



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

**TESIS PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN
SISTEMAS E INFORMÁTICA**

AUTOR: LOZADA PEÑAFIEL, XIMENA NATHALIE

CRUZ TAMAYO, HOLGER DAVID

**TEMA: ANÁLISIS, DISEÑO, CONSTRUCCIÓN E IMPLEMENTACIÓN DE
UN DATA WAREHOUSE PARA TOMA DE DECISIONES Y
CONSTRUCCIÓN DE LOS KPI, PARA LA EMPRESA
KRONOSCONSULTING CIA LTDA.**

DIRECTOR: ING. PÉREZ, WASHINGTON

CODIRECTOR: ING. DE LA TORRE, ANDRÉS

SANGOLQUÍ, MARZO 2014

UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

CERTIFICADO

Ing. Washington Pérez (DIRECTOR DE TESIS)

Ing. Andrés de la Torre (CODIRECTOR DE TESIS)

CERTIFICAN

Que el trabajo titulado “ANÁLISIS, DISEÑO, CONSTRUCCIÓN E IMPLEMENTACIÓN DE UN DATA WAREHOUSE PARA TOMA DE DECISIONES Y CONSTRUCCIÓN DE LOS KPI, PARA LA EMPRESA KRONOSCONSULTING CIA LTDA.”, realizado por la Srta. Ximena Nathalie Lozada Peñafiel y el Sr. Holger David Cruz Tamayo, ha sido guiado y revisado periódicamente y cumple normas estatutarias establecidas de la Universidad de las Fuerzas Armadas - ESPE.

Debido a que constituye un trabajo de excelente contenido científico, que coadyuvará a la aplicación de conocimientos y al desarrollo profesional, Sí recomiendan su publicación.

El mencionado trabajo consta de un documento empastado y un disco compacto el cual contiene los archivos en formato portátil de Acrobat (PDF). Se autoriza a la Srta. Ximena Nathalie Lozada Peñafiel y el Sr. Holger David Cruz Tamayo, que el material se entregue al Ing. Mauricio Campaña, en su calidad de Director de la Carrera.

Sangolquí, 14 de marzo de 2014

Ing. Washington Pérez
DIRECTOR

Ing. Andrés de la Torre
CODIRECTOR

UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

DECLARACIÓN DE RESPONSABILIDAD

Nosotros, Ximena Nathalie Lozada Peñafiel y Holger David Cruz Tamayo

DECLARAMOS QUE:

El proyecto de grado denominado “ANÁLISIS, DISEÑO, CONSTRUCCIÓN E IMPLEMENTACIÓN DE UN DATA WAREHOUSE PARA TOMA DE DECISIONES Y CONSTRUCCIÓN DE LOS KPI, PARA LA EMPRESA KRONOSCONSULTING CIA LTDA.”, ha sido desarrollado con base a una investigación exhaustiva, respetando derechos intelectuales de terceros, conforme las citas que constan el pie de las páginas correspondiente, cuyas fuentes se incorporan en la bibliografía.

Consecuentemente este trabajo es de nuestra autoría.

En virtud de esta declaración, nos responsabilizamos del contenido, veracidad y alcance científico del proyecto de grado en mención.

Sangolquí, 14 de marzo de 2014

Ximena Nathalie Lozada Peñafiel

/CC: 172351432-7

Holger David Cruz Tamayo

CC: 171475800-8

UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE
CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

AUTORIZACIÓN DE PUBLICACIÓN

Nosotros, Ximena Nathalie Lozada Peñafiel y Holger David Cruz Tamayo.

Autorizamos a la UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE, la publicación, en la biblioteca virtual de la Institución del trabajo “ANÁLISIS, DISEÑO, CONSTRUCCIÓN E IMPLEMENTACIÓN DE UN DATA WAREHOUSE PARA TOMA DE DECISIONES Y CONSTRUCCIÓN DE LOS KPI, PARA LA EMPRESA KRONOSCONSULTING CIA LTDA.”, cuyo contenido, ideas y criterios son de nuestra exclusiva responsabilidad y autoría.

Sangolquí, 14 de marzo de 2014

Ximena Nathalie Lozada Peñafiel
CC: 172351432-7

Holger David Cruz Tamayo
CC: 171475800-8

AGRADECIMIENTO

En primer lugar agradezco a nuestro padre Dios por estar vivo con salud y tener la oportunidad de vivir una vida maravillosa a lado de mi familia, amigos y compañeros.

La persona a quien nunca me cansaré de dar las gracias es a mi madre Janeth, quien ha estado conmigo toda mi vida y me ha inculcado los valores, la conducta y directrices que me han formado. Gracias a mi padre César, por enseñarme a ser más fuerte en la vida y darme los consejos necesarios en el momento necesario. A mi hermanita querida Verito que ha sido una fuente, muy cercana a mi corazón, de estímulo constante en cada uno de los pasos que doy.

Agradezco también a mí amada Melie y toda la familia, quienes me tuvieron paciencia, comprensión y amor todo este tiempo de arduas horas de trabajo y mucho sacrificio.

Gracias a mis profesores que han sembrado el conocimiento como parte de de mi formación profesional; y, en especial a mis directores de tesis, que mediante sus guías, hicieron del proyecto de tesis una realidad.

Holger Cruz

DEDICATORIA

Todo el trabajo realizado, es parte de una consecuencia de mucho sacrificio, fuerza y persistencia que en general se la dedico a toda mi familia quienes siempre han anhelado este paso durante mucho tiempo para continuar caminando por nuevos objetivos.

Holger Cruz

AGRADECIMIENTO

Agradezco a Dios por darme la vida y una familia unida que día a día han sido mi apoyo y fortaleza para seguir adelante y cumplir mis sueños.

Gracias a mis padres por sus consejos y por haberme formado en valores para ser una persona de bien, agradezco cada sacrificio que han hecho durante toda mi vida para darme lo mejor, por desvelarse junto a mí en mis largas noches de estudios y enseñarme que un tropezón no es caída, los amo.

A mis hermanas que han estado a mi lado ayudándome en todo mi proceso de formación personal y profesional, por ser mis amigas incondicionales y darme unos hermosos sobrinos que llenan de alegría mi vida y se han convertido en un motor más para seguir superándome.

Agradezco también al amor de mi vida David por ser quien ha estado a mi lado apoyándome y dándome fuerzas en todo momento, gracias por aguantar mi carácter y por ayudarme a culminar este sueño que lo inicié cinco años atrás y hoy es una realidad.

A mis amigos y profesores en especial mi director y codirector de tesis por su exigencia y ejemplo de superación profesional.

Ximena Lozada

DEDICATORIA

A mis padres Héctor Lozada e Isabel Peñafiel por ser el pilar fundamental en toda mi vida, por todo el esfuerzo y sacrificio que siempre han hecho para darme lo mejor y ser un ejemplo de perseverancia y constancia, además por brindarme todo su apoyo incondicional en cada momento.

A mis hermanas Jazmina y Verito por siempre estar a mi lado apoyándome y aconsejándome, por ser mis mejores amigas y mis segundas madres las quiero ñañitas.

A mis sobrinos Sebastián, Sofía, Valentina y Victoria por amarme y hacer que mis días sean más felices a su lado, por ser mi inspiración a lo largo de toda mi carrera.

Ximena Lozada

ÍNDICE DE CONTENIDOS

CAPÍTULO 1	1
1.1. ANTECEDENTES	1
1.2. PLANTEAMIENTO DEL PROBLEMA	4
1.3. JUSTIFICACIÓN	5
1.4. OBJETIVOS	7
1.4.1. OBJETIVO GENERAL	7
1.4.2. OBJETIVOS ESPECÍFICOS	7
1.5. ALCANCE	7
CAPÍTULO 2	10
FUNDAMENTACIÓN TEÓRICA	10
2.1. BASES TEÓRICAS	10
2.1.1. DEFINICIONES IMPORTANTES	10
2.1.1.1. OLTP	10
2.1.1.2. OLAP	11
2.1.1.3. DATA WAREHOUSING	11
2.1.1.4. DATA WAREHOUSE	12
2.1.1.5. OPERATIONAL DATA STORE (ODS)	13
2.1.1.6. DATA MART	13
2.2. INTELIGENCIA DE NEGOCIOS (BUSINESS INTELLIGENCE BI)	14
2.2.1. ORÍGENES	14
2.2.2. DEFINICIÓN	17
2.2.3. PROCESO PARA REALIZAR BUSINESS INTELLIGENT	18
2.3. BODEGA DE DATOS (DATA WAREHOUSE)	20
2.3.1. DEFINICIÓN	20
2.3.2. FUNCIONALIDAD	20
2.3.3. CARACTERÍSTICAS	21
UN DATA WAREHOUSE ES:	21
• ORIENTADO AL NEGOCIO	21
• INTEGRADO	21
• VARIANTE EN EL TIEMPO	21
• NO VOLÁTIL	21
• REDUNDANTE	21
2.3.3.1. ORIENTADA AL NEGOCIO	21
2.3.3.2. INTEGRADA	22
2.3.3.3. VARIANTE EN EL TIEMPO	23
2.3.3.4. NO VOLÁTIL	24
2.3.3.5. REDUNDANCIA	25
2.3.4. ESTRUCTURA	26
2.3.5. OLTP vs DW	27
2.4. ARQUITECTURA DATA WAREHOUSING	29
2.4.1. OLTP	30
2.4.2. LOAD MANAGER	31
2.4.2.1. EXTRACCIÓN	32
2.4.2.2. TRANSFORMACIÓN	33
2.4.2.2.1. TIPOS DE DATOS Y MEDIDA	34
2.4.2.2.2. CODIFICACIÓN	35
2.4.2.2.3. FUENTES MÚLTIPLES	36
2.4.2.2.4. CONVENCIONES DE NOMBRAMIENTO	37
2.4.2.2.5. LIMPIEZA DE DATOS	38

2.4.2.3.	CARGA.....	39
2.4.3.	DATA WAREHOUSE MANAGER.....	39
2.4.3.1.	BASE DE DATOS MULTIDIMENSIONAL.....	41
2.4.3.2.	TABLAS DE DIMENSIÓN.....	43
2.4.3.2.1.	TABLAS DE DIMENSIÓN TEMPORALIDAD.....	44
2.4.3.3.	TABLAS DE HECHOS.....	45
2.4.4.	QUERY MANAGER.....	47
2.4.5.	HERRAMIENTAS DE CONSULTA Y ANÁLISIS.....	48
2.4.6.	USUARIOS.....	49
2.4.7.	VENTAJAS Y DESVENTAJAS DEL DATA WAREHOUSING.....	50
2.5.	INDICADOR DE GESTIÓN (KPI).....	53
2.5.1.	IMPORTANCIA DE LOS INDICADORES DE GESTIÓN EMPRESARIAL.....	53
2.5.2.	LA CLAVE DEL PROCESO DE SELECCIÓN DE KPI.....	54
2.5.3.	CLASIFICACIÓN.....	56
2.5.4.	SISTEMA DE INDICADORES.....	57
2.6.	ANÁLISIS DE LAS METODOLOGÍAS PARA BUSINESS INTELLIGENCE.....	59
2.7.	METODOLOGÍA HEFESTO.....	61
2.7.1.	CARACTERÍSTICAS.....	62
2.7.2.	PASOS Y APLICACIÓN DE LA METODOLOGÍA.....	63
2.7.2.1.	PASO 1. ANÁLISIS DE REQUERIMIENTOS.....	64
2.7.2.1.1.	IDENTIFICAR PREGUNTAS.....	64
2.7.2.1.2.	IDENTIFICAR INDICADORES Y PERSPECTIVAS.....	66
2.7.2.1.3.	MODELO CONCEPTUAL.....	66
2.7.2.2.	PASO 2. ANÁLISIS DE LOS OLTP.....	67
2.7.2.2.1.	CONSTRUCCIÓN DE LOS INDICADORES.....	67
2.7.2.2.2.	ESTABLECER CORRESPONDENCIA.....	69
2.7.2.2.3.	NIVEL DE GRANULARIDAD.....	69
2.7.2.2.4.	MODELO CONCEPTUAL AMPLIADO.....	71
2.7.2.3.	PASO 3. MODELO LÓGICO DEL DW.....	72
2.7.2.3.1.	TIPO DE MODELO LÓGICO.....	72
2.7.2.3.2.	DISEÑO TABLAS DE DIMENSIONES.....	73
2.7.2.3.3.	TABLAS DE HECHOS.....	76
2.7.2.3.4.	UNIONES.....	79
2.7.2.4.	PASO 4. INTEGRACIÓN DE DATOS.....	80
2.8.	HERRAMIENTAS.....	82
2.8.1.	MICROSOFT SQL SERVER.....	82
2.8.2.	MICROSOFT VISUAL ESTUDIO.....	83
2.8.3.	SQL SERVER MANAGEMENT STUDIO.....	84
2.8.4.	DEVEXPRESS SUITE.....	85
2.8.4.1.	DXPERIENCE ENTERPRISE.....	85
2.8.4.2.	DXTREME ENTERPRISE.....	85
2.8.5.	POWERDESIGNER.....	85
CAPÍTULO 3.....	87	
ANÁLISIS Y DISEÑO DEL PROYECTO.....	87	
3.1.	METODOLOGÍA HEFESTO.....	87
3.1.1.	ETAPA I). ANÁLISIS DE REQUERIMIENTOS.....	87
3.1.1.1.	IDENTIFICAR PREGUNTAS.....	87
3.1.1.2.	IDENTIFICAR INDICADORES Y PERSPECTIVAS.....	89
3.1.1.3.	MODELO CONCEPTUAL.....	90
3.1.2.	ETAPA II). ANÁLISIS DE OLTP.....	91
3.1.2.1.	CONSTRUCCIÓN DE LOS INDICADORES.....	91
3.1.2.1.1.	MONTO DE CARTERA VENCIDA RECUPERADA.....	91
3.1.2.1.2.	MONTO DE CARTERA VENCIDA NO RECUPERADA.....	92

3.1.2.1.3.	PAGOS EFECTUADOS.	92
3.1.2.1.4.	EFICIENCIA DEL EJECUTIVO EN RECUPERAR LA CARTERA.	92
3.1.2.1.5.	EFICIENCIA DEL EJECUTIVO EN CONTACTAR LOS CLIENTES.	93
3.1.2.1.6.	EFICIENCIA DE LA EMPRESA EN RECUPERAR LA CARTERA.	93
3.1.2.1.7.	EFICIENCIA DE LA EMPRESA EN CONTACTAR A LOS CLIENTES.	94
3.1.2.2.	ESTABLECER CORRESPONDENCIA.	94
3.1.2.3.	NIVEL DE GRANULARIDAD	104
3.1.2.4.	MODELO CONCEPTUAL AMPLIADO.	110
3.1.3.	ETAPA III). MODELO LÓGICO DEL DATA WAREHOUSE.	111
3.1.3.1.	TIPO DE MODELO LÓGICO DEL DATA WAREHOUSE.	111
3.1.3.2.	TABLAS DE DIMENSIÓN.	111
3.1.3.3.	TABLAS DE HECHOS.	114
3.1.3.4.	UNIONES.	115
3.2.	ANÁLISIS KPI.	115
3.2.1.	DEFINICIÓN DEL OBJETIVO DEL ESTUDIO.	115
3.2.2.	MATRIZ DE KPI.	117
CAPÍTULO 4	118
DESARROLLO Y CONSTRUCCIÓN	118
4.1.	APLICACIÓN METODOLOGÍA HEFESTO.	118
4.1.1.	ETAPA IV). INTEGRACIÓN DE DATOS.	118
4.1.1.1.	CARGA INICIAL.	119
4.1.1.1.1	PROCESO 1. CARGA FUENTES EXTERNAS AL ÁREA DE DESEMBARCO.	119
4.1.1.1.2	PROCESO 2. CARGA DE LAS TABLAS DE DIMENSIÓN AL ÁREA DEL DWH.	131
4.1.1.1.3	PROCESO 3. CARGA DE LA TABLA DE HECHOS AL ÁREA DEL DWH.	132
4.1.1.2.	ACTUALIZACIÓN.	133
4.1.1.3.	CONSTRUCCIÓN DEL CUBO DE RECUPERACIÓN DE CARTERA.	134
4.1.1.3.1	CREACIÓN DEL PROYECTO ANALYSIS SERVICES MULTIDIMENSIONAL.	135
4.1.1.3.2	CONFIGURACIÓN DE LA CONEXIÓN A LA BASE DEL DWH.	136
4.1.1.3.3	CONFIGURACIÓN DE LA VISTA BASE DEL CUBO MULTIDIMENSIONAL.	137
4.1.1.3.4	CREACIÓN DE LAS DIMENSIONES DEL CUBO MULTIDIMENSIONAL.	138
4.1.1.3.5	CREACIÓN DEL CUBO MULTIDIMENSIONAL.	141
4.1.1.3.6	CREACIÓN DE JERARQUÍAS.	144
4.1.1.3.7	CREACIÓN DE INDICADORES.	145
4.1.1.3.8	CREACIÓN DE KPI.	147
4.2.	APLICACIONES PARA USUARIOS FINALES.	149
4.2.1.	TABLERO DE CONTROL.	149
4.2.2.	APLICACIÓN WEB.	151
CAPÍTULO 5	155
CONCLUSIONES Y RECOMENDACIONES	155
5.1.	CONCLUSIONES.	155
5.2.	RECOMENDACIONES	156
BIBLIOGRAFÍA	157

LISTADO DE TABLAS

TABLA 1. ANÁLISIS DATA WAREHOUSE NO REDUNDANTE	26
TABLA 2. ANÁLISIS ESTRUCTURA DWH. (DARIO, 2009).....	27
TABLA 3. OLTP VS. DWH.....	28
TABLA 4. CARACTERÍSTICAS DE LA LIMPIEZA DE DATOS. (DARIO B. , 2009).....	38
TABLA 5. COMPARACIÓN METODOLOGÍAS HEFESTO VS. SAS.	61
TABLA 6. TABLA DE DIFERENCIACIÓN ENTRE INDICADORES Y REPORTES.	89
TABLA 7. TABLA DE IDENTIFICADORES Y PERSPECTIVAS A SER ANALIZADAS.....	90
TABLA 8. DESCRIPCIÓN DE LOS KPI Y SU PERSPECTIVA.....	116
TABLA 9. DESCRIPCIÓN DE LOS KPI Y SU PERSPECTIVA.....	117
TABLA 10. ESQUEMA DE PROCESOS ETL DE CARGA FUENTE EXTERNA.	119
TABLA 11. ESQUEMA DE PROCESOS ETL DE CARGA TABLAS DIMENSIONES.....	131
TABLA 12. ESQUEMA DE PROCESOS ETL DE CARGA TABLA DE HECHOS.....	132

LISTADO DE FIGURAS

FIGURA 1. ORGANIGRAMA DE LA EMPRESA TOPNOTCH BUSINESS.....	3
FIGURA 2. PROCEDIMIENTO “GENERACIÓN REPORTES GERENCIALES”.....	3
FIGURA 3 ESQUEMA PROPUESTO.....	9
FIGURA 4. FASES DE BI (BERNABEU, PROCESO DE BI, 2009).	19
FIGURA 5. CARACTERÍSTICAS DEL DATA WAREHOUSE.....	21
FIGURA 6. DATA WAREHOUSE, VARIANTE EN EL TIEMPO.	23
FIGURA 7. DATA WAREHOUSE, NO VOLÁTIL.....	25
FIGURA 8. DATA WAREHOUSE, ESTRUCTURA NIVELES DE ESQUEMATIZACIÓN.....	26
FIGURA 9. DATA WAREHOUSING, ARQUITECTURA.....	30
FIGURA 10. SELECCIÓN OLTP.....	31
FIGURA 11. SELECCIÓN LOAD MANAGER.	32
FIGURA 12. TRANSFORMACIÓN: MEDIDA DE ATRIBUTOS.....	35
FIGURA 13. TRANSFORMACIÓN: CODIFICACIÓN	36
FIGURA 14. TRANSFORMACIÓN: FUENTES MÚLTIPLES.....	37
FIGURA 15. TRANSFORMACIÓN: CONVENCIONES DE NOMBRAMIENTO.....	37
FIGURA 16. CARACTERÍSTICAS DE LA CARGA, ANÁLISIS TIPOS DE CARGA.	39
FIGURA 17. SELECCIÓN DATA WAREHOUSE MANAGER.....	41
FIGURA 18. CUADRO COMPARATIVO ENTRE LA IMPLEMENTACIÓN ROLAP Y MOLAP.....	43
FIGURA 19. TABLAS DE DIMENSIÓN ZONA GEOGRÁFICA, CLIENTE Y TIEMPO.....	44
FIGURA 20. REPRESENTACIÓN MODELO FÍSICO DE TIPO ESTRELLA (HECHOS Y DIMENSIONES).....	46
FIGURA 21. ELEMENTO QUERY MANAGER EN LA ARQUITECTURA DATA WAREHOUSING.....	48
FIGURA 22. ELEMENTO HERRAMIENTAS DE CONSULTA Y ANÁLISIS, ARQUITECTURA DWH.....	49
FIGURA 23. ELEMENTO USUARIOS, EN LA ARQUITECTURA DATA WAREHOUSING.....	50
FIGURA 24. LISTADO DE VENTAJAS QUE OFRECE UN DATA WAREHOUSING.	51
FIGURA 25. LISTADO DE DESVENTAJAS DE UN DATA WAREHOUSING.	52
FIGURA 26. CLASIFICACIÓN DE LOS INDICADORES.....	56
FIGURA 27. CLASIFICACIÓN DE LOS INDICADORES POR EL ÁMBITO DE CONTROL.....	56
FIGURA 28. CLASIFICACIÓN DE LOS INDICADORES POR DIMENSIÓN.....	57
FIGURA 29. SEMÁFORO MEDICIÓN VISUAL DE DESEMPEÑO.....	59
FIGURA 30. METODOLOGÍA HEFESTO.....	64
FIGURA 31. ANÁLISIS DE REQUERIMIENTOS, OBTENCIÓN DE INDICADORES Y PERSPECTIVAS.....	66
FIGURA 32. MODELO CONCEPTUAL.....	67
FIGURA 33. EJEMPLO. MATRIZ DE MAPEO FUENTE OLTP / VARIABLES.....	69
FIGURA 34. EJEMPLO. MATRIZ DE MAPEO OLTP, DESCRIPCIÓN DE CAMPO.....	70
FIGURA 35. MODELO CONCEPTUAL AMPLIADO.	71

