiene un conjunto de componentes y herramientas que facilitaron en gran medida el desarrollo de la solución propuesta. Con ella se logró realizar la extracción, integración, estandarización, transformación y modelación de los datos; y así, cargarlos en un repositorio centralizado, para luego representarlos en forma de indicadores que facilitarán el análisis de los mismos, por parte de la gerencia del sistema de publicación. Esta herramienta se puede considerar una de las plataformas tecnológicas, entre las desarrolladas bajo Software Libre, más potentes para el procesamiento analítico de información estratégica dentro de una organización y para la solución de Inteligencia de Negocios. Sin embargo, dicha herramienta presentó limitaciones al momento de implementar los tacómetros en el cuadro de mando; por lo tanto, dichos tacómetros fueron desarrollados manualmente por medio de un archivo *xactions*.

Como bien se mencionó, el método propuesto por Beltran (2000) permitió la construcción de un conjunto de indicadores de gestión que permitiera monitorear el rendimiento del proceso de Revisión Editorial. Asimismo, el método utilizado para el desarrollo de la solución, considerado como una adaptación de los enfoques de Ralph Kimball y Bill Inmon, facilitó el diseño y desarrollo del almacén de datos, debido a la simplicidad y flexibilidad en cada una de las fases. Por lo que a medida que surgían nuevos requerimientos durante la etapa de desarrollo, se podían incorporar

los mismos sin necesidad de realizar cambios que afectaran la implementación ya establecida. Por lo tanto, se puede establecer que el método de desarrollo utilizado para llevar a cabo este Trabajo Especial de Grado cubrió las expectativas de los usuarios en cuanto a tiempo de respuesta sobre nuevos requerimientos.

La implementación de indicadores de gestión, a través una plataforma de Inteligencia de Negocios, trae consigo diferentes beneficios para las organizaciones, permitiendo facilitar el proceso analítico de los datos, relacionados a las transacciones cotidianas, realizado por la media y alta gerencia de una revista digital. De igual forma, proveen información relevante, oportuna y automatizada a la organización, como apoyo para la toma de decisiones asociada al proceso. Además, permite la posibilidad de realizar análisis comparativos para detectar tendencias y variaciones, a lo largo del tiempo. Por otra parte, ha permitido la disminución de errores humanos, ya sean por transcripción de datos, al momento de elaborar los reportes e indicadores, basados en la información contenida en los sistemas transaccionales, hojas de cálculo, archivos planos, entre otros; garantizando así, la integridad y calidad de los datos.

En conclusión, la solución de Inteligencia de Negocios, por medio de un sistema de indicadores de gestión, permite dar soporte a la media y alta gerencia de una organización, al momento de tomar decisiones, de tal forma que las mismas sean acertadas en base a situaciones reales por las que puede atravesar en un momento determinado. Igualmente, se extingue la necesidad en las organizaciones de tomar decisiones intuitivamente, que pueden no corresponder con la realidad del negocio. Y de esta forma, se genera una ventaja competitiva, de gran valor, para las empresas y organizaciones que deseen mejorar su productividad.

Referencias Bibliográficas

- 1Keydata.com (2012) 1KeyData. Recuperado el 3 de marzo de 2012, de: <http://www.1keydata.com>
- Ateico (2007). “Indicadores de Gestión”. Recuperado el 16 de Febrero de 2012: <http://www.aiteco.com/indicador.htm>
- Barruecos, J. (2000). “Revistas Electrónicas: Normalización y Perspectivas”. Universidad de Valencia, España. Recuperado el 22 de Abril de 2011, de <http://www.uv.es/~barrueco/badajoz.pdf>
- Beltran. J. (2000). “Indicadores de Gestión”. Bogotá: 3R Editores. 2da. Edición LTDA. Recuperado el 10 de febrero, de: <http://es.scribd.com/doc/2463985/Beltran-Jaramillo-Jesus-Mauricio-Indicadores-de-gestion>
- Campanario, J. (1998). “Peer review for journals as it stands today - Part 1”. Revista Española de Documentación Científica. Recuperado el 16 de febrero de 2012, de <http://www2.uah.es/jmc/ai4.pdf>
- Campanario, J. (2002). “El sistema de revisión por expertos (peer review): muchos problemas y pocas soluciones”. Revista Española de Documentación Científica. Recuperado el 16 de febrero de 2012, de <http://www2.uah.es/jmc/an24.pdf>
- Cano, J. (s.f.). “Business Intelligence: Competir con Información”. Recuperado el 14 de Febrero de 2012 desde <http://www.dataprix.com/files/BI-Business-Intelligence-Competir-Con-Informacion.pdf>
- Cary & Anderson (1992). “Harvard graphics 3: the complete reference”. Osborne McGraw-Hill. ISBN 0078817498.
- CINDOC & CSIC (2004). Centro de Información y Documentación Científica (CINDOC) y el Consejo Superior de Investigaciones Científicas (CSIC). “Revista Científicas Electrónicas: Estado del Arte”. Recuperado el 22 de Abril de 2011, de http://www.tecnociencia.es/e-revistas/especiales/revistas/pdf/e-revistas_informe.pdf
- Cortes E. (2007). Copyright SAP BI. “Business Intelligence”. Extraído el 14 de Marzo de 2012, de: <http://www.sap.com/spain>