FIGURA 36. DISEÑO DIMENSIÓN CLIENTES.....	74
FIGURA 37. DISEÑO DIMENSIÓN PUBLICACIONES.....	74
FIGURA 38. DISEÑO DIMENSIÓN ZONAS.....	74
FIGURA 39. DISEÑO DIMENSIÓN TIEMPO.....	75
FIGURA 40. TABLA DE HECHOS CASO 1 (RICARDO, 2010).....	77
FIGURA 41. TABLA HECHOS CASO 2 (RICARDO, 2010).....	77
FIGURA 42. DISEÑO TABLA DE HECHOS CASO DE ESTUDIO.....	79
FIGURA 43: UNIONES TABLA DE HECHOS VENTAS.....	80
FIGURA 44. MODELO CONCEPTUAL ALTO NIVEL.....	91
FIGURA 45. DIAGRAMA ENTIDAD RELACIÓN DE LA BASE DE DATOS BD_COMUN_TOP.....	95
FIGURA 46. CORRESPONDENCIA ENTRE EL DIAGRAMA ENTIDAD RELACIÓN DE LA BASE DE DATOS BD_COMUN_TOP Y EL MODELO CONCEPTUAL.....	96
FIGURA 47. COLUMNAS DE LA FUENTE CARTERA TOP NOTCH.....	97
FIGURA 48. CORRESPONDENCIA ENTRE LAS COLUMNAS DE LA FUENTE CARTERA TOP NOTCH Y EL MODELO CONCEPTUAL.....	98
FIGURA 49. CORRESPONDENCIA ENTRE LAS COLUMNAS DE LA FUENTE REPORTE DE EJECUTIVO Y EL MODELO CONCEPTUAL.....	99
FIGURA 50. CORRESPONDENCIA ENTRE LAS COLUMNAS DE LA FUENTE “OTROS MONTOS RECUPERADOS” Y EL MODELO CONCEPTUAL.....	99
FIGURA 51. MODELO CONCEPTUAL AMPLIADO.....	111
FIGURA 52. TABLA DIMENSIÓN CLIENTES.....	112
FIGURA 53. TABLA DIMENSIÓN EJECUTIVOS.....	113
FIGURA 54. TABLA DIMENSIÓN FECHA.....	113
FIGURA 55. TABLA DIMENSIÓN CIUDADES.....	114
FIGURA 56. TABLA DE HECHOS RECUPERACIÓN CARTERA.....	115
FIGURA 57. UNIONES DE LA TABLA DE DIMENSIONES Y LA TABLA DE HECHOS.....	115
FIGURA 58. FLUJO DE LAS ÁREAS DE DESEMBARCO.....	118
FIGURA 59. TAREA PROCESO ETL ASIGNACIÓN PARÁMETROS INICIALES.....	120
FIGURA 60. TAREAS PROCESO ETL DESEMBARCO CARTERA.....	120
FIGURA 61. CONSULTA FECHA EN LA QUE SE ESTÁ PROCESANDO.....	121
FIGURA 62. TAREAS PROCESO ETL DESEMBARCO CLIENTES.....	123
FIGURA 63. TAREAS PROCESO ETL DESEMBARCO GESTIÓN.....	124
FIGURA 64. TAREAS PROCESO ETL DESEMBARCO CIUDADES.....	126
FIGURA 65. CONSULTA EXTRACCIÓN CIUDADES.....	127
FIGURA 66. TAREAS PROCESO ETL DESEMBARCO RECUPERACIÓN ADICIONAL.....	128
FIGURA 67. TAREAS PROCESO ETL DESEMBARCO EJECUTIVOS.....	130
FIGURA 68. CONSULTA EXTRACCIÓN EJECUTIVOS.....	130
FIGURA 69. FLUJO DE TAREAS CARGA TABLAS DE DIMENSIÓN AL DWH.....	131

FIGURA 70. FLUJO DE TAREAS CARGA TABLA DE HECHOS AL DWH.	133
FIGURA 71. DISEÑO DEL MODELO DEL CUBO MULTIDIMENSIONAL RECUPERACIÓN CARTERA. ...	135
FIGURA 72. SELECCIÓN DEL PROYECTO PARA CREACIÓN DEL CUBO MULTIDIMENSIONAL.....	136
FIGURA 73. CARPETAS DE TRABAJO EXPLORADOR DE SOLUCIONES.....	136
FIGURA 74. PARÁMETROS DE CONEXIÓN A LA BASE DEL DWH.	137
FIGURA 75. SELECCIÓN DE TABLAS DE DIMENSIÓN Y HECHOS.	138
FIGURA 76. SELECCIÓN MÉTODO GENERACIÓN DIMENSIONES.	139
FIGURA 77. SELECCIÓN FUENTE Y NOMBRE PARA LA DIMENSIÓN CIUDADES.	139
FIGURA 78. SELECCIÓN FUENTE Y NOMBRE PARA LA DIMENSIÓN CLIENTES.	140
FIGURA 79. SELECCIÓN FUENTE Y NOMBRE DIMENSIÓN EJECUTIVOS.....	140
FIGURA 80. SELECCIÓN FUENTE Y NOMBRE DIMENSIÓN FECHAS.....	140
FIGURA 81. SELECCIÓN MÉTODO DE CREACIÓN CUBO MULTIDIMENSIONAL.	141
FIGURA 82. SELECCIÓN DE LA TABLA FUENTE DEL GRUPO DE MÉTRICAS.	142
FIGURA 83. SELECCIÓN DIMENSIONES DEL CUBO DE RECUPERACIÓN DE CARTERA.	143
FIGURA 84. ASIGNACIÓN DEL NOMBRE DEL CUBO PARA LA RECUPERACIÓN DE CARTERA.	143
FIGURA 85. ASIGNACIÓN DEL NOMBRE DEL CUBO PARA LA RECUPERACIÓN DE CARTERA.	144
FIGURA 86. JERARQUIZACIÓN DIMENSIÓN FECHAS.	145
FIGURA 87. CÓDIGO MDX PARA CONSTRUIR EL INDICADOR “EFICIENCIA RECUPERACION CARTERA EJECUTIVO X FECHA”.	146
FIGURA 88. CÓDIGO MDX PARA CONSTRUIR EL INDICADOR “EFICIENCIA CONTACTAR CLIENTES”.....	146
FIGURA 89. CÓDIGO MDX PARA CONSTRUIR EL INDICADOR “EFICIENCIA RECUPERACION CARTERA EMPRESA X FECHA”.	147
FIGURA 90. CÓDIGO MDX PARA CONSTRUIR EL KPI “KPI INCREMENTO RECUPERACION CARTERA”.	148
FIGURA 91. CÓDIGO MDX PARA CONSTRUIR EL KPI “KPI INCREMENTO CONTACTAR A LOS CLIENTES”.....	149
FIGURA 92. GRUPO DE BOTONES PARA EL MANEJO DE FICHEROS.	150
FIGURA 93. GRUPO DE BOTONES PARA EL MANEJO DE LA FUENTE DE DATOS.....	150
FIGURA 94. GRUPO DE BOTONES PARA EL MANEJO DE LA FUENTE DE DATOS.....	150
FIGURA 95. GRUPO DE BOTONES PARA EL MANEJO DE LA FUENTE DE DATOS.....	151
FIGURA 96. GRUPO DE BOTONES DE LA APLICACIÓN WEB.....	151
FIGURA 97. PERFIL INICIO DE LA APLICACIÓN WEB.	152
FIGURA 98. PERFIL TABLA CRUZADA DE LA APLICACIÓN WEB.	153
FIGURA 99. PERFIL KPI DE LA APLICACIÓN WEB.	153
FIGURA 100. PERFIL REPORTES DE FUENTE Y BODEGA APLICACIÓN WEB.....	154

LISTADO DE FÓRMULAS

FÓRMULA 1. CÁLCULO DEL DESEMPEÑO POSITIVO DE UN EMPLEADO	58
FÓRMULA 2. CÁLCULO DEL DESEMPEÑO NEGATIVO DE UN EMPLEADO	58
FÓRMULA 3. CÁLCULO DEL DESEMPEÑO DE UN EMPLEADO, PARA LA EMPRESA TOP NOTCH BUSINESS.	93
FÓRMULA 4. CÁLCULO DEL INDICADOR EFICIENCIA DEL EJECUTIVO EN CONTACTAR A LOS CLIENTES, PARA LA EMPRESA TOP NOTCH BUSINESS.	93
FÓRMULA 5. CÁLCULO DEL INDICADOR EFICIENCIA EN RECUPERAR LA CARTERA A NIVEL EMPRESARIAL, PARA LA EMPRESA TOP NOTCH BUSINESS.	93
FÓRMULA 6. CÁLCULO DEL INDICADOR EFICIENCIA EN CONTACTAR A LOS CLIENTES, PARA LA EMPRESA TOP NOTCH BUSINESS.....	94

LISTADO DE ANEXOS

ANEXO 1

ENTREVISTA DE ANÁLISIS EMPRESA TOPNOTCH BUSINESS

ANEXO 2

MATRIZ DE MAPEO FUENTE A BODEGA

ANEXO 3

LISTADO DE ABREVIATURAS ESTANDARIZADAS PARA EL DESARROLLO

ANEXO 4

MANUAL DE USUARIO SISTEMA SATB - BI - DASHBOARD

ANEXO 5

MANUAL DE USUARIO SISTEMA SATB - BI - WEB

ANEXO 6

MANUAL TÉCNICO SISTEMA SATB

ANEXO 7

MANUAL ADMINISTRACIÓN SISTEMA SATB

GLOSARIO DE ABREVIATURAS

BI	Business Intelligence - Inteligencia de Negocios
DEVEXPRESS	Development Express to Visual Studio .NET
DWH	Data Warehouse.
ETL	Extract, transformation and Load.
HOLAP	Hybrid Online Analytical Process.
KPI	Key Performance Indicators.
MOLAP	Multidimensional Online Analytical Processing.
OLAP	Procesamiento de transacciones en línea (Online Transaction Processing) o sistema transaccional.
ROLAP	Processing Analytical On Line Relational.
SATB	Sistema de Análisis Topnotch Business.
SQL	Structured query language.
TBSEARCH	Sistema de búsqueda de clientes y ejecutivos de la empresa Topnotch Business
WEB	Vocablo inglés que significa “red”, “telaraña” o “malla”. El concepto se utiliza en el ámbito tecnológico para nombrar a una red informática

RESUMEN

El presente proyecto muestra el desarrollo de una aplicación de Inteligencia de Negocios, para la empresa KRONOSCONSULTING CIA LTDA., que incluye el análisis, diseño y construcción de un Data Warehouse y Cubo multidimensional que permite el análisis de indicadores y Key Performance Indicators (KPI), aplicados a la recuperación de cartera y la eficiencia en el contacto de clientes deudores. El objetivo es organizar y centralizar la información de la empresa en un solo repositorio, optimizando el tiempo de generación de indicadores situacionales y reportes de la empresa. La primera etapa enmarca los objetivos, la situación actual y el alcance que se pretenden al finalizar el presente trabajo. Por otra parte se explica el fundamento teórico y metodológico que guiará la construcción de la bodega de información y los indicadores. La metodología seleccionada es HEFESTO versión 2.0, en la cual detalla todas las etapas necesarias para la construcción de un Data Warehouse corporativo y los conceptos de Business Intelligence. En la etapa de construcción se trabajó con las herramientas de la suite de Microsoft SQL Management SQL 2012, Visual Studio 2010 y el componente DevExpress en la versión 13.1.

Palabras Clave:

- Inteligencia de Negocio
- Indicadores,
- Bodega de Datos,
- Cuadro de Mando
- HEFESTO

ABSTRACT

This white paper shows a Business Intelligence (BI) application development for the KRONOSCONSULTING CIA LTDA. Company. This includes analysis, design and construction of a data warehouse and multidimensional cube, that allows the indicators and Key Performance Indicators (KPI) analysis of the company applied to loan recovery and efficiency in customer contact. The main objective of the project is to organize and centralize business information into a single repository, optimizing the time in the generation of situational indicators and company reports. The first part let know us the objectives, today situation and the scope of the Project. By other hand it explains the methodology and theoretical fundamentals for the construction of the DWH and the indicators. The methodology selected was HEFESTO 2.0, it describe all phases to construct DWH and Business Intelligence solution. The tools used were suite de Microsoft SQL Management SQL 2012, Visual Studio 2010 y DevExpress 13.1 component.

Keywords:

- Business Intelligence
- KPI
- Data Warehouse
- Dashboard
- HEFESTO

CAPÍTULO 1

Análisis, diseño, construcción e implementación de un Data Warehouse para toma de decisiones y construcción de los KPI, para la empresa KRONOSCONSULTING CIA LTDA.

1.1. Antecedentes

La consultora KRONOSCONSULTING CIA LTDA.¹, dedicada a realizar consultorías en varios ámbitos de los negocios, pretende como parte de un nuevo proyecto de consultoría, dar soporte tecnológico a la empresa Topnotch Business para resolver su necesidad de evaluación del desempeño de la organización y sus empleados con una gestión eficiente de la información actual. La consultora gestionará, administrará y controlará cada fase del proyecto, interviniendo y extendiendo el soporte en todos los ámbitos de acción y a todos los participantes.

La empresa Topnotch Business², se creó en Quito, Ecuador en el año 2005, dedicada a la cobranza externa legal de Movistar y otros clientes con menor participación. Esta necesidad de negocio surgió debido a que la empresa Movistar no tenía la capacidad para cobrar las cuentas vencidas de sus clientes.

¹ Nombre propio de la Empresa Consultora dueña del Proyecto.

² Nombre propio de la Empresa Cliente a la cual va dirigido el Proyecto.

La empresa se inició y se mantiene con el manejo de varios tipos de cartera vencida y cada una de estas se gestionan por uno o varios ejecutivos de cobranza. Estos contratos se reparten de forma empírica, a cada ejecutivo según el administrador va adquiriendo experiencia de sus colaboradores con respecto a la ciudad en la que son más aptos, careciendo de una estrategia de seguimiento y evaluación de desempeño al personal para su asignación posterior de los contratos; por lo cual, hace algún tiempo, ha impactado en el alcance de los objetivos del personal y empresariales fijados al comienzo de cada año, debido a lo complicado que se hace la tarea de evaluación del personal en forma manual.

Actualmente toda la información que maneja la empresa Topnotch Business se realiza en forma manual mediante hojas electrónicas en Excel; las mismas que representan sus contratos, carteras³ cobradas, carteras pendientes y nómina de ejecutivos; y, mediante el sistema "TBSEARCH"⁴ la base de sus clientes. El volumen de datos que maneja la empresa es muy alto al momento de recolectar la información para la construcción de los reportes; generando así un alto costo en tiempo de procesamiento, como en recursos.

Los reportes que genera la empresa con periodicidad quincenal y mensual para el comité gerencial, se generan de forma tardía; adicionalmente reportes semanales adicionados por cada ejecutivo no se logran consolidar; lo cual hace imposible un manejo manual adecuado para

³ Listado de operaciones de débito o crédito realizadas por una persona en una empresa.

⁴ Sistema informático que realiza búsquedas de clientes y ejecutivos de la empresa Topnotch Business

presentar la información y generar los “Indicadores de Gestión”⁵ claves de éxito de la empresa.

A continuación se representa gráficamente la situación actual de la empresa, su organigrama y el esquema del proceso de generación de reportes gerenciales.

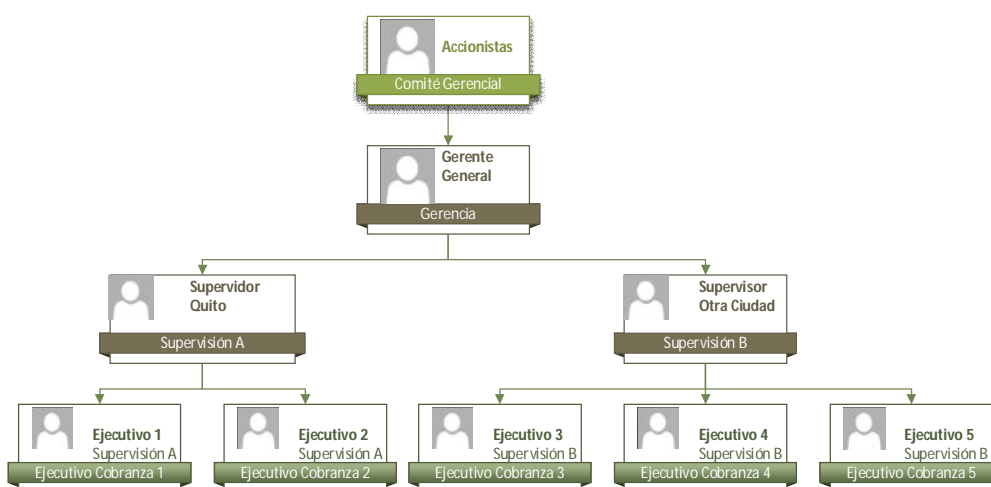


Figura 1. Organigrama de la Empresa Topnotch Business

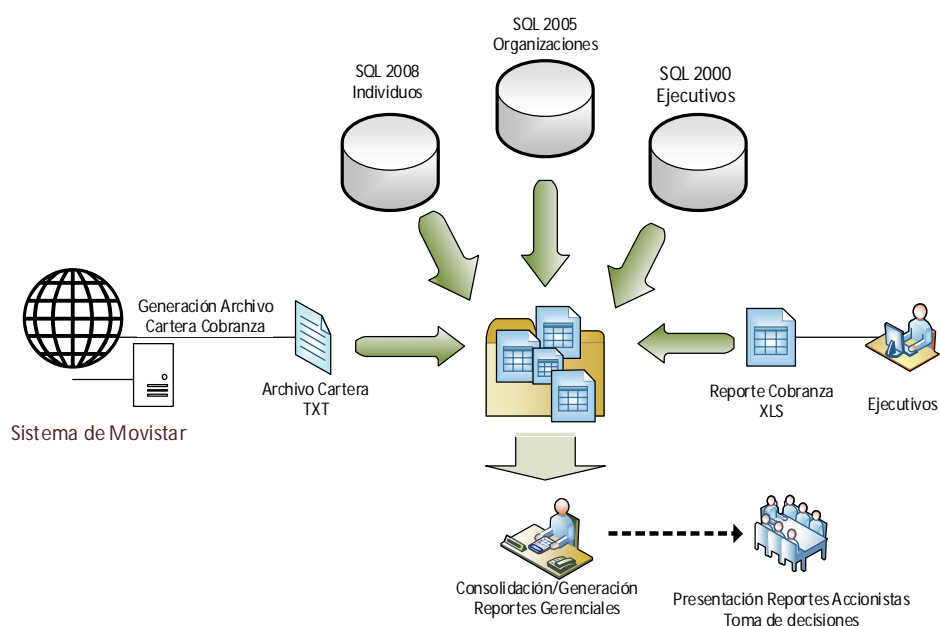


Figura 2. Procedimiento “Generación reportes Gerenciales”.

⁵ Parámetro de análisis que mide el estado actual de un proceso, actividad o escenario para poder tomar decisiones.

Se propone, por las gerencias y el grupo de proyecto de tesis que lo abalan, realizar una propuesta informática que ayude al adecuado almacenamiento, procesamiento y presentación de la información y la generación de los KPI⁶, para la empresa Topnotch Business colaborando de esta manera con la toma de decisiones pertinentes dentro de la organización.

1.2. Planteamiento del problema

La empresa Topnotch Business carece de una herramienta que permita evaluar el desempeño obtenido por cada uno de los empleados y de la empresa a nivel general, de una forma clara y oportuna. Por tal motivo, existe un alto riesgo de plantear estrategias empresariales erróneas, que no permitan alcanzar los objetivos empresariales planteados, y sin poder tener una evaluación continua sobre el desempeño de sus empleados.

Topnotch Business necesita conocer cuál es el porcentaje de recuperación de cartera vencida en un determinado tiempo y por cada ejecutivo de cartera, adicionalmente la construcción de los reportes gerenciales en los tiempos establecidos y presentación de los Indicadores de Gestión para medir el desempeño.

⁶ *Key Performance Indicators*, los indicadores clave de desempeño son métricas financieras o no financieras, utilizadas para cuantificar objetivos que reflejan el rendimiento de una organización, y que generalmente se recogen en su plan estratégico

Acorde a esta necesidad la empresa requiere la construcción de un almacén de datos (Data Warehouse) que permita generar tanto los reportes gerenciales como la construcción de KPI de los ejecutivos de una forma confiable, rápida y precisa.

1.3. Justificación

Desde que se inició la era de la computadora, las organizaciones han usado los datos desde sus sistemas operacionales para atender sus necesidades de análisis de información. Algunas proporcionan acceso directo a la información contenida dentro de las aplicaciones; otras, han extraído los datos desde sus bases de datos para combinarlos de varias formas no estructuradas y analizar la información, en su intento por tomar decisiones eficaces y oportunas.

El Data Warehouse, es actualmente, el centro de atención de las grandes instituciones, porque provee un ambiente base para el análisis predictivo de la información que está siendo administrada por diversas aplicaciones/herramientas operacionales.

Subsiguiente a implementar una bodega de datos, es necesario calcular y presentar el resultado de este análisis de una forma sencilla y a nivel general, en este caso al Comité Gerencial de la empresa.

Relevancia Social. Ayudar a la empresa, empleados y personas relacionadas, a mejorar su desempeño con un estímulo de eficacia y mejor gestión de la empresa.

Importancia Tecnológica. El proyecto SATB⁷, se presenta como solución tecnológica que permita almacenar, procesar y presentar toda la información necesaria para atender los requerimientos más importantes de análisis de la empresa. De esta manera se obtiene una ayuda tecnológica, que hoy carecen, como instrumento organizado, de acceso inmediato e integrador del proceso de generación de indicadores, KPI, necesarios para el análisis situacional y toma de decisiones, en la Organización. Además se complementará el estudio de cómo trabajar en la nueva versión de SQL Server 2012 con herramientas de Análisis de Datos.

Importancia Económica. Al analizar la información que proveerá el Data Warehouse por medio de los reportes y de los KPI, se mejora considerablemente la toma de decisiones y con ello se incurrirá en menores gastos, mejor administración de los recursos y generación de un mayor porcentaje de recuperación de la cartera vencida.

Beneficiarios:

- Directos
 - Topnotch Business.

- Indirectos
 - Consultora “KRONOSCONSULTING CIA LTDA.”

 - Sociedad en General.

⁷ Denominación que se le ha dado al proyecto actual de construcción del Data Warehouse e Indicadores de desempeño de los empleados.

1.4. Objetivos

1.4.1. Objetivo general

Desarrollar una aplicación de *Business Intelligence* utilizando la metodología Hefesto con el uso de la herramienta orientada a BI SQL 2012 para la empresa KRONOSCONSULTING CIA LTDA.

1.4.2. Objetivos específicos

- Revisar los conceptos de Data Warehousing, Business Intelligence y los fundamentos de la Metodología Hefesto.
- Análisis y diseño del proyecto de Business Intelligence SATB, aplicando la Metodología Hefesto.
- Desarrollo y construcción de los indicadores KPI del proyecto de Business Intelligence SATB, aplicando la Metodología Hefesto.

1.5. ALCANCE

El presente proyecto de tesis concentrará el estudio de la información del cliente más importante como es Movistar, sin alterar la fuente de información con la que opera la empresa Topnotch Business. Por lo cual se realizará:

La identificación de cada fuente de datos que se va a utilizar para la integración de la información en el Data Warehouse, la información se encuentra en Hojas de Excel y serán copiadas con periodicidad que se definirá en el desarrollo del proyecto en una ruta FTP compartida.

Levantamiento de los Requerimientos y creación de las matrices de Mapeo de las fuentes.

- Construcción del Data Warehouse acoplado a los requerimientos funcionales.
- Construcción de los reportes de la información almacenada en el Data Warehouse, estos en primera instancia se detallan a continuación:
 - Monto de cartera vencida recuperada por la empresa en un periodo de tiempo.
 - Monto de cartera vencida recuperada por el ejecutivo en un periodo de tiempo.
 - Pagos realizados por los clientes en periodo de tiempo.
 - Porcentaje de Clientes contactados en un periodo de tiempo.
 - Porcentaje de Clientes contactados por ejecutivo en un periodo de tiempo.
 - Listado de Clientes pendientes de Notificación.
 - Listado de Direcciones y números de contacto del Cliente
 - Construcción de los KPI,
 - Eficiencia del ejecutivo en recuperar la cartera.

- Eficiencia del ejecutivo en contactar los clientes.
- Eficiencia de la empresa en recuperar la cartera.
- Eficiencia de la empresa en contactar los clientes.

Generación de los respaldos periódicos del Data Warehouse para asegurar la disponibilidad de la información en casos de criticidad o desastre ya sea externa o interna.

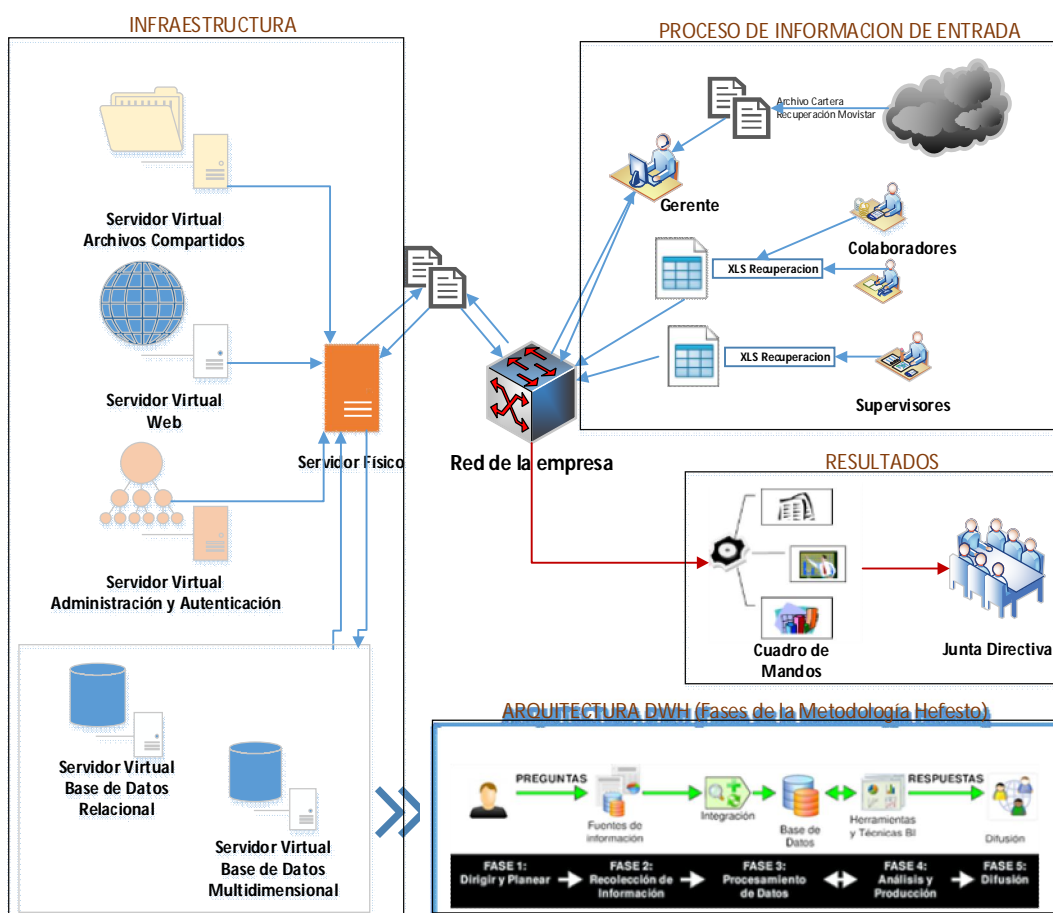


Figura 3 Esquema Propuesto.

CAPÍTULO 2

FUNDAMENTACIÓN TEÓRICA

2.1. Bases teóricas

2.1.1. Definiciones Importantes

Al comenzar se necesita tener cada uno de los términos y su significado, que se incluyen en las fases de implantación de una herramienta Business Intelligence, para ello se describen a continuación.

2.1.1.1. OLTP

Es la sigla en inglés de Procesamiento de Transacciones En Línea (Online Transaction Processing) es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional). Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor ya que suelen ser utilizados por empresas con una red informática distribuida. (Wikipedia, OLTP, 2013).

2.1.1.2. OLAP

Es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de la llamada Inteligencia empresarial (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para cumplir este objetivo se debe usar estructuras multidimensionales (o Cubos OLAP) que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales (OLTP); los mismos sirvan para crear informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares.

La razón de usar OLAP para las consultas es la rapidez de respuesta. Una base de datos relacional almacena datos en tablas discretas normalizadas. Esta estructura en sistemas OLTP es buena, pero para las complejas consultas multitabla es relativamente lenta; por lo tanto, las bases de datos multidimensional es el mejor modelo para búsquedas (aunque peor desde el punto de vista operativo).

La principal característica que potencia a OLAP, es que es lo más rápido a la hora de ejecutar sentencias SQL de tipo SELECT, en contraposición con OLTP que es la mejor opción para operaciones de tipo INSERT, UPDATE Y DELETE (Wikipedia, OLTP, 2013).

2.1.1.3. Data Warehousing

Se entiende por Data Warehousing como el proceso de extraer y filtrar datos de las operaciones comunes de la organización, procedentes de los

distintos sistemas de información operacionales y/o sistemas externos, para transformarlos, integrarlos y almacenarlos en un depósito o almacén de datos (Data Warehouse, en inglés) con el fin de acceder a ellos para dar soporte en el proceso de toma de decisiones de una organización. Es decir, la finalidad es convertir los datos operacionales en información relacionada y estructurada, homogénea y de mayor calidad, identificada convenientemente y que se mantenga en el tiempo, es decir, los datos más recientes no sustituyen a los precedentes, pero tampoco se acumulan de cualquier manera, sino que se suelen mantener con un mayor nivel de detalle los datos actuales, y de manera más agregada los datos anteriores (Curto, 2007).

2.1.1.4. Data Warehouse

Según W. H. Inmon (considerado por muchos el padre del Data Warehouse), un Data Warehouse es un conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, que tienen por objetivo dar soporte a la toma de decisiones.

Según Ralph Kimball (considerado el principal promotor del enfoque dimensional para el diseño de almacenes de datos), un Data Warehouse es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis.

Un Data Warehouse proporciona una visión global, común e integrada de los datos de la organización, independiente de cómo se vayan a utilizar

posteriormente por los consumidores o usuarios, con las propiedades siguientes: estable, coherente, fiable y con información histórica. Al abarcar un ámbito global de la organización y con un amplio alcance histórico, el volumen de datos puede ser muy grande (centenas de terabytes). Las bases de datos relacionales son el soporte técnico más comúnmente usado para almacenar las estructuras de estos datos y sus grandes volúmenes. Normalmente en el almacén de datos habrá que guardar información histórica que cubra un amplio período de tiempo. Pero hay ocasiones en las que no se necesita la historia de los datos, sino sólo sus últimos valores, siendo además admisible generalmente un pequeño desfase o retraso sobre los datos operacionales. En estos casos el almacén se llama almacén operacional (ODS, Operational Data Store) (Curto, 2007).

2.1.1.5. Operational Data Store (ODS)

ODS es un contenedor de datos activos, es decir datos operacionales que ayudan al soporte de decisiones y a la operación. Está entre un OLAP y un OLTP. Su función es integrar los datos al igual que en el Data Warehouse pero con una ventana de actualización muy pequeña (del orden de minutos) y con mucho menos detalle. (Wikipedia, Almacén operacional de los datos, 2013).

2.1.1.6. Data Mart.

Un Data Mart es un subconjunto de datos del Data Warehouse con el objetivo de responder a un determinado análisis, función o necesidad y con

una población de usuarios específica. Al igual que en un Data Warehouse, los datos están estructurados en modelos de estrella o copo de nieve y un Data Mart puede ser dependiente o independiente de un Data Warehouse. Por ejemplo, un posible uso sería para el Data Mining.

¿Qué diferencia existe entonces entre un Data Mart y un Data Warehouse? Su alcance. El Data Mart está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización. Es el almacén natural para los datos departamentales. En cambio, el ámbito del Data Warehouse es la organización en su conjunto. Es el almacén natural para los datos corporativos comunes. (Curto, 2007)

2.2. Inteligencia de Negocios (Business Intelligence BI)

2.2.1. Orígenes.

Howard Dresner, es a quien se le atribuye la creación del término Business Intelligent, en el año de 1989, cuando era analista del Instituto Gartner. Así como él, los norteamericanos adquirieron fama por el desarrollo de las modernas herramientas que se utilizan para el Business Intelligent.

Yves-Michel Marti, científico, profesor y fundador de Egideria, una de las más grandes empresas europeas de consultoría en Business Intelligence, reivindica, que el Viejo Continente, es la cuna y la pionera en la aplicación del concepto de BI. Según Marti, en sus estudios sobre economía inteligente, uno de los ejemplos señala que a fines del siglo XVI, la reina

Isabel I, con el fin de ocupar territorios conquistados, determinó que la base de la fuerza inglesa fuera “la información y el comercio”, y le solicitó al filósofo Francis Bacon que inventara un sistema dinámico de información, que fue ampliamente aplicado por los ingleses

La era antes del BI se sitúa en un pasado cercano, aproximadamente cuarenta o cincuenta años atrás, durante los años 60 y 70 del siglo XX. En esa época, las computadoras dejaron de ocupar salas gigantescas, mientras las empresas comenzaban a considerar a los datos como una posible e importante fuente de información para la toma de decisiones, aún sin existir los recursos para el análisis consistente de los datos.

El panorama comenzó a cambiar en la década del 70, con el surgimiento de las tecnologías de almacenamiento y de acceso a datos – DASD (Direct Access Storage Device – Dispositivo de Almacenamiento de Acceso Directo), y SGBD (Sistema de Gestión de Bases de Datos), dos siglas cuyo principal objetivo era establecer una única fuente de datos para todo el procesamiento. A partir de ese momento, la computadora comenzó a ser vista como un coordinador central de las actividades corporativas y la base de datos fue considerada un recurso básico para la ventaja competitiva.

A comienzos de los 90, la mayoría de las grandes empresas contaba solamente con Centros de Información (CI), que aun manteniendo un *stock* de datos, tenían disponibles en pequeña medida la información. Aun así, los CI suplían, de cierta manera, las necesidades de los gerentes, al proporcionar informes y análisis gerenciales. El mercado se volvía más

complejo y la Tecnología de la Información mejoraba los programas informáticos, que generaban información no sólo más precisa sino también en el momento oportuno para definir acciones que mejoraran el desempeño de las empresas.

Entre 1992 y 1993, surgió Data Warehouse, una gran base de datos informativa, es decir, un repositorio de datos consolidados, limpios y estandarizados. Los especialistas lo consideran la pieza esencial para un proyecto de Business Intelligence. Ese repositorio no tiene que ser, necesariamente, un Data Warehouse, sino algo menos complejo, como por ejemplo un Data Mart (base de datos diseñada especialmente para temas o áreas específicas), o inclusive una base de datos relacional común, separada del entorno transaccional (operativo) y dedicada a almacenar la información utilizada como base para elaborar diferentes análisis y proyecciones.

El desarrollo tecnológico que permitió crear herramientas para facilitar la captación, extracción, almacenamiento, filtrado, disponibilidad y personalización de los datos, llevó a que las corporaciones se interesaran en soluciones de BI en forma más contundente, principalmente alrededor de 1990, cuando el concepto de Business Intelligence comenzó a difundirse como una evolución del EIS, Executive Information System, creado a fines de la década del 70 por investigadores del MIT (Massachusetts Institute of Technology-EE.UU.).

Con el correr de los años, el término Business Intelligence adquirió mayor alcance e incluyó una serie de herramientas, como el mismo EIS y las soluciones DSS (Decision Support System – Sistema de Soporte de Decisiones), Planillas Electrónicas, Generadores de Consultas y de Informes, Data Marts, Data Mining y Herramientas OLAP (Online Analytical Process). Todas buscan impulsar la agilidad comercial, dinamizar la toma de decisiones y perfeccionar las estrategias de relación con los clientes.

Actualmente, las pequeñas, medianas y grandes empresas necesitan del BI para las más diversas situaciones, desde la toma de decisiones hasta la optimización del trabajo, reducción de costos, pronóstico de crecimiento, elaboración de estrategias, rápida detección de desviaciones de presupuesto, identificación de tendencias de ventas, seguimiento eficiente de los objetivos planificados, disponibilidad de cuadros comparativos de rendimiento de empleados, asociados y colaboradores, alineamiento del día a día con estrategias planteadas a futuro.

Para que un proyecto de BI contribuya al mejor desempeño de una empresa, deben analizarse dos factores: el costo y los objetivos que se desea alcanzar - es decir, se debe alinear el proyecto con los intereses y las estrategias de la corporación. Tomado de (NextGenerationCenter, s.f.)

2.2.2. Definición.

Business Intelligence (BI) o inteligencia de negocios, se lo define como un concepto que integra el procesamiento y el almacenamiento de grandes

cantidades de datos, con el objetivo de poder transformar toda esta información en conocimiento de alto valor que permita tomar eficientemente decisiones estratégicas y en tiempo real, a través de un sencillo análisis y exploración.

2.2.3. Proceso para realizar Business Intelligent.

El proceso se divide en cinco fases, las cuales son:

FASE 1: Dirigir y Planear. En esta fase inicial es donde se deberán recolectar los requerimientos de información específicos de los diferentes usuarios, así como entender sus diversas necesidades, para que luego en conjunto con ellos se generen las preguntas que les ayudarán a alcanzar sus objetivos. (Ricardo, 2010)

FASE 2: Recolección de Información. Es aquí en donde se realiza el proceso de extracción de la información tanto interna como externa, los datos que serán necesarios para encontrar las respuestas a las preguntas planteadas en la Fase 1. (Ricardo, 2010)

FASE 3: Procesamiento de Datos. En esta fase se integran y cargan los datos en crudo en un formato utilizable para el análisis. Esta actividad puede realizarse mediante la creación de una nueva base de datos, agregando datos a una base de datos ya existente o bien consolidando la información expuesta en la Fase 2. (Ricardo, 2010)

FASE 4: Análisis y Producción. Ahora, se procederá a trabajar sobre los datos extraídos e integrados, utilizando herramientas y técnicas propias de la tecnología BI, para crear inteligencia de negocio. Como resultado final de esta fase se obtendrán las respuestas a las preguntas, mediante la creación de reportes, indicadores de rendimiento, cuadros de mando, gráficos estadísticos, etc. (Ricardo, 2010)

FASE 5: Difusión. Finalmente, se les entregará a los usuarios que lo requieran las herramientas necesarias, que les permitirán explorar los datos de manera sencilla e intuitiva. (Ricardo, 2010)

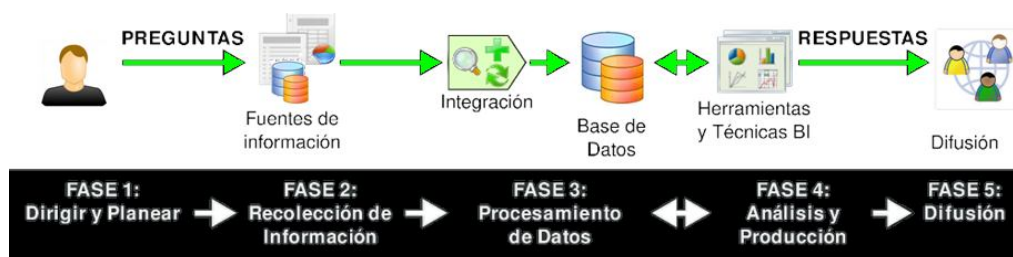


Figura 4. Fases de BI

Fuente: (Bernabeu, Proceso de BI, 2009).

2.3. Bodega de datos (Data Warehouse)

2.3.1. Definición

Según la definición realizada por William Harvey Inmon, también conocido como el padre del Data Warehousing, un Data Warehouse es una colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia.

En otras palabras es una bodega o almacén único de datos multidimensional o relacional, que almacena datos extraídos y transformados, para su consolidación, integración y centralización; estos pueden provenir de distintas fuentes y tipos, haciendo así posible el posterior análisis y exploración.

2.3.2. Funcionalidad

El Data Warehousing posibilita la extracción de datos de sistemas operacionales y fuentes externas, permite la integración y homogenización de los datos de toda la empresa, provee información que ha sido transformada y resumizada, para que ayude en el proceso de toma de decisiones estratégicas. (Bernabeu, HEFESTO: Metodología para la Construcción de un Data Warehouse, 2010)

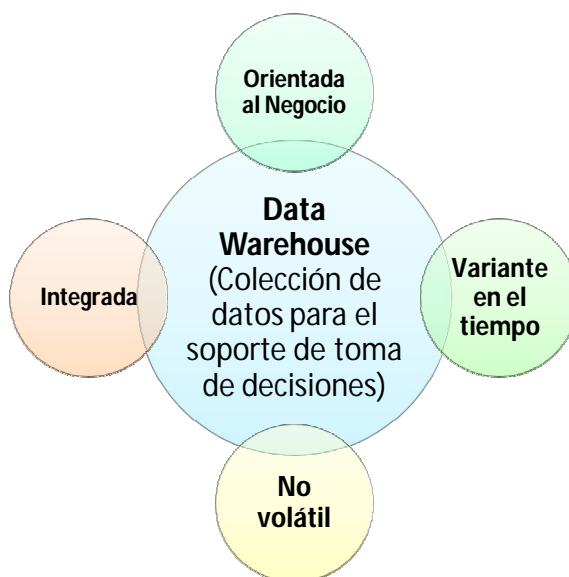


Figura 5. Características del Data Warehouse.

2.3.3. Características

Un Data Warehouse es:

- Orientado al negocio.
- Integrado.
- Variante en el tiempo.
- No volátil.
- Redundante.

2.3.3.1. Orientada al negocio

La primera característica del Data Warehouse, es que la información se clasifica en base a los aspectos que son de interés para la organización. Esta clasificación afecta directamente al diseño y la implementación del almacén de datos, debido a que la estructura del mismo difiere

considerablemente a la de los clásicos procesos operacionales orientados a las aplicaciones. (Bernabeu, HEFESTO: Metodología para la Construcción de un Data Warehouse, 2010)

Los procesos orientados a la aplicación están relacionados entre los sistemas transaccionales y el Data Warehouse, ya que en los dos se necesita una alta accesibilidad a los datos debido a un mayor desempeño y velocidad en la ejecución de consultas, la premisa de esta relación se debe a que en un Data Warehouse la información está desnormalizada⁸, es decir, con redundancia y que la misma este dimensionada bajo perspectivas; esto para evitar recorrer por toda la base de datos cuando se realiza las consultas generales o específicas. Un Data Warehouse permite realizar estas tareas de consulta de una forma más rápida y eficaz, con ello poder satisfacer una alta demanda y muy recurrente, en complejos análisis mensuales; optimizando el tiempo de respuesta e igualándolo al de un sistema transaccional.

2.3.3.2. Integrada

Esta característica implica que todos los datos fuentes producidos por distintos departamentos, secciones y aplicaciones (internos o externos), deben ser consolidados en una instancia antes de ser agregados al Data Warehouse, y deben por lo tanto ser analizados para asegurar su consistencia, calidad y limpieza. Cuenta con diversas técnicas y

⁸ Desnormalización, es el proceso de invertir las transformaciones realizadas durante la normalización; o sea se debe eliminar las relaciones y redundar la información dentro de una misma tabla.

subprocesos para llevar a cabo sus tareas. Una de estas técnicas son los procesos ETL: Extracción, Transformación y Carga de Datos. (Bernabeu, HEFESTO: Metodología para la Construcción de un Data Warehouse, 2010)

La integración de datos, ayuda a estandarizar las convenciones de nombres, unidades de medidas y codificaciones, de los múltiples datos disímiles de las fuentes.