- Cutro, L. (2010). *"Tendencias de BI Business Intelligence"*. Recuperado el 10 de Marzo de 2012, de: <http://www.dataprix.com/blogs/alfonsocutro/tendencias-bi-business-intelligence>
- Dataprix.com (2009) Dataprix. Recuperado el 4 de marzo de 2012, de: <http://www.dataprix.com/>
- De Pablos, Albarrán & Castilla (2001) *"El proceso de implantación del Data Warehouse en la organización: Análisis de un caso"*. Investigaciones Europeas de Dirección y Economía de la Empresa Vol. 4. Recuperado el 3 de marzo de 2012, de: <http://www.aedem-virtual.com/articulos/iedee/v04/043073.pdf>
- De Robbio, A. (2001). *"Revistas electrónicas en el ciberespacio"*. Recuperado el 18 de Abril de 2011, de <http://didattica.spbo.unibo.it/bibliotime/num-iv-3/derobbio.htm>
- EcuRed.cu (2012). EcuRed. Recuperado el 21 de Mayo de 2012, de: <http://www.ecured.cu/index.php>
- ETLTools.org (2009). ETL and Data Warehousing portal. Recuperado el 3 de marzo de 2012, de: <http://www.etltools.org/>
- Herrera, C. (2007). *"Todo lo que querías saber sobre Datawarehouse (I)"*. Recuperado el 21 de Mayo de 2012, de: <http://www.adictosaltrabajo.com/tutoriales/pdfs/datawarehouse.pdf>
- Hill, G. (2008). *"A guide to Enterprise Reporting"*. Recuperado el 8 de Marzo de 2012, de: <http://ghill.customer.netSPACE.net.au/reporting/definition.htm>
- ICSU (2001). Consejo Internacional para la Ciencia. *"Para publicaciones científicas"*. Recuperado el 16 de Febrero de 2012, de http://biblioteca.ucv.cl/guidelines/guidelines.htm#2_2_2
- Inmon, B. (1998). *"Data Mart Does Not Equal Data Warehouse"*. Recuperado el 13 de Marzo de 2012, de: <http://www.information-management.com/infodirect/19991120/1675-1.html>
- Instituto de Almacenamiento de Datos (2002) The Data Warehousing Institute Faculty Newsletter. Recuperado el 17 de Marzo de 2012, de: <http://www.mioaug.org/Presentations/BI-Readiness-White-Paper.pdf>

- Kaplan, R., & Norton, D. (2009.). *"Clarifying and communicating vision and strategy in action the BSC framework"*. Recuperado el 23 de Marzo de 2012, de: http://www.valuebasedmanagement.net/methods_balancedscorecard.html
- Kimball, R. (1997). *"A Dimensional Model Manifesto"*. Recuperado el 2 de Marzo de 2012, de: http://www.ralphkimball.com/html/articles_search/articles1997/9708d15.html
- Kimball, R., & Ross, M. (2002). *"The Data Warehouse Toolkit (2da ed.)"*. Wiley Computer Publishing.
- Larson, B. (2006). *Delivering Business Intelligence with Microsoft SQL Server 2005*. California: McGraw-Hill/Osborne.
- Laudon, K. & Laudon J. (2004). *"Sistemas de Información Gerencial"*. Pearson Prentice Hall, México. Recuperado el 3 de Marzo de 2012, de: <http://elibrolibre.com.ar/descargas/laudon.pdf>
- Lopez & Estrada (2006). *"La edición y publicaciones universitarias"*. Universidad Nacional Autónoma de México, México. Recuperado el 12 de Marzo de 2012, de: <http://www.edicion.unam.mx/html/2.html>
- Loshin, D. (2003). *"Business Intelligence: The Savvy Manager's Guide"*. Morgan Kaufmann Publishers.
- Moraga, M (2001). *"El modelo de datos jerárquico"*. Universidad de Castilla - La Mancha. Recuperado el 20 de Marzo de 2012, de: http://alarcos.inf-cr.uclm.es/doc/bda/doc/trab/T0001_MAMoraga.pdf
- Moss, L. & Atre, S. (2003). *"Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications"*. Boston: Addison Wesley.
- NASE (2010). *"Inteligencia de Negocios"*. Recuperado el 28 de Marzo de 2012, de: http://www.nase-it.com/index.php?option=com_content&view=article&id=84&Itemid=231
- Oppel, A. (2009). *"Databases"*. Estados Unidos de Norteamérica: McGraw Hill.
- Pachecho, J., Castañeda, W., & Caicedo, C. H. (2002). *"Indicadores Integrales de Gestión"*. Mc Graw Hill.
- Pentaho (2005). Recuperado el 9 de Febrero de 2012, de: <http://www.pentaho.com/>

- Pérez, C. (2006). *“Los Indicadores de gestión”*. Recuperado el 26 de Enero de 2012, de: <http://www.escuelagobierno.org/inputs/los%20indicadores%20de%20gestion.pdf>
- Rasmussen, N., Goldy, P., & Solli, P. (2002). *“Financial Business Intelligence”*. John Wiley & Sons, inc.
- Sánchez, R. (2003). *“To bi or not to bi”*. Recuperado el 28 de Marzo de 2012, de: <http://www.gestiopolis.com/recursos/documentos/fulldocs/ger1/tbi.htm>
- Saroka, R. (2002). *“Sistemas de Información en la era digital”*. Fundación OSDE, Buenos Aires. Recuperado el 16 de Febrero de 2012, de: http://www.fundacionosde.com.ar/pdf/biblioteca/Sistemas_de_informacion_en_la_era_digital-Modulo_I.pdf
- Strateby (2010) Stratebi Open Business Intelligence. *“Comparativa BI Open Source”*. Recuperado el 2 de febrero de 2012, de: http://www.stratebi.es/todobi/jun10/Comparativa_OSBI.pdf
- Williams, S., & Williams, N. (2007). *“The Profit Impact of Business Intelligence”*. Morgan Kaufmann Publishers.
- Wolff, C. (2002). *“Modelamiento Multidimensional”*. Recuperado el 9 de marzo de 2012, de: <http://www.inf.udec.cl/revista/ediciones/edicion4/modmulti.PDF>