2.3.3.3. Variante en el tiempo



Figura 6. Data Warehouse, variante en el tiempo.

Esta cualidad que no se encuentra en fuentes de datos operacionales, garantiza poder desarrollar análisis de la dinámica de la información, como pronósticos y análisis de tendencias y patrones, todo esto a partir de una base estadística de información, pues ella es procesada como una serie de instantáneas, cada una representando un periodo de tiempo. Gracias al sello de tiempo se podrá tener acceso a diferentes versiones de la misma información.

Es importante tener en cuenta la granularidad de los datos, así como también la intensidad de cambio natural del comportamiento de los fenómenos de la actividad que se desarrolle, para evitar crecimientos incontrolables y desbordamientos de la base de datos.

El intervalo de tiempo y periodicidad de los datos debe definirse de acuerdo a la necesidad y requisitos de los usuarios. (Bernabeu, HEFESTO: Metodología para la Construcción de un Data Warehouse, 2010)

2.3.3.4. No volátil

La información es útil para el análisis y la toma de decisiones solo cuando es estable. Los datos operacionales varían momento a momento, en cambio, los datos una vez que entran en el Data Warehouse no cambian.

La actualización, o sea, insertar, eliminar y modificar, se hace de forma muy habitual en el ambiente operacional sobre una base, registro por registro, en cambio en el depósito de datos, la manipulación básica de estos es mucho más simple, debido a que solo existen dos tipos de operaciones: la carga de datos y el acceso a los mismos.

Por esta razón es que en el Data Warehouse no se requieren mecanismos de control de concurrencia y recuperación. (Bernabeu, HEFESTO: Metodología para la Construcción de un Data Warehouse, 2010)

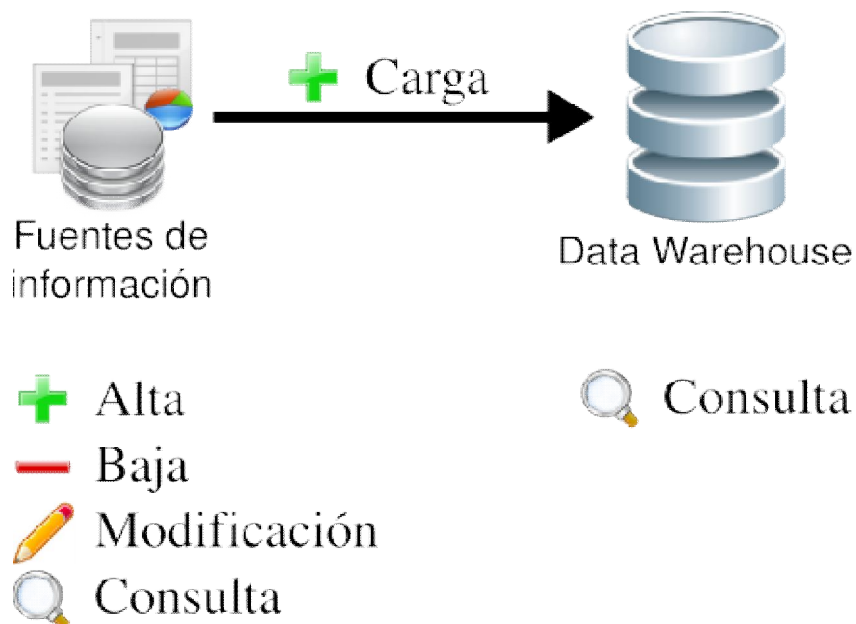


Figura 7. Data Warehouse, no volátil.

Fuente: (Bernabeu, HEFESTO: Metodología para la Construcción de un Data Warehouse, 2010).

2.3.3.5. Redundancia

Un Data Warehouse es redundante en un nivel muy bajo, aunque este recibe la información histórica de diferentes fuentes de sistemas operacionales, la redundancia de datos entre ambos ambientes es menor que 1%. Para comprender mejor este análisis, se debe considerar lo siguiente:

Tabla 1. Análisis Data Warehouse no Redundante

No.	Consideraciones
1	El periodo de tiempo es muy diferente entre los dos ambientes.
2	Los datos son procesados y experimentan una transformación considerable, antes de ser cargados al DWH. La mayor parte de los datos se alteran significativamente al ser seleccionados, consolidados y movidos al depósito.
3	El Data Warehouse contiene la información de forma resumida y sumariada que no se encuentra en el ambiente operacional.
4	En el ambiente operacional existe información que se filtra antes de pertenecer al Data Warehouse. Estos datos nunca ingresarán, debido a que no es una información relevante para el proceso de toma de decisiones.

2.3.4. Estructura

En el Data Warehouse los datos se estructuran de manera muy particular y existen diferentes niveles de esquematización y detalle que los delimitan. El nivel de granularidad se obtiene a través de “tablas de hechos” agregadas y/o pre agregadas.

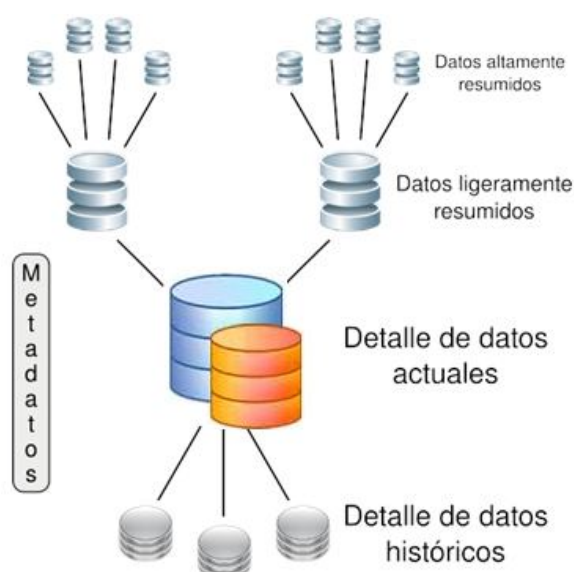


Figura 8. Data Warehouse, estructura niveles de esquematización

Fuente: (Bernabeu, Estructura Data Warehouse, 2009).

A continuación, en la Tabla 2., se describen cada uno de los elementos que conforman la estructura del Data Warehouse, detallados en la Figura 8.

Tabla 2. Análisis Estructura DWH. (Dario, 2009)

No.	Elemento	Consideraciones
1	Datos actuales	Los datos residentes poseen el más bajo nivel de granularidad, o sea, se almacenan a nivel de detalle. Su administración sea costosa y compleja, con el fin de conseguir que el acceso a la información sea sencillo y rápido, ya que son bastante voluminosos
2	Datos históricos	Representan aquellos datos antiguos a nivel de detalle, que no son frecuentemente consultados y se almacenan de forma externa, ya que son pesados y poco requeridos.
3	Datos ligeramente resumidos	Son los que provienen desde un bajo nivel de detalle y sumarizan o agrupan los datos bajo algún criterio o condición de análisis. Ventas en un determinado mes.
4	Datos altamente resumidos	Son aquellos datos que se compactan aún más a los datos ligeramente resumidos. Se guardan en disco y son muy fáciles de acceder.
5	Metadatos	Representan la información acerca de los datos. De muchas maneras se sitúa en una dimensión diferente al de otros datos del DWH, ya que su contenido no es tomado directamente desde el ambiente operacional.

2.3.5. OLTP vs DWH

Para la comparación entre OLTP y Data Warehouse se presenta una tabla comparativa entre los dos ambientes, así resumiendo sus diferencias.

Tabla 3. OLTP Vs. DWH.

No		OLTP	Data Warehouse
1	Objetivo	Soportar actividades transaccionales diarias.	Consultar y analizar la información estratégica y táctica.
2	Tipo de datos	Operacionales.	Para la toma de decisiones.
3	Modelo de datos	Normalizado.	Desnormalizada.
4	Consulta	SQL.	SQL más extensiones.
5	Datos Consultados	Actuales.	Actuales e históricos.
6	Horizonte tiempo	60 – 90 días.	5 – 10 años.
7	Tipos de consultas	Repetitivas, predefinidas	No previsibles, dinámicas.
8	Nivel Almacenamiento	Nivel de detalle.	Nivel de detalle y diferentes niveles de sumarización.
9	Acciones disponibles	Alta, baja, modificación y consulta.	Carga y consulta.
10	Número de Transacciones	Elevado.	Medio o bajo.
11	Tamaño	Pequeño – Mediano.	Grande.
12	Tiempo de Respuesta	Pequeño (segundos - minutos)	Variable (minutos - horas).
13	Orientación	Orientado a las aplicaciones.	Orientado al negocio.
14	Sello de Tiempo	La clave puede tener o no un elemento de tiempo.	La clave tiene un elemento de tiempo.
15	Estructura	Generalmente estable.	Generalmente varía de acuerdo a su propia evolución y actualización.

El cuadro analizado denota la gran diferencia y el por qué se debe cambiar nuestro lineamiento tecnológico al de un Data Warehouse si se requiere analizar los datos.

2.4. Arquitectura Data Warehousing

La arquitectura comprende todos los elementos que interactúan entre sí, y su función específica en un proceso de Data Warehousing. El proceso de general se detalla a continuación:

- La fuente para la obtención de los datos son extraídos de diferentes fuentes; tales como: aplicaciones, bases de datos, archivos, etc. La información obtenida generalmente reside en diferentes tipos de sistemas, orígenes y arquitecturas y tienen formatos muy variados.
- Los datos son integrados, transformados y limpiados, para luego ser cargados en el Data Warehouse.
- La información del DWH se estructura en cubos multidimensionales, mediante los cuales se prepara la información para responder a consultas dinámicas con un buen comportamiento. Pero también pueden utilizarse otros tipos de estructuras de datos para representar la información del DWH, como por ejemplo Business Models.
- Los usuarios acceden a los cubos multidimensionales, Business Models (u otro tipo de estructura de datos) del DWH utilizando diversas herramientas de consulta, exploración, análisis y reportes.

La figura 9., representa los elementos que intervienen en el proceso de Data Warehousing, que en los literales puntos posteriores serán detallados.

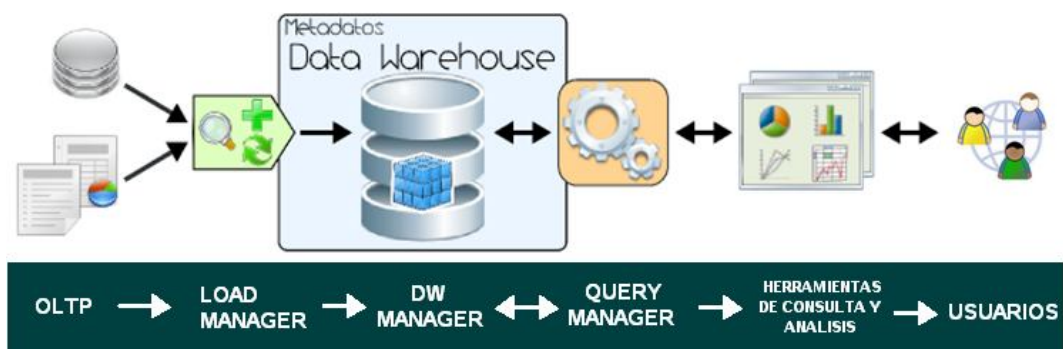


Figura 9. Data Warehousing, arquitectura

Fuente: (Bernabeu, Arquitectura Data Warehouse, 2009).

2.4.1. OLTP

OLTP (Online Transaction Processing), esta fase analiza toda aquella información transaccional que genera la empresa diariamente, además, de las fuentes externas que se extrae información. Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran:

- Archivos de textos.
- Hojas de cálculos.
- Hipertextos.
- Bases de datos transaccionales.

Estos objetos contienen informes semanales, mensuales, anuales, etc., que sirven para el análisis de las variable del negocio.

En la Figura 10 se enmarca donde se encuentra el primer elemento de la arquitectura

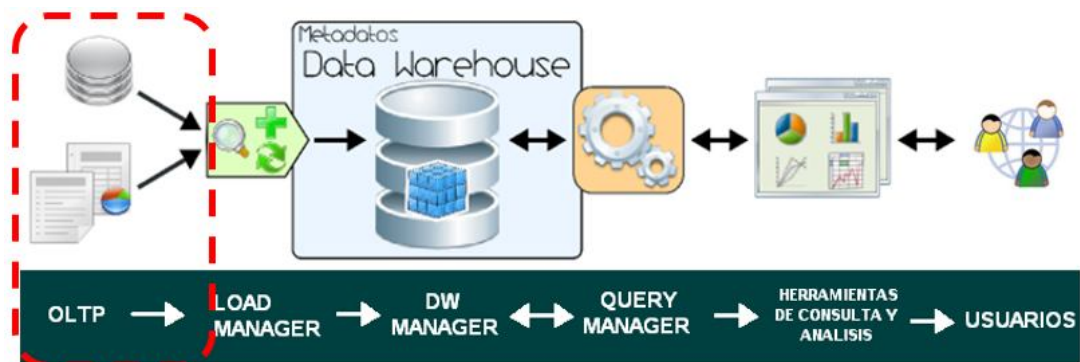


Figura 10. Selección OLTP

Fuente: (Bernabeu, OLTP, 2009).

2.4.2. Load Manager

La fase de Integración de Datos agrupa una serie de técnicas y subprocesos que ayudan a cumplir con las tareas relacionadas a la extracción, manipulación, control, integración, depuración de datos, carga y actualización el Data Warehouse. Es decir, todas las tareas que se realizan desde el momento que se toman datos de varias fuentes hasta que se cargan en el Data Warehouse. Este proceso es también conocido como los procesos ETL (Extracción, Transformación y Carga).

El proceso de Extracción, se realiza mediante un grupo de técnicas específicas para tomar la información indicada, en base a los diferentes criterios de negocio y así mantenerlos en un almacenamiento intermedio.

Subsiguiente proceso es la Transformación, en el cual se procede a realizar aquellas técnicas que analizan los datos para verificar que sean correctos y válidos.

Finalmente se tiene el proceso de Carga de Datos, se agruparán por ejemplo técnicas propias de la carga y actualización del DWH.

A continuación, se detallará cada uno de estos procesos; se presenta el procedimiento que lleva a cabo el ETL y se enumerarán cuáles son sus principales tareas

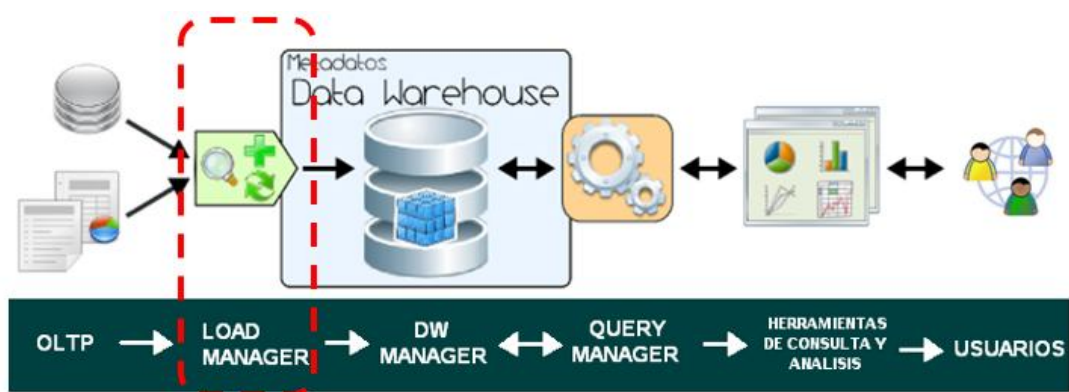


Figura 11. Selección Load Manager.

2.4.2.1. Extracción

Basándose en los requisitos de los usuarios, se estudian las diversas fuentes OLTP que se tengan a disposición, y en esta subfase se extrae la información que sea más importante y se considere relevante al caso.

Si los datos operacionales residen en un SGBD relacional, el proceso de extracción se puede reducir a consultas en SQL o rutinas programadas. En cambio, si se encuentran en un sistema no convencional o fuentes externas, ya sean textuales, hipertextuales, hojas de cálculos, etc., la obtención de los mismos puede ser un tanto más dificultoso, debido a que, por ejemplo, se tendrán que realizar cambios de formato y/o descarga de información a partir de alguna herramienta específica. (Dario B. , 2009)

Una vez que los datos son seleccionados y extraídos, se guardan en un almacenamiento intermedio, lo cual permite, entre otras ventajas:

- No depender de la disponibilidad de los OLTP.
- Facilitar la integración de las diversas fuentes, internas y externas.
- Almacenar y gestionar los metadatos que se generarán en los procesos ETL.
- Manipular los datos sin interrumpir, paralizar los OLTP, ni tampoco el DWH.

Esta área de almacenamiento intermedio, también conocida como área de Desembarco, es generalmente una base de datos donde la información puede ser almacenada en tablas temporales, tablas auxiliares, etc. estos datos son los que poblarán el Data Warehouse luego de su transformación.

2.4.2.2. Transformación

Como parte de la Integración de datos la Transformación convierte la información inconsistentes en un conjunto de datos compatibles y congruentes, para que puedan ser cargados en el Data Warehouse. Estas acciones se llevan a cabo debido a las variadas fuentes de las que se extrae la información, y es muy importante conciliar un formato y forma única, definiendo estándares para que todos los datos que ingresarán al Data Warehouse estén integrados y estructurados, bajo un lineamiento general de la empresa de forma técnica.

Los casos más comunes en los que se deberá realizar integración, son los siguientes:

- Tipos de Datos y Medida.
- Codificación.
- Fuentes múltiples.
- Convenciones de nombramiento.

Además de lo antes mencionado, esta subfase es la encargada de realizar, entre otros, los procesos de Limpieza de Datos (Data Cleansing) y Calidad de Datos.

2.4.2.2.1. Tipos de Datos y Medida

Los tipos de datos y su longitud que comúnmente se utiliza para representar los atributos de una entidad, varían considerablemente entre sí, a través de los diferentes OLTP. Así se tiene que al registrar el tamaño o la longitud de un activo determinado, de acuerdo a la aplicación que se emplee para tal fin, las unidades de depreciación puede ser expresada en años, meses, días, etc.

Como razonamiento de lo expuesto, se debe estandarizar los tipos de datos de los atributos y sus longitudes, para que todas las fuentes de datos expresen sus valores de igual manera.

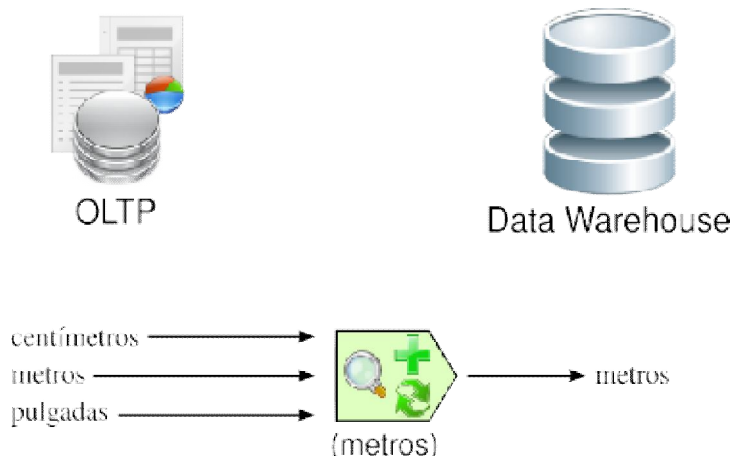


Figura 12. Transformación: medida de atributos

Fuente: (Bernabeu, OLTP, 2009).

2.4.2.2.2. Codificación

Una inconsistencia muy típica que se encuentra al intentar integrar varias fuentes de datos, es la de contar con más de una forma de codificar un atributo en común, así, en el campo “estado”, algunos diseñadores completan su valor con “0” y “1”, otros con “Apagado” y “Encendido”, otros con “off” y “on”, etc. Lo que se debe realizar en estos casos, es seleccionar o recodificar estos atributos, para que cuando la información llegue al DWH, esté integrada de manera uniforme. (Dario B. , 2009)

En la Figura 13., se puede apreciar la estandarización al seleccionar una de varias formas de codificar, entonces cuando surge una codificación diferente a la seleccionada, se procede a su transformación.

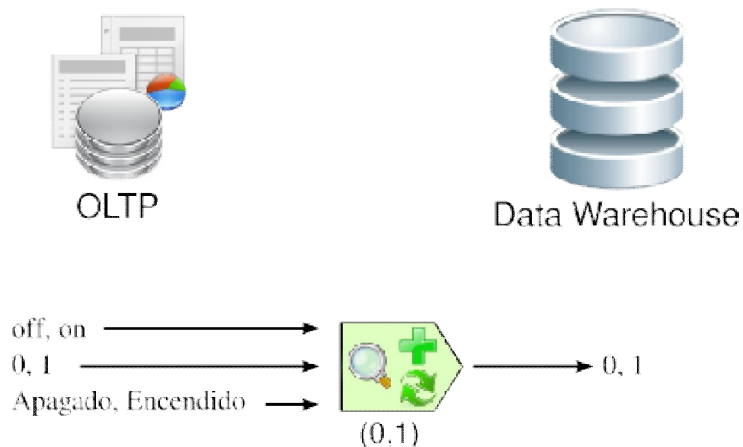


Figura 13. Transformación: codificación

Fuente: (Bernabeu, OLTP, 2009).

2.4.2.2.3. Fuentes múltiples.

Esta característica en un Data Warehouse, deriva de tener un mismo elemento en varias fuentes. El concepto de fuentes múltiples infiere que la información no se halla en un solo repositorio de información, sino que, se puede encontrar en varios tipos de repositorios de diferente tecnología, estos pueden ser, archivos planos, archivos Excel, bases de datos relacionales, base de datos orientadas a objetos.

Para resolver este problema, se debe elegir aquella fuente que se considere más fiable y apropiada, claro, luego de un análisis muy enfocado a la forma de realizar los procesos del negocio con el usuario.

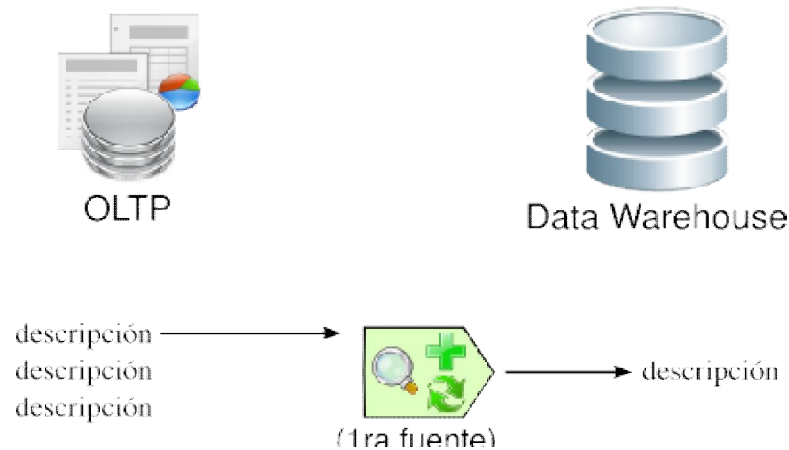


Figura 14. Transformación: fuentes múltiples

Fuente: (Bernabeu, OLTP, 2009).

2.4.2.2.4. Convenciones de nombramiento.

En la mayoría de los casos, al construir un Data Warehouse, un mismo atributo es nombrado de diversas maneras en las diferentes fuentes de extracción; por ejemplo, cuando se refiere al nombre de un cliente; así se tiene: “nombre”, “nombre_cliente”, “razon_social”, “cliente”, etc., se debe crear y utilizar la convención para el nombramiento del atributo que sea más comprensible para los usuarios.

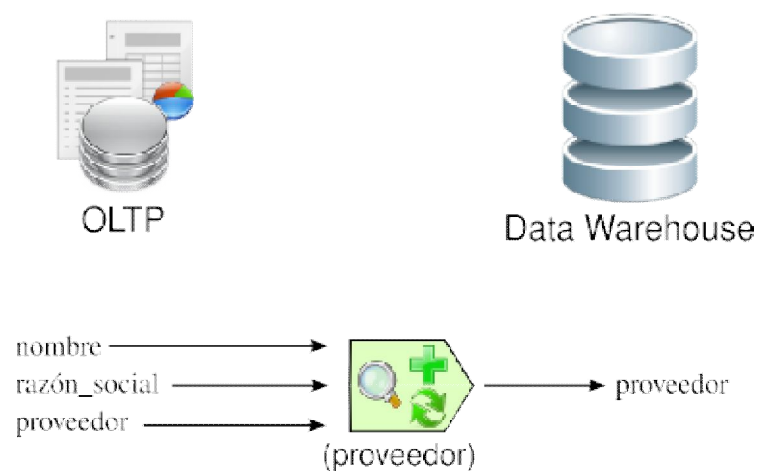


Figura 15. Transformación: convenciones de nombramiento

Fuente: (Bernabeu, OLTP, 2009).

2.4.2.2.5. Limpieza de datos

EL objetivo principal es filtrar los datos erróneos, inconsistentes e irrelevantes construyendo estrategias y acciones particulares para cada uno de los conceptos que se requiera cargar en el Data Warehouse.

Las acciones más típicas que se pueden llevar a cabo al encontrarse con Datos Anómalos (Outliers) son:

Tabla 4. Características de la Limpieza de datos.

Fuente: (Dario B. , 2009)

No.	Datos Anómalos (Outliers)	Datos Faltantes (Missing Values)
1	Ignorarlos.	Ignorarlos.
2	Eliminar la columna.	Eliminar la columna.
3	Filtrar la columna.	Filtrar la columna.
4	Filtrar la fila errónea, ya que a veces su origen, se debe a casos especiales.	Filtrar la fila errónea, ya que a veces su origen, se debe a casos especiales.
5	Reemplazar el valor	Reemplazar el valor
6	Discretizar los valores de las columnas. EJ.: de 1 a 2, poner "bajo"; de 3 a 7, "óptimo"; de 8 a 10, "alto".	Esperar hasta que los datos faltantes estén disponibles.

Un punto muy importante que se debe tener en cuenta al elegir alguna acción, es el de, identificar el porqué de la anomalía, para luego actuar en consecuencia, con el fin de evitar que se repitan, agregándole de esta manera más valor a los datos de la organización.

2.4.2.3. Carga

Esta función se encarga de realizar las tareas relacionadas con la carga inicial y carga diaria; siendo esta última dividida en dos: la Carga de registros nuevos y Actualización de los registros antiguos. Con más detalle se analiza los tipos de carga en la Figura 16.

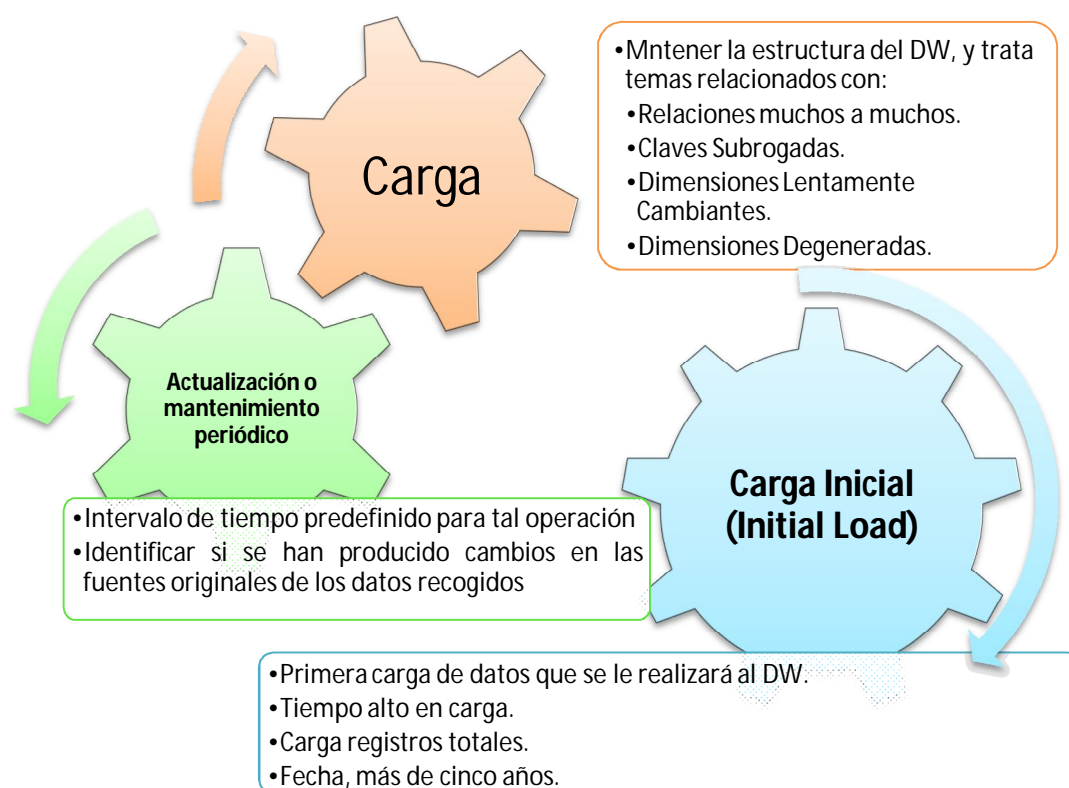


Figura 16. Características de la carga, análisis tipos de carga.

2.4.3. Data Warehouse Manager

El Data Warehouse Manager tiene como características y funciones:

- Generalmente se forma al combinar un SGBD con software y aplicaciones dedicadas de bases multidimensionales.

- Almacena la información de forma multidimensional, es decir, a través de tablas de hechos y tablas de dimensiones con o sin jerarquía.
- Gestiona las diferentes estructuras de datos que se construyan sobre el Data Warehouse, como los Cubos Multidimensionales y Business Models.
- Gestiona y mantiene metadatos de las estructuras creadas.
- Además, el DWH Manager se encarga de las siguientes actividades:
 - Transformar e integrar la información de las diversas fuentes de extracción y de almacenamiento intermedio en un modelo adecuado para la toma de decisiones.
 - Realizar todas las funciones de definición y manipulación del repositorio de datos, tanto para los temporales como para la estructura Data Warehouse, para poder soportar todos los procesos de gestión del mismo.
 - Ejecutar y definir las políticas de particionamiento de tablas. El objetivo de la tarea es obtener una mayor eficiencia y rendimiento en las consultas al aminorar el manejo del volumen de los datos. Esta política debe aplicarse sobre la tabla de hechos que es en la que se almacena toda la información que será analizada y bajo un proceso de análisis con el usuario sobre la forma de consultar la información.

- Realizar copias de respaldo, ya sean, incrementales o totales de la información del Data Warehouse, de una forma calendarizada y planificada.

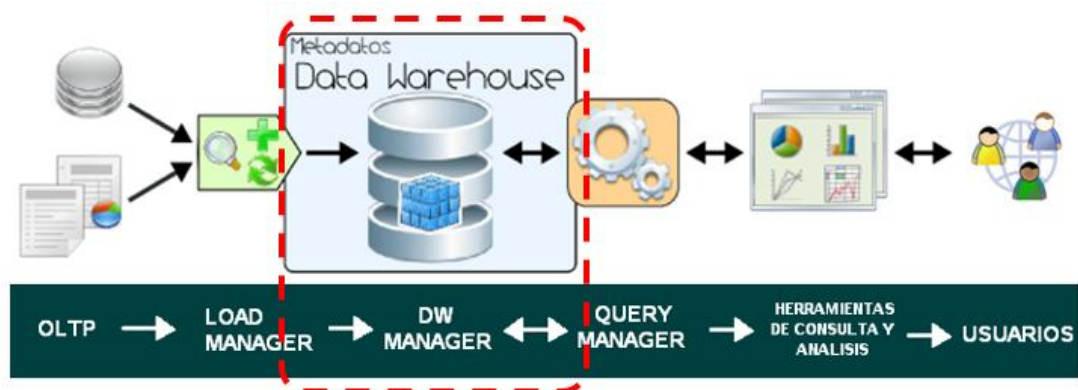


Figura 17. Selección Data Warehouse Manager.

2.4.3.1. Base de datos multidimensional

Una base de datos multidimensional es aquella base de datos en donde su información se almacena en una estructura multidimensional, conformada por tablas de dimensión y tablas de hechos. Se ha determinado conceptualmente tres variantes de modelamiento, puntualizando las consultas de soporte de decisión, estas son:

Esquema en Estrella (Star Scheme). El esquema en estrella está formada por una tabla de hechos base y de varias tablas de dimensiones relacionadas a esta, a través de las claves de relación. Es el esquema más simple de interpretar y permite optimizar los tiempos de respuesta ante las consultas de los usuarios. Este esquema, siendo el más eficiente, es soportado por la mayoría de las herramientas de consulta y análisis, y los metadatos son fáciles de documentar y mantener, pero hay que tomar en

cuenta que es el menos robusto para la carga de la información y es el más lento de construir.

Esquema Copo de Nieve (*Snowflake Scheme*). El esquema copo de nieve consiste en una tabla de hechos central relacionada con una o más tablas de dimensiones, las que a su vez pueden estar relacionadas con otras tablas de dimensiones. Este esquema representa el modelo en estrella pero con las dimensiones organizadas en forma jerarquías.

Esquema Constelación o copo de estrellas (*Star flake Scheme*). El esquema constelación está compuesto por una serie de esquemas en estrella. Consta de una tabla de hechos principal y una o más tablas de hechos auxiliares, relacionadas con sus respectivas tablas de dimensiones.

La implementación de estos esquemas puede ser diversa, pero generalmente requieren que toda la estructura de datos este desnormalizada totalmente o en un pequeño grado de normalización dependiendo del esquema; esto, con el objetivo de mejorar la ejecución de consultas y el análisis. Los diferentes tipos de implementación son los siguientes:

- Relacional — ROLAP. Esquema frecuentemente utilizado y el más antiguo con estructuras relacionales entre entidades.
- Multidimensional — MOLAP. Esquema actual que permite analizar la información cambiando la forma de la consulta en un Cubo.
- Híbrido — HOLAP. Combina los dos esquemas antes mencionados.

<u>ROLAP</u>	<u>MOLAP</u>
<p>Brinda mucha flexibilidad, ya que los cubos son generados dinámicamente al momento de ejecutar las consultas.</p>	<p>Cada vez que se requiere o es necesario realizar cambios sobre algún cubo, se debe tener que recalcularlo totalmente, para que se reflejen las modificaciones. Provocando de esta manera una disminución importante en cuanto a flexibilidad.</p>
<p>Los datos de los cubos se deben calcular cada vez que se ejecuta una consulta sobre ellos. Esto provoca que ROLAP no sea muy eficiente en cuanto a la rapidez de respuesta ante las consultas de los usuarios.</p>	<p>Las consultas son respondidas con mucha rapidez, ya que los mismos no deben ser calculados en tiempo de ejecución, obteniendo de esta manera un muy buen rendimiento.</p>

Figura 18. Cuadro comparativo entre la implementación ROLAP y MOLAP.

2.4.3.2. Tablas de Dimensión

Las tablas de dimensión permiten definir cómo están los datos organizados lógicamente y describen la forma para analizar el contexto del negocio; además se caracterizan por contener datos cualitativos. En la Figura 19., se muestran algunos ejemplos de clarificación.

<u>Ejemplo 1</u>	<u>Ejemplo 2</u>	<u>Ejemplo3</u>																					
<table border="1"> <thead> <tr> <th>LOCALIZACION</th> </tr> </thead> <tbody> <tr> <td><u>localizacionId</u> <pi></td> </tr> <tr> <td>pais</td> </tr> <tr> <td>provincia</td> </tr> <tr> <td>canton</td> </tr> <tr> <td>ciudad</td> </tr> <tr> <td>parroquia</td> </tr> <tr> <td>localizacionId <pi></td> </tr> </tbody> </table>	LOCALIZACION	<u>localizacionId</u> <pi>	pais	provincia	canton	ciudad	parroquia	localizacionId <pi>	<table border="1"> <thead> <tr> <th>TIEMPO</th> </tr> </thead> <tbody> <tr> <td><u>temporalidadId</u> <pi></td> </tr> <tr> <td>anio</td> </tr> <tr> <td>semestre</td> </tr> <tr> <td>trimestre</td> </tr> <tr> <td>mes</td> </tr> <tr> <td>dia</td> </tr> <tr> <td>temporalidadId <pi></td> </tr> </tbody> </table>	TIEMPO	<u>temporalidadId</u> <pi>	anio	semestre	trimestre	mes	dia	temporalidadId <pi>	<table border="1"> <thead> <tr> <th>CLIENTE</th> </tr> </thead> <tbody> <tr> <td><u>clienteId</u> <pi></td> </tr> <tr> <td>cedula</td> </tr> <tr> <td>nombreCompleto</td> </tr> <tr> <td>clienteId <pi></td> </tr> </tbody> </table>	CLIENTE	<u>clienteId</u> <pi>	cedula	nombreCompleto	clienteId <pi>
LOCALIZACION																							
<u>localizacionId</u> <pi>																							
pais																							
provincia																							
canton																							
ciudad																							
parroquia																							
localizacionId <pi>																							
TIEMPO																							
<u>temporalidadId</u> <pi>																							
anio																							
semestre																							
trimestre																							
mes																							
dia																							
temporalidadId <pi>																							
CLIENTE																							
<u>clienteId</u> <pi>																							
cedula																							
nombreCompleto																							
clienteId <pi>																							

Figura 19. Tablas de Dimensión zona geográfica, cliente y tiempo.

Como se puede observar en la Figura 19., cada tabla posee un identificador único como clave primaria y al menos un atributo de referencia que describe los criterios de análisis relevantes a el enfoque de análisis, estos son por lo general de tipo texto. Cada tabla de dimensión podrá contener los siguientes campos:

- Clave principal, que de la unicidad de la tabla.
- Claves foráneas, en una jerarquía.
- Datos de referencia primarios, datos que identifican la dimensión, por ejemplo, nombre del cliente.
- Datos de referencia secundarios: datos que complementan la descripción de la dimensión, por ejemplo, e-mail del cliente o fax del cliente.

2.4.3.2.1. Tablas de Dimensión Temporalidad.

En un Data Warehouse la creación y el mantenimiento de una tabla de dimensión de temporalidad es casi obligatorio, debido a su importancia en el

análisis de los datos; la definición de granularidad y estructuración de la misma depende de la dinámica del negocio. Esta tabla determina la ocurrencia de un hecho en una temporalidad específica, representando de esta manera, diferentes fotos de una misma situación en temporalidades distintas.

Es importante tener en cuenta que la dimensión de temporalidad ayuda de sobre manera a los usuarios a realizar el análisis sobre las operaciones realizadas teniendo en cuenta mes del año en que se produjeron, día, quincena, trimestre, semestre, año, estación, etc.

Así mismo, si se requiere analizar los datos por fecha expresada en (año, mes, día, etc.) y por hora expresada en (hora, minuto, segundo, etc.), lo más recomendable es confeccionar dos tablas de dimensión tiempo; una contendrá los datos referidos a la fecha y la otra los referidos a la hora. (Dario B. , 2009)

La recomendación es almacenar en la tabla de dimensión tiempo un campo que se refiera al día Juliano⁹. Almacenar este campo permitirá realizar consultas que involucren condiciones de filtrado de fechas desde-hasta, mayor que, menor que, etc., de manera sencilla. (Dario B. , 2009)

2.4.3.3. Tablas de Hechos

Las tablas de Hechos son la base central de los esquemas multidimensionales porque contienen datos cuantitativos, los cuales son

⁹ El día juliano se representa a través de un número secuencial e identifica unívocamente cada día.

instantáneos en el tiempo; y, permiten al usuario filtrarlos, agruparlos y explorarlos a través de condiciones definidas en las tablas de dimensiones.

Toda la información presente en las tablas de hechos constituyen el volumen de la bodega, y pueden estar compuestos por millones de registros dependiendo del nivel de detalle y la historia que el negocio requiere alcanzar.

Para el registro del hecho, este debe poseer una clave primaria única compuesta por las claves primarias de las tablas de dimensiones relacionadas. En la Figura 20., se clarifica el concepto con un ejemplo del diagrama físico que se genera.

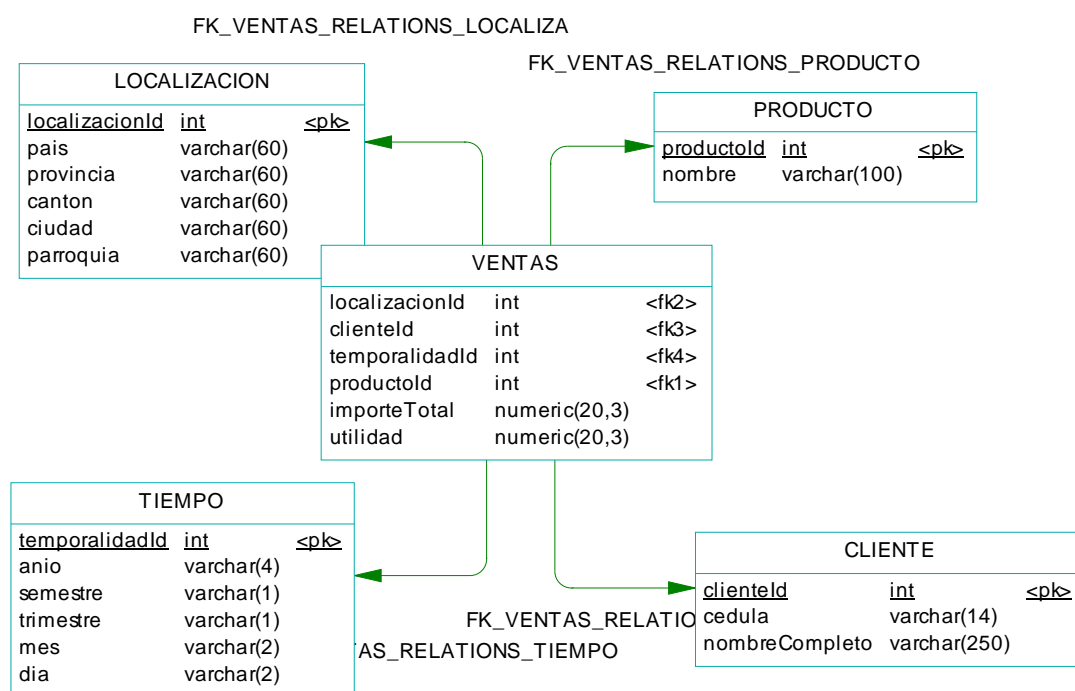


Figura 20. Representación modelo físico de tipo estrella (Hechos y dimensiones).

2.4.4. Query Manager

Esta función realiza las operaciones, tales como soportar los procesos de gestión y ejecución de consultas relacionales (*Join* y agregaciones), y de consultas propias del análisis de datos (*drill-up* y *drill-down*).

El componente recibe las consultas de los usuarios, las aplica a la estructura de datos correspondiente (cubo multidimensional, Business Models, etc.) y devuelve los resultados obtenidos. Generalmente esta tarea consiste en la obtención de indicadores a partir de los datos que se encuentran en la tabla de hechos; estos restringidos por las propiedades o condiciones de los atributos que hayan sido creados.

Las operaciones que se pueden realizar sobre modelos multidimensionales y que son las que verdaderamente les permitirán a los usuarios explorar e investigar los datos en busca de respuestas, son:

- **Drill-down.** Permite apreciar los datos en un mayor detalle.
- **Drill-up.** Permite apreciar los datos en menor nivel de detalle.
- **Drill-across.** Funciona de forma similar a *drill-down* pero agrega un atributo a la consulta como nuevo criterio de análisis.
- **Roll-across.** Funciona de forma similar a *drill-down* pero quita un atributo a la consulta eliminando un criterio de análisis.
- **Pivot.** Permite seleccionar el orden de visualización de los atributos e indicadores.

- **Page.** Presenta el cubo dividido en secciones, a través de los valores de un atributo como páginas de un libro.
- **Drill-through.** Permite apreciar los datos en su máximo nivel de detalle.

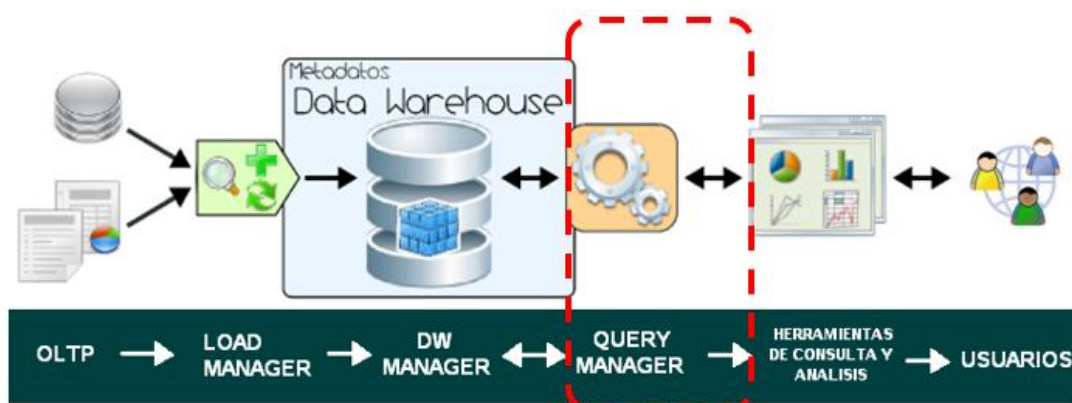


Figura 21. Elemento Query Manager en la Arquitectura Data Warehousing

Fuente: (Bernabeu, Query Manager, Arquitectura Data Warehouse, 2009).

2.4.5. Herramientas de consulta y análisis.

Las herramientas se basan en sistemas para realizar consultas sobre los datos que se encuentran en el Data Warehouse. Estas consultas se realizan a través de una interfaz gráfica y una serie de pasos, los usuarios generan consultas al *Query Manager*, este a su vez realiza la extracción de información al *Data Warehouse Manager* y los resultados obtenidos se los devuelve a la herramienta que se los solicitó para ser presentada al usuario. Las herramientas más usadas son:

- Reportes y Consultas.
- Dashboards.
- Data Mining.

EIS (Executive Information System) es un componente que proporciona medios sencillos para consultar, analizar y acceder a la información del negocio que se encuentra almacenada en el Data Warehouse.

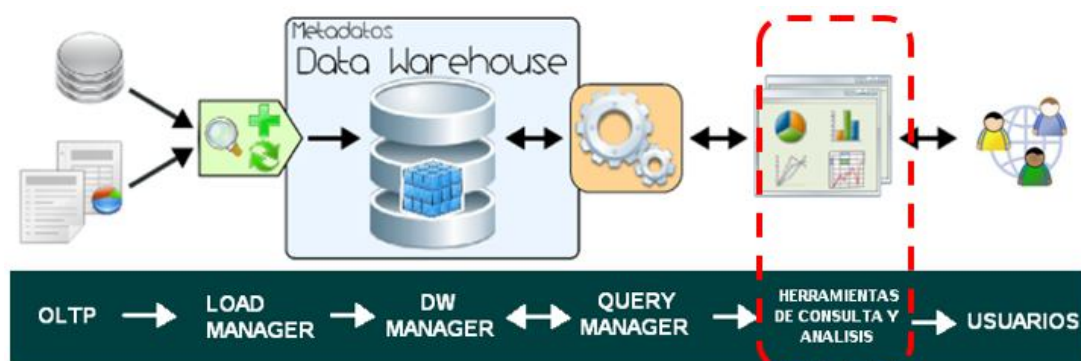


Figura 22. Herramientas de consulta y análisis, Arquitectura DWH

Fuente: (Bernabeu, Herramientas de consulta y análisis, Arquitectura Data Warehouse, 2009).

2.4.6. Usuarios.

Los usuarios que posee el Data Warehouse son aquellos que se encargan de tomar decisiones y de planificar las actividades del negocio a través de las herramientas mencionadas buscando respuestas a la situación del negocio; siendo estos parte clave del proceso de análisis inicial en la integración, limpieza de datos, etc., para poder conseguir que la información posea toda la calidad posible.

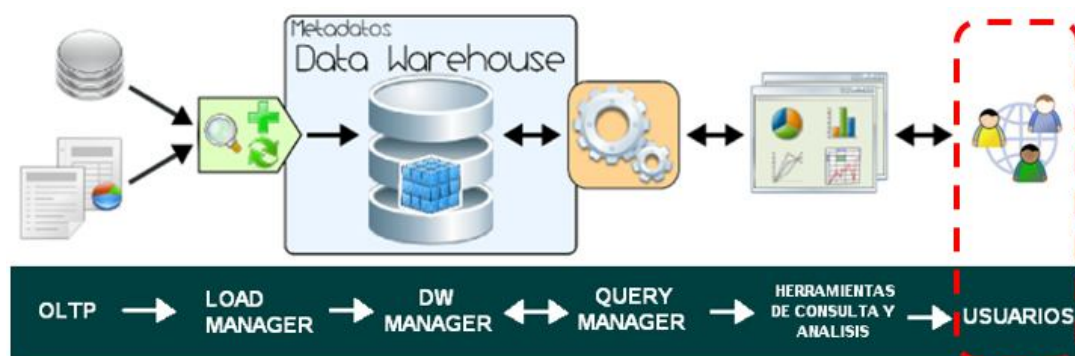
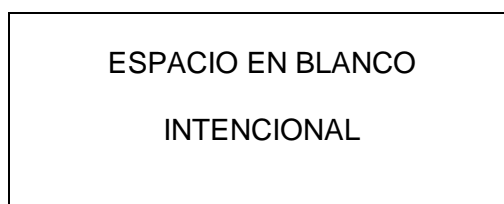


Figura 23. Elemento Usuarios, en la Arquitectura Data Warehousing

Fuente:(Bernabeu, Arquitectura Data Warehouse, 2009).

2.4.7. Ventajas y Desventajas del Data Warehousing

A continuación se presenta un esquema de las ventajas y desventajas al implementar un Data Warehousing:



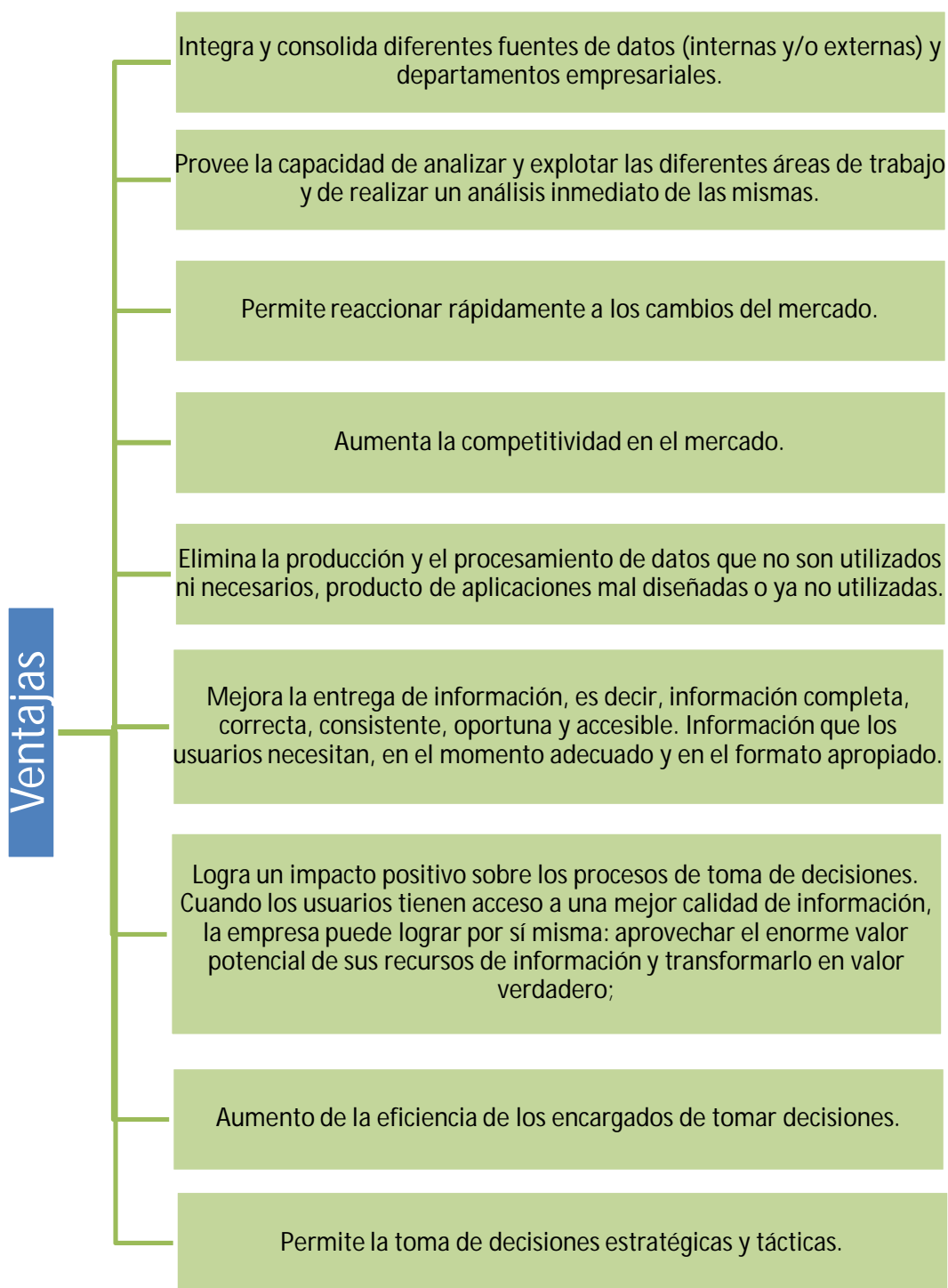


Figura 24. Listado de ventajas que ofrece un Data Warehousing.

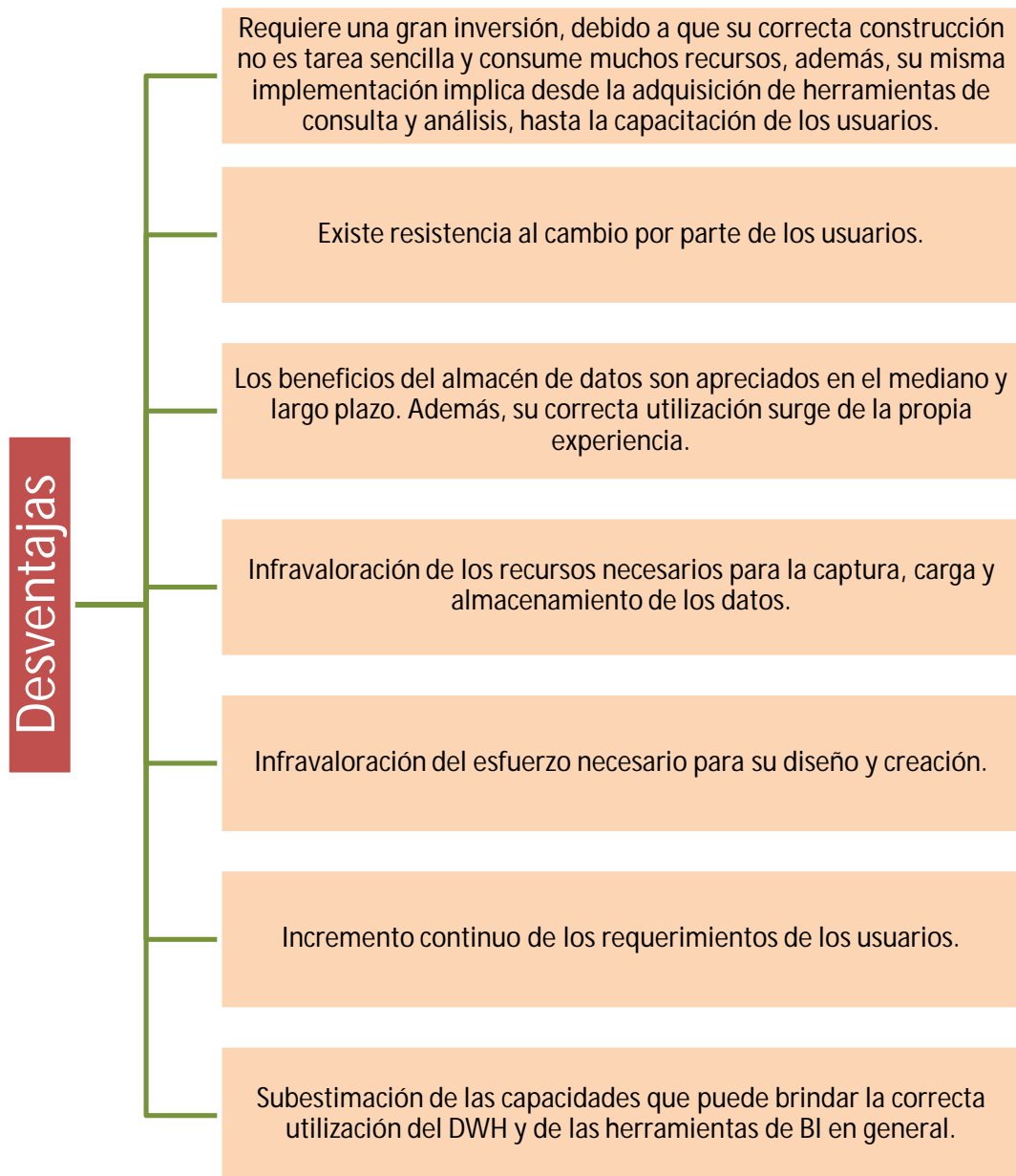


Figura 25. Listado de desventajas de un Data Warehousing.

2.5. Indicador de gestión (KPI).

KPI (Key Performance Indicators), o Indicadores Clave de Desempeño, miden el nivel del desempeño de un proceso, enfocándose en el "cómo" evaluando el resultado de los procesos y su cumplimiento. (Sixtinagroup)

Los indicadores clave de desempeño son métricas, usadas para evaluar el rendimiento de un proceso, persona o entidad organizacional de forma cuantificativa en base a sus objetivos planteados, que generalmente se recogen en su plan estratégico.

Los indicadores son indispensables para la mejora continua, puesto que "lo que no se mide no se puede controlar, y lo que no se controla no se puede gestionar".

Por tanto KPI no es un término tecnológico, generado por el *Business Intelligence*, sino, es un concepto ligado a la *Gestión Empresarial*. No obstante, el desarrollo de la tecnología y de especialidades como el *Business Intelligence*, han permitido que el proceso para medir, controlar y presentar la información se haga de un modo mucho más eficiente y rápido.

2.5.1. Importancia de los Indicadores de Gestión Empresarial

Los indicadores estratégicos son importantes en la Gestión Empresarial debido a que:

- Ayudan a interpretar lo que está ocurriendo en la organización.
- Definen la necesidad de introducir cambios y/o mejoras a un proceso determinado o actuación.

- Sirven como apoyo importante al proceso de toma de decisiones estratégicas a los comités gerenciales, especialmente cuando las variables se salen de los límites establecidos, o se requiere proponer una nueva meta.
- Facilitan el compromiso de mejores resultados.

2.5.2. La Clave del Proceso de Selección de KPI.

Las mediciones más comunes apuntan a tener indicadores como los siguientes:

- Calidad de los productos y servicios.
- Rentabilidad del negocio.
- Cumplimiento de planificaciones.
- Productividad de los empleados.
- Tiempos de desarrollo de trabajos.
- Eficacia de los procesos.
- Eficacia en los cobros.
- Crecimiento.
- Control de costos.
- Uso de los recursos.
- Nivel de innovación y desempeño de la infraestructura tecnológica.

De forma generalizada, los indicadores ayudan a las entidades organizacionales a determinar si los recursos y costos se están manejando acertadamente, ayudando a que el comité gerencial tenga una noción clara de lo que acontece en un momento específico para tomar medidas correctivas y oportunas.

Para poder definir KPI se suele aplicar el acrónimo SMART, ya que los indicadores tienen que ser:

- **e**Specíficos (Specific)
- **M**edibles (Measurable)
- **A**lcanzables (Achievable)
- **R**ealista (Realistic)
- **a** Tiempo (Timely)

ESPACIO EN BLANCO
INTENCIONAL

2.5.3. Clasificación.

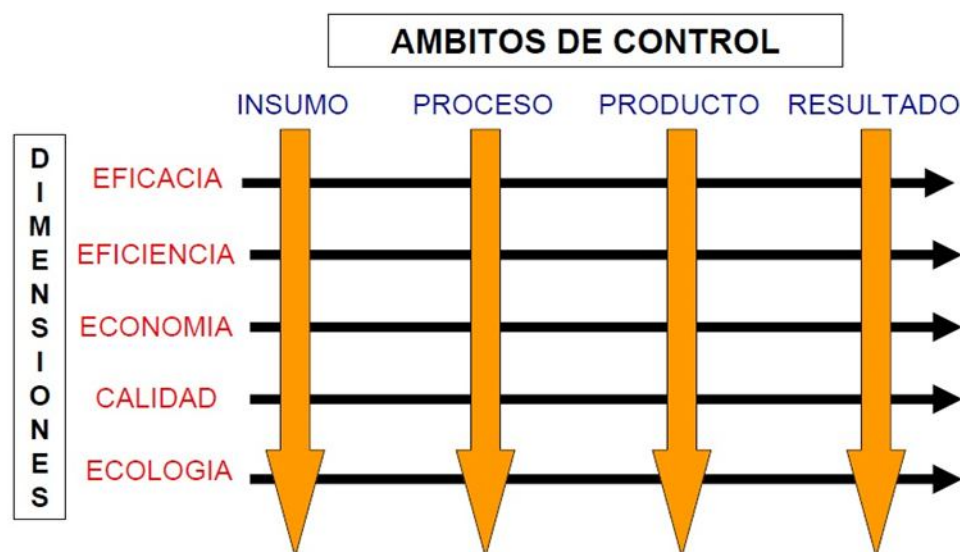


Figura 26. Clasificación de los indicadores

Fuente: (Bernabeu, Clasificación KPI, 2013).

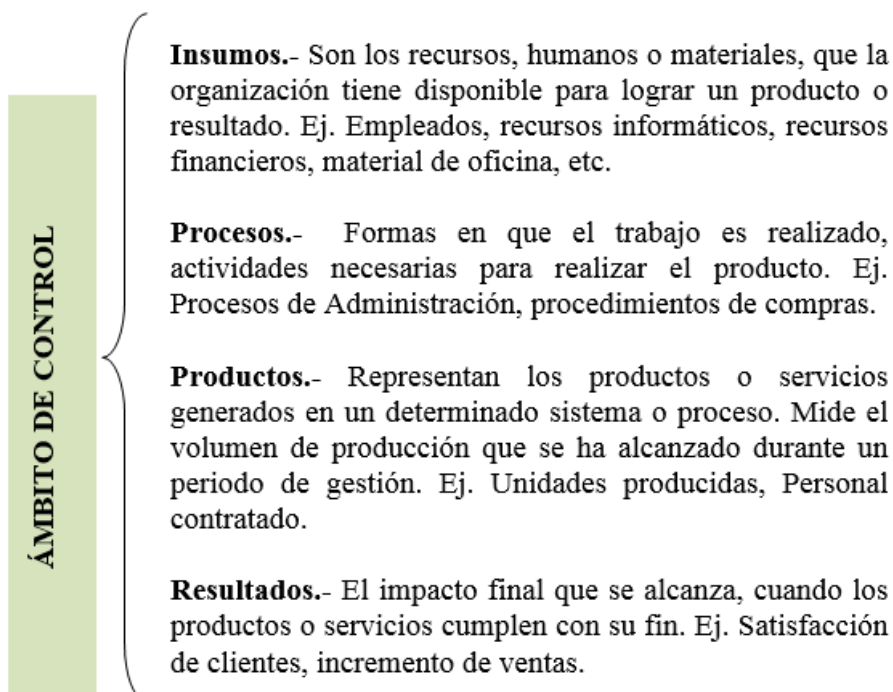


Figura 27. Clasificación de los indicadores por el ámbito de control

Fuente: (Bernabeu, Clasificación KPI, 2013).

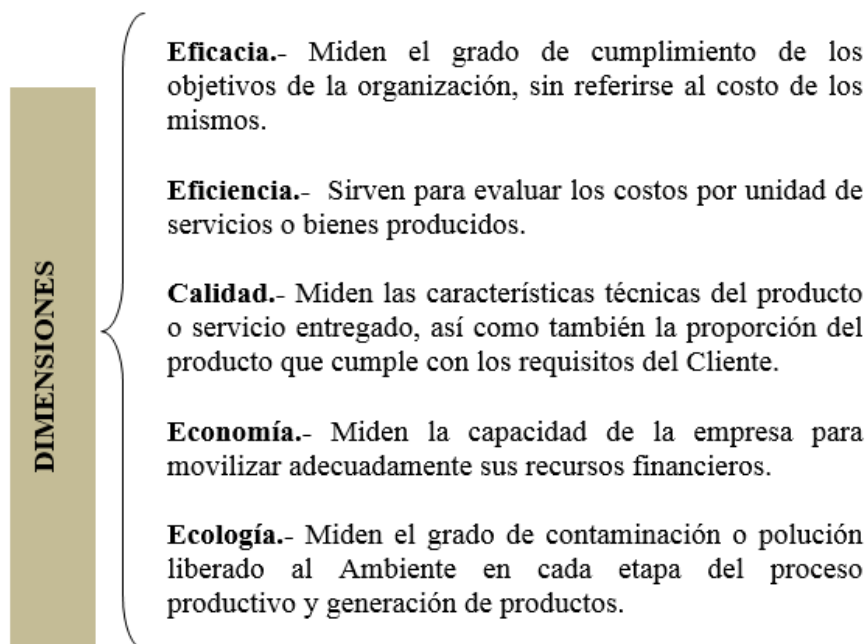


Figura 28. Clasificación de los indicadores por dimensión

Fuente: (Bernabeu, Clasificación KPI, 2013).

2.5.4. Sistema de Indicadores

Un sistema de Indicadores está estructurado por: el Indicador, el nivel base, el valor actual, la meta, y el uso de semáforos para la evaluación del desempeño del Indicador.

Las tareas y metas que se propone alcanzar una unidad organizacional, deben expresarse de una forma medible, que permitan evaluar el grado de cumplimiento o avance de los mismos. Es aquí donde el uso de **indicadores** tiene mayor fortaleza. Los indicadores pueden ser de tipo positivos o negativos.

Indicadores Positivos.- Son aquellos indicadores, que un **aumento** en su valor o tendencia, estaría indicando un avance hacia la situación deseada. El nivel de cumplimiento o desempeño, se mide mediante la fórmula 1:

$$Desempeño = \left[\frac{Valor - Base}{Meta - Base} \right] * 100\%$$

Fórmula 1. Cálculo del desempeño positivo de un empleado
(SIXTINA CONSULTING GROUP, 2012).

Indicadores Negativos.- Son aquellos indicadores, que una **disminución** de su valor o tendencia, estaría indicando un avance en la situación deseada. Su nivel de desempeño se mide por:

$$Desempeño = \left[\frac{Base - Valor}{Base - Meta} \right] * 100\%$$

Fórmula 2. Cálculo del desempeño negativo de un empleado
(SIXTINA CONSULTING GROUP, 2012).

Base.- Nivel Base, refiérase a la métrica inicial o nivel estándar que toma el indicador, y representa el desempeño logrado antes del efecto de mejora de las iniciativas estratégicas.

Valor.- Es el valor actual que representa las mediciones período a período del indicador, las cuales se ven afectadas por los efectos de las iniciativas estratégicas.

Meta.- Es el nivel esperado del indicador que la organización desea lograr luego de ejecutar exitosamente las acciones de mejora.

Semáforos.- Los semáforos representan una manera más fácil de observar el nivel de desempeño de los indicadores; donde el color verde

representa un desempeño esperado o mejorado, el color amarillo un desempeño que puede alarmar el objetivo atado al indicador y el rojo indica un desempeño inaceptable.

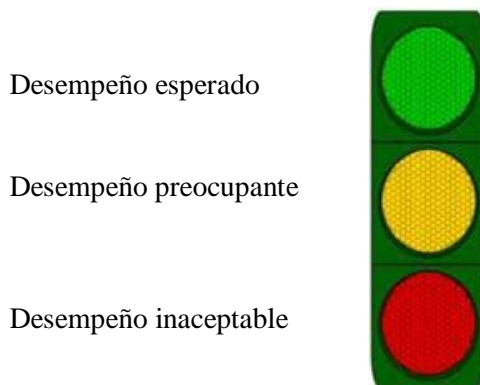


Figura 29. Semáforo medición visual de desempeño.

2.6. Análisis de las Metodologías para Business Intelligence.

Dentro de este análisis se basarán las dos metodologías ágiles más conocidas para el desarrollo de proyectos BI, estas son Metodología Hefesto y la Metodología desarrollada por SAS (The SAS Rapid Data Warehouse Methodology).

Metodología Hefesto, Como una gran aportación se menciona a continuación, fue creada por Bernabéu Ricardo Darío (disponible con licencia GNU FDL). El libro es un resumen muy completo de todo lo relacionado con el Business Intelligence y los DWH, y puede ser un punto de partida de gran calidad para entrar en materia. En la segunda parte del libro se desarrolla la metodología Hefesto, creada y revisada por esta persona, que además ha compartido con todo el mundo con licencia GNU su completo trabajo.

La metodología está orientada a la construcción de DWH para Análisis Dimensional (OLAP) y comprende las siguientes fases:

- Análisis de Requerimientos.
- Análisis de los OLTP.
- Modelo Lógico del DWH.
- Procesos ETL.

Metodología SAS, creada por SAS Institute, la cual es iterativa, y está basada en el desarrollo incremental del proyecto de Data Warehouse dividido en cinco fases que son:

- Definición de los objetivos.
- Definición de los requerimientos de información.
- Diseño y modelización.
- Implementación.
- Revisión.

Tabla 5. Comparación metodologías Hefesto vs. SAS.

No.		Hefesto	SAS	Metodología más adecuada para el proyecto
1	Tamaño del Proyecto	Pequeños - Medianos	Medianos-Grandes	Hefesto
2	Tiempo en el análisis y diseño	Medio una sola vez	Extenso por ser iterativo	Hefesto
3	Tiempo en construcción.	Medio	Medio	Ambos
4	Etapas de implantación	NO	SI	SAS
5	Guías y prácticas se aplican a SQL	SI	Algunas	Hefesto
6	Fácil entendimiento principiantes	SI	NO	Hefesto
7	Revisión Post Implantación	NO	SI	SAS

Una vez realizada la comparación entre las dos metodologías se puede seleccionar Hefesto como la metodología más compatible para desarrollar el proyecto SATB.

2.7. Metodología Hefesto.

La Metodología Hefesto está enfocada en dar lineamientos para la construcción de Data Warehouse de una manera ordenada, entendible, útil y sencilla, para que quien la revise sepa lo que está haciendo.

La construcción e implementación de un Data Warehouse posee una gran adaptabilidad a cualquier ciclo de vida de desarrollo de software; excepto en algunas fases en particular, las acciones que se han de realizar serán muy diferentes. Lo que se debe tener muy en cuenta, es no entrar en la utilización de metodologías que requieran fases extensas para reunir los requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas.

2.7.1. Características.

Esta metodología cuenta con las siguientes características:

- Está basada en los requerimientos de los usuarios, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa para que tome decisiones respecto al comportamiento y funciones del Data Warehouse.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el DWH y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- Se aplica tanto para Data Warehouse como para Data Mart.

2.7.2. Pasos y aplicación de la Metodología.

Para comenzar a utilizar Hefesto, se debe considerar un marco previo de análisis, describiendo las características principales de la empresa, como son: misión, visión, objetivos, organigrama, políticas, estrategias, metas, procesos involucrados; esto para comprender mejor el funcionamiento y accionar de la empresa, interpretando cada decisión que se tome con respecto a la implementación y diseño del DWH. En la Figura 30., se muestra el esquema metodológico.

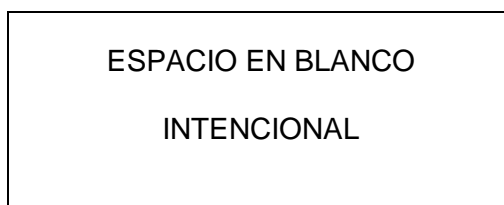




Figura 30. Metodología Hefesto

Fuente: (Ricardo, 2010).

2.7.2.1. Paso 1. Análisis de Requerimientos.

2.7.2.1.1. Identificar preguntas

Se recolectan las necesidades de información de los usuarios y se obtienen las preguntas claves del negocio. Estas puedan llevarse a cabo usando varias técnicas como: entrevistas, lluvia de ideas, cuestionarios, observaciones, etc.

El objetivo principal de esta fase, es la de obtener e identificar las necesidades de información clave de alto nivel, que es esencial para llevar a cabo las metas y estrategias de la empresa, y que facilitará una eficaz y eficiente toma de decisiones. Estas necesidades pueden ser guiadas por los procesos principales existentes en la empresa, generando así variables de análisis como las ventas por una determinada fecha, se debe tener en cuenta que no todo se puede abarcar en este análisis sino establecer prioridades de análisis, para luego en una siguiente fase ir incrementado más procesos de análisis. Un ejemplo del resultado de análisis sería:

- Se desea conocer la cantidad de unidades facturadas de las publicaciones a los clientes, en cada zona y en un periodo de tiempo determinado.
- Se desea conocer la cantidad de unidades de reposición de las publicaciones a los clientes, en cada zona y en un periodo de tiempo determinado.
- Se desea conocer la cantidad de unidades vendidas de las publicaciones a los clientes, en cada zona y en un periodo de tiempo determinado.
- Se desea conocer el porcentaje de devoluciones de las publicaciones realizadas por los clientes, en cada zona y en un periodo de tiempo determinado, etc.

2.7.2.1.2. Identificar indicadores y perspectivas

Luego, se deben identificar los indicadores, generalmente valores numéricos, resultantes de los interrogativos y sus respectivas perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del DWH, con el fin de contestar las preguntas planteadas.

Preguntas	Indicadores	Perspectivas
<ul style="list-style-type: none"> • <i>Unidades facturadas</i> de las <i>publicaciones</i> a los <i>clientes</i>, en cada <i>zona</i> y en un periodo de <i>tiempo</i> determinado • <i>Unidades de reposición</i> de las <i>publicaciones</i> a los <i>clientes</i>, en cada <i>zona</i> y en un periodo de <i>tiempo</i> determinado • <i>Unidades vendidas</i> de las <i>publicaciones</i> a los <i>clientes</i>, en cada <i>zona</i> y en un periodo de <i>tiempo</i> determinado. • <i>Porcentaje de devoluciones</i> de las <i>publicaciones</i> a los <i>clientes</i>, en cada <i>zona</i> y en un periodo de <i>tiempo</i> determinado 	<ul style="list-style-type: none"> • Unidades Facturadas. • Unidades de Reposición. • Unidades Vendidas. • Porcentaje de devoluciones 	<ul style="list-style-type: none"> • Clientes • Publicaciones • Zonas • Tiempo

Figura 31. Análisis de requerimientos, obtención de indicadores y perspectivas.

La Figura 31., representa el análisis realizado para obtener los indicadores y perspectivas en base a las preguntas realizadas en el punto 2.7.2.1.1.

2.7.2.1.3. Modelo Conceptual

El modelo conceptual se basa en el análisis del anterior punto, el cual permite identificar el indicador que se va a modelar en conjunto a sus perspectivas y medidas. Así se tiene el siguiente modelo:

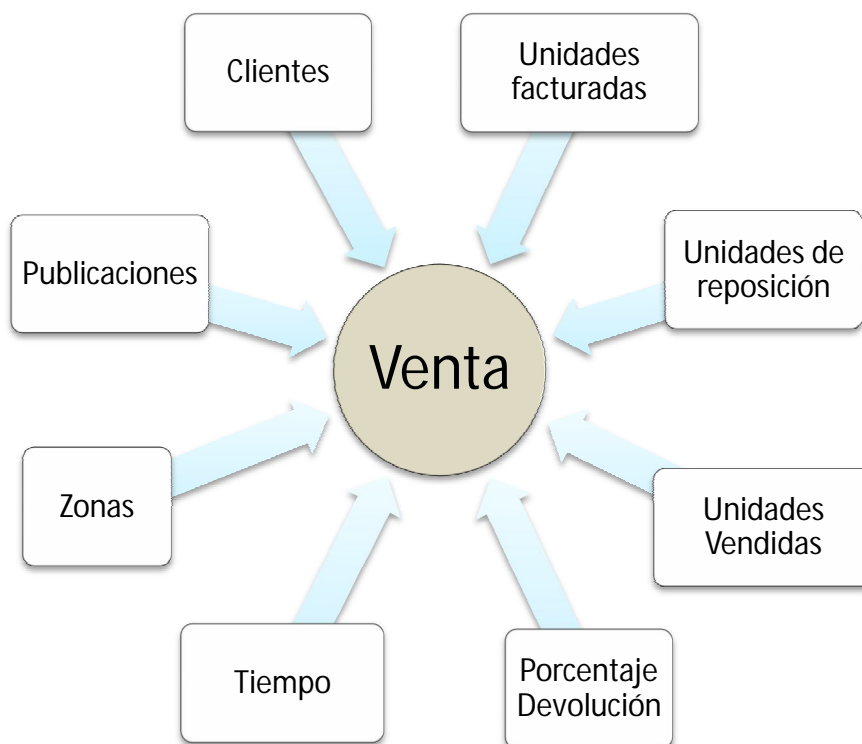


Figura 32. Modelo Conceptual.

2.7.2.2. Paso 2. Análisis de los OLTP.

Se analiza los OLTP para determinar, cómo se construirán los indicadores, señalar las correspondencias con los datos fuentes y para seleccionar los campos de estudio de cada perspectiva.

2.7.2.2.1. Construcción de los Indicadores

- Unidades Facturadas:

Hechos: Unidades Facturadas

Función de sumarización: SUM

Este indicador representa la sumatoria de las unidades que se han facturado de una publicación en particular.

- Unidades de Reposición:

Hechos: Unidades de Reposición

Función de sumarización: SUM

Este indicador representa la sumatoria de las unidades facturadas como reposición de una publicación en particular.

- Unidades Vendidas:

Hechos: Unidades Facturadas + Unidades de Reposición - Unidades Acreditadas

Función de sumarización: SUM

Este indicador representa la sumatoria de las unidades que se han facturado de una publicación en particular, incluyendo las reposiciones y descontando las unidades acreditadas de la misma.

- Porcentaje de Devoluciones:

Hechos: $(\text{Unidades Acreditadas} / (\text{Unidades Facturadas} + \text{Unidades de Reposición})) * 100$

Función de sumarización: SUM

Este indicador representa la relación entre las sumatorias de las unidades acreditadas y las unidades facturadas, incluyendo las reposiciones, de cada publicación, y se obtiene al dividir estas cantidades, respectivamente y multiplicarlas por 100.

2.7.2.2.2. Establecer Correspondencia.

Para realizar este paso se debe tener un esquema claro del diagrama del sistema operacional que tiene la empresa a ser analizada.

Se puede implementar una matriz de mapeo de fuente, indicando de qué base de datos, tablas y campos; es la correspondencia de cada perspectiva o indicador.

TIPO	VARIABLE	TABLA	TIPO FUENTE	BASE DE DATOS	RUTA	CAMPOS	FORMULA
PERSPECTIVA	CLIENTES	VENEDORES	BASE DE DATOS	BANCS	BDD=P012BA	TODOS	
PERSPECTIVA	PUBLICACIONES	INGRESOS	BASE DE DATOS	BANCS	BDD=P012BA	TODOS	
PERSPECTIVA	ZONAS	RECORRIDOS	BASE DE DATOS	BANCS	BDD=P012BA	TODOS	
INDICADOR	UNIDADES FACTURADAS	HISTORICOS	BASE DE DATOS	PORTAL	BDD=P012BA	CANTFACT	
INDICADOR	PORCENTAJE DEVOLUCIÓN	HISTORICOS	BASE DE DATOS	PORTAL	BDD=P012BA	DEVOLUCION CANTFACT RPOSICION	$i = \frac{DEVOLUCION}{(CANTFACT + RPOSICION)}$

Figura 33. Ejemplo. Matriz de mapeo Fuente OLTP / Variables.

2.7.2.2.3. Nivel de granularidad.

Para definir el nivel de granularidad se debe examinar y seleccionar los campos que contendrá cada perspectiva, ya que será a través de estos por los que se manipularán y filtrarán los indicadores. Primero viene la descripción del campo; para esto, de la matriz Fuente OLTP definida, se puede agregar una columna en la cual se describa el concepto del campo mediante una previa investigación de su sentido, ya sea consultado a los expertos, diccionarios o mediante reuniones con el usuario.

TIPO	VARIABLE	TABLA	TIPO FUENTE	BASE DE DATOS	RUTA	CAMPOS	FORMULA	DESCRIPCION
PERSPECTIVA	CLIENTES	VENEDORES	BASE DE DATOS	BANCS	BDD=P012B	CODVEND		Representa el código del vendedor en el sistema operacional.
PERSPECTIVA	CLIENTES	VENEDORES	BASE DE DATOS	BANCS	BDD=P012B	RAZONSOC		Si es una empresa representa el nombre de esta.
PERSPECTIVA	CLIENTES	VENEDORES	BASE DE DATOS	BANCS	BDD=P012B	TELEFONO		Número convencional o celular del vendedor.
PERSPECTIVA	CLIENTES	VENEDORES	BASE DE DATOS	BANCS	BDD=P012B	PAGOATRAS		Último pago realizado de una venta realizada.
PERSPECTIVA	CLIENTES	VENEDORES	BASE DE DATOS	BANCS	BDD=P012B	CONSIGNA		Monto consignado para operar.

Figura 34. Ejemplo. Matriz de mapeo OLTP, descripción de campo.

Luego de exponer frente al usuario los datos existentes, explicando su significado, valores posibles y características, este debe decidir cuáles son los que se considera relevantes para consultar los indicadores y cuáles no.

Con respecto a la perspectiva “Tiempo”, que es la pieza fundamental que determinará la granularidad del depósito de datos, los datos más típicos que pueden emplearse son los siguientes:

- Año
- Semestre
- Cuatrimestre
- Trimestre
- Número de mes
- Nombre del mes
- Quincena
- Decena

- Semana
- Número de día
- Nombre del día
- Estación del año

2.7.2.2.4. Modelo Conceptual ampliado.

Es el Modelo Conceptual original, pero incluido los atributos que servirán para el análisis del cliente.

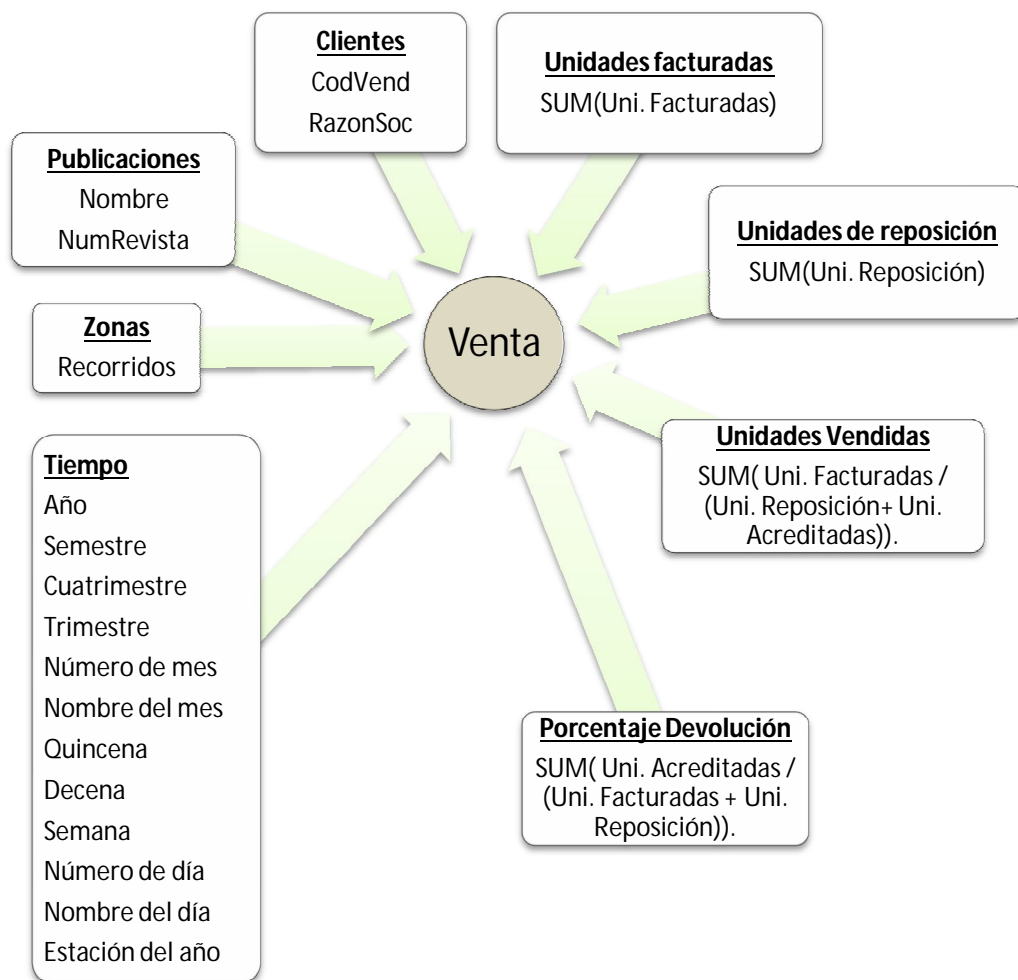


Figura 35. Modelo Conceptual ampliado.

2.7.2.3. Paso 3. Modelo lógico del DWH

Este paso, modelo lógico¹⁰, define cuál será el tipo de esquema que se implementará. Después, se confeccionarán las tablas de dimensiones y las tablas de hechos, para luego efectuar sus respectivas uniones.

Se debe seleccionar cuál será el tipo de esquema que se utilizará para contener la estructura del depósito de datos, que se adapte mejor a los requerimientos y necesidades del usuario. Es muy importante definir objetivamente que tipo de esquema se empleará [15], ya que esta decisión afectará considerablemente la elaboración del modelo lógico de la bodega de datos.

2.7.2.3.1. Tipo de Modelo lógico.

Para el caso de estudio se define un esquema de tipo estrella¹¹, debido a las siguientes características que ayudan a la construcción del modelo por sus ventajas:

- Posee los mejores tiempos de respuesta.
- Su diseño es fácilmente modificable.
- Existe paralelismo entre su diseño y la forma en que los usuarios visualizan y manipulan los datos.
- Simplifica el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

¹⁰ Modelo Lógico, es la representación de una estructura de datos, que puede procesarse y almacenarse en algún SGBD.

¹¹ Los esquemas de modelamiento se revisaron en el punto 2.4.3.1. de este documento.

2.7.2.3.2. Diseño tablas de dimensiones.

Este paso, se aplicará por igual a todos los tipos de esquemas lógicos. Lo primero que se hará será crear las dimensiones del mismo, para ello se tomará cada perspectiva con sus atributos relacionados y se les realizará el siguiente proceso:

- Se elegirá un nombre que identifique la dimensión.
- Se añadirá un campo que represente su clave principal.
- Se redefinirán los nombres de los atributos si es que no son lo bastante explicativos.

A continuación, se presentan los diseños de las tablas de dimensiones del caso de estudio presentado:

Perspectiva “Clientes”,

La nueva tabla de dimensión tiene el nombre “CLIENTE”.

Se le agregó una clave principal con el nombre “idCliente”.

Se modificó el nombre del campo “CodVend” por “CodCliente”.

Se modificó el nombre del campo “RazonSoc” por “Cliente”.

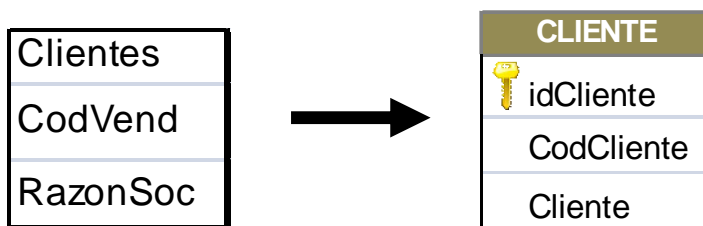


Figura 36. Diseño dimensión Clientes.

Perspectiva “Publicaciones”,

La nueva tabla de dimensión tiene el nombre “PUBLICACION”.

Se le agregó una clave principal con el nombre “idPublicacion”.

Se modificó el nombre del campo “Nombre” por “Publicacion”.

Se modificó el nombre del campo “NumRevista” por “Edicion”.

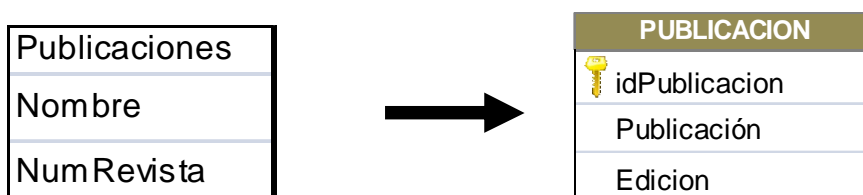


Figura 37. Diseño dimensión Publicaciones.

Perspectiva “Zonas”,

La nueva tabla de dimensión tiene el nombre “ZONA”.

Se le agregó una clave principal con el nombre “idZona”.

Se modificó el nombre del campo “Recorrido” por “Zona”.

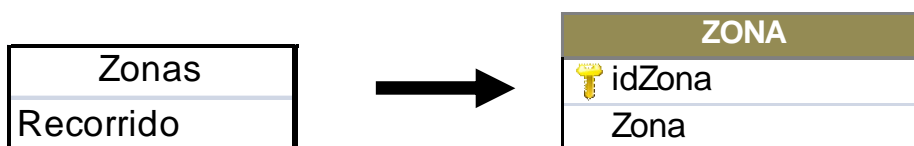


Figura 38. Diseño dimensión Zonas.

Perspectiva “Tiempo”,

La nueva tabla de dimensión tiene el nombre “FECHA”.

Se le agregó una clave principal con el nombre “idFecha”.

Se modificó el nombre del campo “Año” por “Anio”.

Se modificó el nombre del campo “Número del mes” por “NumeroMes”.

Se modificó el nombre del campo “Nombre del mes” por “NombreMes”.

Se modificó el nombre del campo “Quincena” por “QuincenaMes”.

Se modificó el nombre del campo “Decena” por “DecenaMes”.

Se modificó el nombre del campo “Semana” por “SemanaMes”.

Se modificó el nombre del campo “Número del día” por “NumeroDia”.

Se modificó el nombre del campo “Nombre del día” por “NombreDia”.

Se modificó el nombre del campo “Estación del año” por “Estacion”.



Figura 39. Diseño dimensión Tiempo.

2.7.2.3.3. Tablas de hechos.

En este paso, se definirá las tablas de hechos, que contendrán los indicadores de estudio. Para los esquemas en estrella y copo de nieve, se realizará lo siguiente:

- Al igual que las dimensiones, se le deberá asignar un nombre a la tabla de hechos que en este caso represente la información analizada, área de investigación, negocio enfocado, etc.
- Se definirá su clave primaria, que se compone de la combinación de las claves primarias de cada dimensión que se utilizará para generar las consultas.
- Se renombrarán las tablas de hechos o indicadores si es que no llegasen a ser lo suficientemente explícitos.

Casos que se deben considerar al momento de definir las tablas de hechos:

Caso 1: Si en dos o más preguntas figuran los mismos indicadores pero con diferentes dimensiones de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición.



Figura 40. Tabla de Hechos Caso 1 (Ricardo, 2010).

Caso 2: Si en dos o más preguntas figuran diferentes indicadores con diferentes dimensiones de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición.

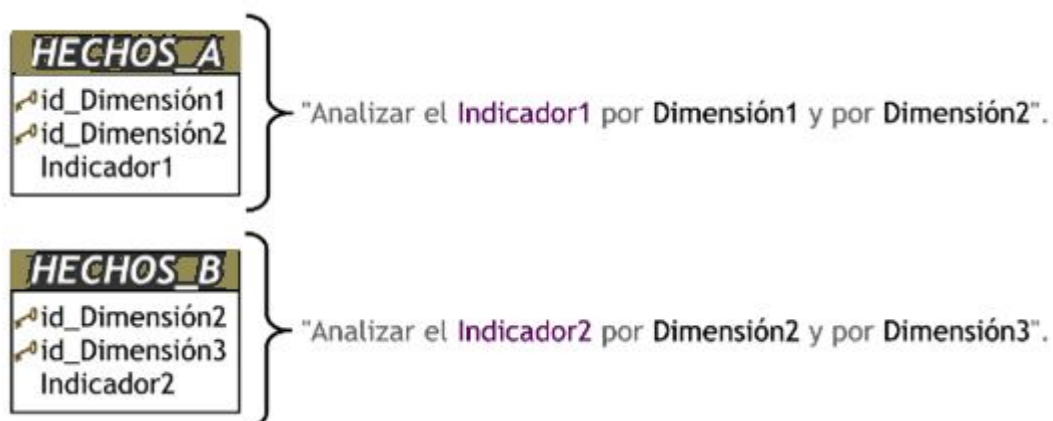


Figura 41. Tabla Hechos Caso 2 (Ricardo, 2010).

Caso 3: Si el conjunto de preguntas cumplen con las condiciones de los dos puntos anteriores se deberán unificar aquellos interrogantes que posean diferentes indicadores pero iguales dimensiones de análisis, para luego reanudar el estudio de las preguntas.

Para los esquemas constelación se tiene como objetivo realizar las siguientes actividades:

- Las tablas de hechos se deben confeccionar teniendo en cuenta el análisis de las preguntas realizadas por el usuario en pasos anteriores y sus respectivos indicadores y dimensiones.
- Cada tabla de hechos debe poseer un nombre que la identifique, contener sus indicadores correspondientes y su clave debe estar formada por la combinación de las claves de las dimensiones que intervendrán.
- Antes de continuar, vale la pena recordar que las perspectivas fueron convertidas en dimensiones en el paso anterior, razón por la cual, las preguntas realizadas por el usuario son examinadas a través de indicadores y dimensiones.

A continuación se presenta la tabla de hechos, del caso de estudio:

La tabla de hechos tiene el nombre "VENTAS".

Su clave principal es la combinación de las claves principales de las tablas de dimensiones antes definidas: "idCliente", "idPublicacion", "idZona" e "idFecha".

Los hechos creados se corresponden con los indicadores y son renombrados:

“Unidades Facturadas” por “CantFac”.

“Unidades de Reposición” por “CantRep”.

“Importe Total de Ventas” por “Importe”.

“Porcentaje de Devoluciones” por “PorcDev”.

VENTAS	
🔑	idCliente
🔑	idPublicacion
🔑	idFecha
🔑	idZona
	CantFac
	CantRep
	Importe
	PorcDev

Figura 42. Diseño Tabla de Hechos caso de estudio.

2.7.2.3.4. Uniones.

Para los tres tipos de esquemas, se realizarán las uniones correspondientes entre sus tablas de dimensiones y sus tablas de hechos; estas relaciones son iguales a las que se realizan en los SGBD. Como se muestra en el caso de estudio:

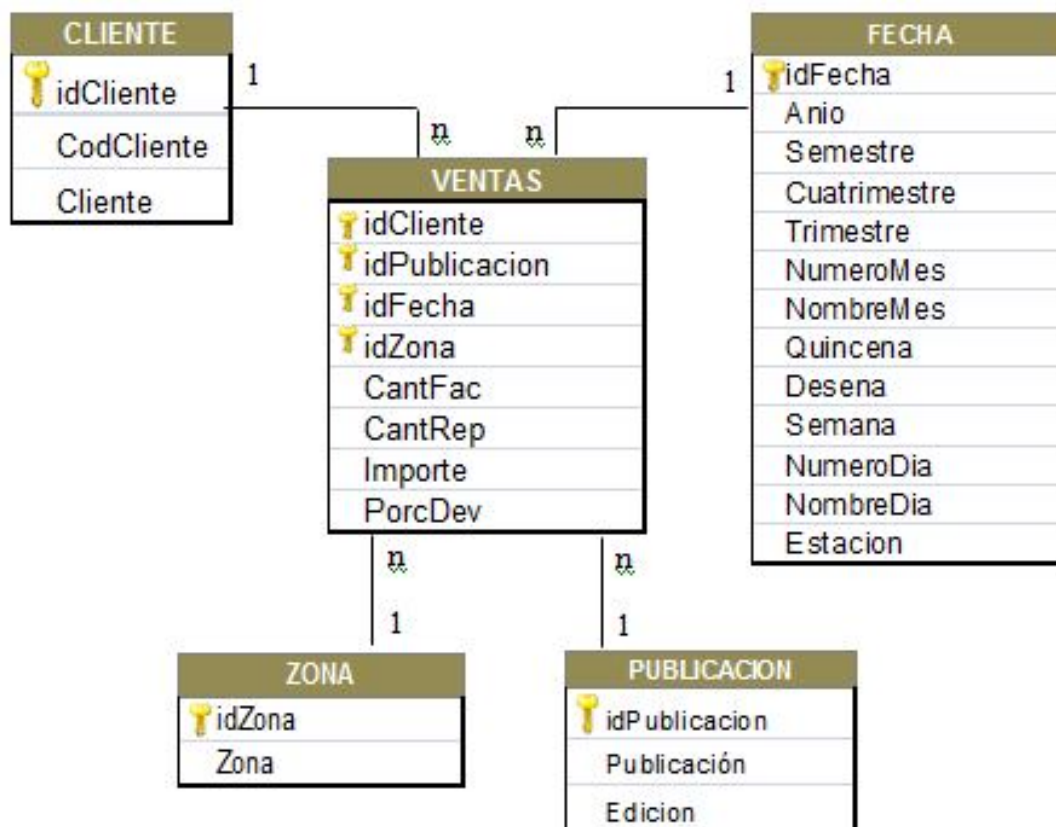


Figura 43: Uniones tabla de hechos Ventas.

2.7.2.4. Paso 4. Integración de Datos

Por último, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc., se definirán políticas y estrategias para la Carga Inicial del DWH y su respectiva actualización.

En esta parte del proyecto, se debe tener completamente definido el modelo lógico del DWH, para posteriormente definir las operaciones que permitan poblar el DWH; a estas operaciones se las conoce como ETL y su propósito es el de asegurar la integridad de la información en la bodega de datos.

Una vez construido el modelo se debe realizar la Carga Inicial y/o actualización ^[16] de la información; la primera vez será la Carga Inicial¹² y posteriormente solo actualizaciones. Previo a poblar la bodega de datos se debe tener en cuenta las siguientes consideraciones:

- Evitar que el DWH sea cargado con valores faltantes o anómalos.
- Establecer condiciones y restricciones para asegurar que solo se utilicen los datos de interés.
- Cuando se trabaja con un esquema constelación, tener presente que varias tablas de dimensiones serán compartidas con diferentes tablas de hechos, ya que ciertas restricciones se pueden contraponer al aplicar a cierta tabla de hechos con respecto a otra.
- Primero se cargarán los datos de las dimensiones y luego los de las tablas de hechos. En el caso en que se esté utilizando un esquema copo de nieve, cada vez que existan jerarquías de dimensiones, se comenzarán cargando las tablas de dimensiones del nivel más general al más detallado.
- Es importante tener presente, que al cargar los datos en las tablas de hechos pueden utilizarse preagregaciones ¹³, ya sea al nivel de granularidad de la misma o a otros niveles diferentes.

¹² Refiérase al punto 2.4.2.3. de este documento donde se explica la carga inicial.

¹³ Tablas que se usan para almacenar información en forma resumida, con un nivel de agregación mayor al obtenido inicialmente.

Finalmente al concluir todas las etapas de la metodología Hefesto se cuenta con toda la información de análisis y diseño necesaria para empezar la construcción del sistema.

2.8. Herramientas.

Las herramientas en que se basa la construcción del sistema se detallan a continuación.

2.8.1. Microsoft SQL Server

Microsoft SQL Server es un sistema para la gestión de bases de datos producido por Microsoft basado en el modelo relacional. Sus lenguajes para consultas son T-SQL y ANSI SQL¹⁴.

Microsoft SQL Server 2012 está basado en una plataforma *Cloud-Ready*, o sea una plataforma lista para la nube que ayuda a las empresas a construir soluciones basadas en la nube con mucha rapidez, hacia el almacenamiento y servicios en la nube, además con todas las herramientas y seguridad requerida para hacerlo. Otra característica del SQL Server 2012 es que mejora el rendimiento y disponibilidad de sus aplicaciones es el SQL Server *AlwaysOn*; con este componente, se podrá tener el “*Uptime*”¹⁵ y se logra reducir el “*Downtime*”¹⁶ gracias a la función integrada de alta disponibilidad y recuperación de desastres de forma tal que, su aplicación este siempre

¹⁴ Fuente de consulta Wikipedia.

¹⁵ *Uptime*, es una medida en que una máquina, generalmente una computadora, ha trabajado y está disponible.

¹⁶ *Downtime*, se refiere a periodos de tiempo cuando un sistema no está disponible.

disponible y todos sus datos dentro de la misma estén siempre seguros. Gracias a la mejora en rendimiento y refinación en el manejo de recursos, es posible reducir el número de equipos inactivos mientras se ahorra costos de TI.

“SQL Server 2012 Edición Business Intelligence” es una plataforma completa con la cual las organizaciones pueden crear y desplegar soluciones de BI seguras, escalables y manejables.

2.8.2. Microsoft Visual Estudio

Microsoft Visual Studio es un entorno de desarrollo integrado (IDE, por sus siglas en inglés) para sistemas operativos Windows. Soporta varios lenguajes de programación tales como Visual C++, Visual C#, Visual J#, y Visual Basic .NET, al igual que entornos de desarrollo web como ASP.NET. Aunque actualmente se han desarrollado las extensiones necesarias para muchos otros.

Visual Studio permite a los desarrolladores crear aplicaciones, sitios y aplicaciones web, así como servicios web en cualquier entorno que soporte la plataforma .NET (a partir de la versión .NET 2002).

“Visual Studio 2010 Professional”, es la herramienta esencial para las personas que realizan tareas de desarrollo básico, la cual simplifica la compilación, la depuración y el despliegue de las aplicaciones en una variedad de plataformas incluyendo *SharePoint* y la *Nube*. También viene con el soporte integrado para el desarrollo con pruebas y con las

herramientas de depuración que ayudan a garantizar soluciones de alta calidad en la construcción de proyectos tecnológicos.

2.8.3. SQL Server Management Studio

SQL Server Management Studio, es un entorno integrado para obtener acceso, configurar, administrar y desarrollar todos los componentes de SQL Server, el cual combina un amplio grupo de herramientas gráficas con una serie de editores de *script* enriquecidos que permiten a desarrolladores y administradores de todos los niveles obtener acceso SQL Server.

SQL Server Management Studio combina las características del Administrador corporativo, el Analizador de consultas y *Analysis Manager*, herramientas incluidas en versiones anteriores de SQL Server, en un único entorno. Además, SQL Server Management Studio funciona con todos los componentes de SQL Server, como *Reporting Services* e *Integration Services*. De este modo, los desarrolladores pueden disfrutar de una experiencia familiar y los administradores de bases de datos disponen de una herramienta única y completa que combina herramientas gráficas fáciles de usar con funciones avanzadas de *scripting*.¹⁷

Permite la creación de proyectos BI, con todas las fases necesarias desde la construcción del Data Warehouse usando técnicas de conexión avanzadas a las fuentes, construyendo y administrando los ETL, hasta presentar todo el análisis al usuario final.

¹⁷ Fuente obtenida de la página web oficial de Microsoft.

2.8.4. DevExpress Suite

Es el conjunto de herramientas de desarrollo de *software* más completa para desarrolladores .NET., que permite crear aplicaciones para Windows, Web, Móviles y Tablets.

2.8.4.1. DXperience Enterprise

El principal objetivo es la generación de soluciones para el desarrollo de aplicaciones de negocio para Windows Forms, y soluciones interactivas para la Web que están preparadas para las nuevas tecnologías del futuro.¹⁸

2.8.4.2. DXtreme Enterprise

Las herramientas DXtreme permiten a desarrolladores crear soluciones innovadoras para aplicaciones multiplataforma y para dispositivos con Windows 8, iPad y smartphones como el iPhone y Android.

Con las herramientas de DXtreme los desarrolladores crean aplicaciones mediante HTML5, CSS y JavaScript para implementar aplicaciones web interactivas y crear la mejor experiencia para el usuario.¹⁹

2.8.5. PowerDesigner

PowerDesigner es una herramienta de diseño que permite el modelado de diagramas de relaciones, entidades y luego transformarlos en Diagramas físicos, y la administración de metadatos.

¹⁸ Fuente obtenida de la página web <http://www.sas.com.mx/Software/DevExpress>.

¹⁹ Fuente obtenida de la página web <http://www.sas.com.mx/Software/DevExpress>.

Además cuenta con la opción de crear *scripts* para crear sus bases de datos con todas sus tablas para diferentes motores de la bases de datos

PowerDesigner permite:

- Aumentar la productividad. Alinea el negocio y la tecnología de la información (TI) para ayudar a mejorar la productividad global de la organización.
- Soportar entornos abiertos. Proporciona un apoyo abierto para entornos heterogéneos.
- Incluir funciones de personalización. Es altamente personalizable para ayudar a hacer cumplir las normas y garantizar su cumplimiento.
- Diseñar diagramas para la empresa compleja. Facilita la implementación de arquitectura empresarial debido a la captura de forma intuitiva de las intersecciones entre todas las capas de la arquitectura y la perspectiva de la Empresa.

CAPÍTULO 3

ANÁLISIS Y DISEÑO DEL PROYECTO

3.1. Metodología Hefesto.

3.1.1. Etapa I). Análisis de Requerimientos.

En el análisis realizado se ha encontrado dos tipos de requerimientos:

Requerimientos de consulta o reportes generales, que se generan a partir del Data Warehouse y solo serán descritos dentro de la metodología.

Requerimientos de análisis del negocio, en los cuales se aplica la metodología Hefesto.

3.1.1.1. Identificar preguntas

Las necesidades claves obtenidas del análisis del negocio de la empresa Topnotch Business son las siguientes:

- Se desea conocer el monto de cartera vencida recuperada por la empresa en un periodo de tiempo y zona.
- Se desea conocer el monto de cartera vencida recuperada por el ejecutivo en un periodo de tiempo y zona.

- Se desea conocer el monto de cartera vencida no recuperada por el ejecutivo en un periodo de tiempo y zona.
- Se desea conocer los pagos realizados por los clientes en un periodo de tiempo y zona.
- Se desea conocer porcentaje de Clientes contactados en un periodo de tiempo.
- Se desea conocer porcentaje de Clientes contactados por ejecutivo en un periodo de tiempo.
- Se desea conocer el listado de Clientes pendientes de Notificación.
- Se desea conocer el listado de Direcciones y números de contacto del Cliente
- Se desea conocer la eficiencia del ejecutivo en recuperar la cartera.
- Se desea conocer la eficiencia del ejecutivo en contactar los clientes.
- Se desea conocer la eficiencia de la empresa en recuperar la cartera.
- Se desea conocer eficiencia de la empresa en contactar los clientes.

A continuación se presenta una matriz separando indicadores de consultas directas a la base de datos:

Tabla 6. Tabla de diferenciación entre indicadores y reportes.

No.	CONSULTAS	INDICADORES
1	Se desea conocer porcentaje de Clientes contactados en un periodo de tiempo.	Se desea conocer el monto de cartera vencida recuperada por la empresa en un periodo de tiempo y por zona.
2	Se desea conocer porcentaje de Clientes contactados por ejecutivo en un periodo de tiempo.	Se desea conocer el monto de cartera vencida recuperada por el ejecutivo en un periodo de tiempo y por zona.
3	Se desea conocer el listado de Clientes pendientes de Notificación.	Se desea conocer el monto de cartera vencida no recuperada por el ejecutivo en un periodo de tiempo y zona.
4	Se desea conocer el listado de Direcciones y números de contacto del Cliente	Se desea conocer los pagos realizados por los clientes en un periodo de tiempo y zona.
5		Se desea conocer la eficiencia del ejecutivo en recuperar la cartera.
6		Se desea conocer la eficiencia del ejecutivo en contactar los clientes.
7		Se desea conocer la eficiencia de la empresa en recuperar la cartera.
8		Se desea conocer eficiencia de la empresa en contactar los clientes.

Las necesidades de información están acorde a los objetivos y estrategias de la empresa, ya que con esta información se proveerá un ámbito para la toma de decisiones, que para Topnotch Business será de gran ayuda ya que le permitirá analizar el desempeño de los ejecutivos de cobranza, para así lograr obtener una ventaja competitiva.

3.1.1.2. Identificar indicadores y perspectivas

La tabla 7., representa el análisis realizado para obtener los indicadores y perspectivas en base de las preguntas realizadas en el punto 3.1.1.

Tabla 7. Tabla de identificadores y perspectivas a ser analizadas.

No.	PREGUNTAS	INDICADORES	PERSPECTIVAS
1	Monto de cartera vencida recuperada por la empresa en un periodo de tiempo y zona.	Monto de cartera vencida recuperada	Clientes
2	Monto de cartera vencida recuperada por el ejecutivo en un periodo de tiempo y zona		Tiempo
3	Monto de cartera vencida no recuperada por ejecutivo en un periodo de tiempo y zona.	Monto de cartera vencida no recuperada	Ejecutivos
4	Pagos realizados por los clientes en un periodo de tiempo y zona.	Pagos efectuados	Zona
5	Eficiencia del ejecutivo en recuperar la cartera.	Eficiencia del ejecutivo en recuperar la cartera.	
6	Eficiencia del ejecutivo en contactar los clientes.	Eficiencia de la empresa en contactar los clientes.	
7	Eficiencia de la empresa en recuperar la cartera.	Eficiencia de la empresa en recuperar la cartera.	
8	Eficiencia de la empresa en contactar los clientes.	Eficiencia del ejecutivo en contactar los clientes.	

3.1.1.3. Modelo Conceptual

En esta etapa, se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior.

Modelo Conceptual: descripción de alto nivel de la estructura de la base de datos

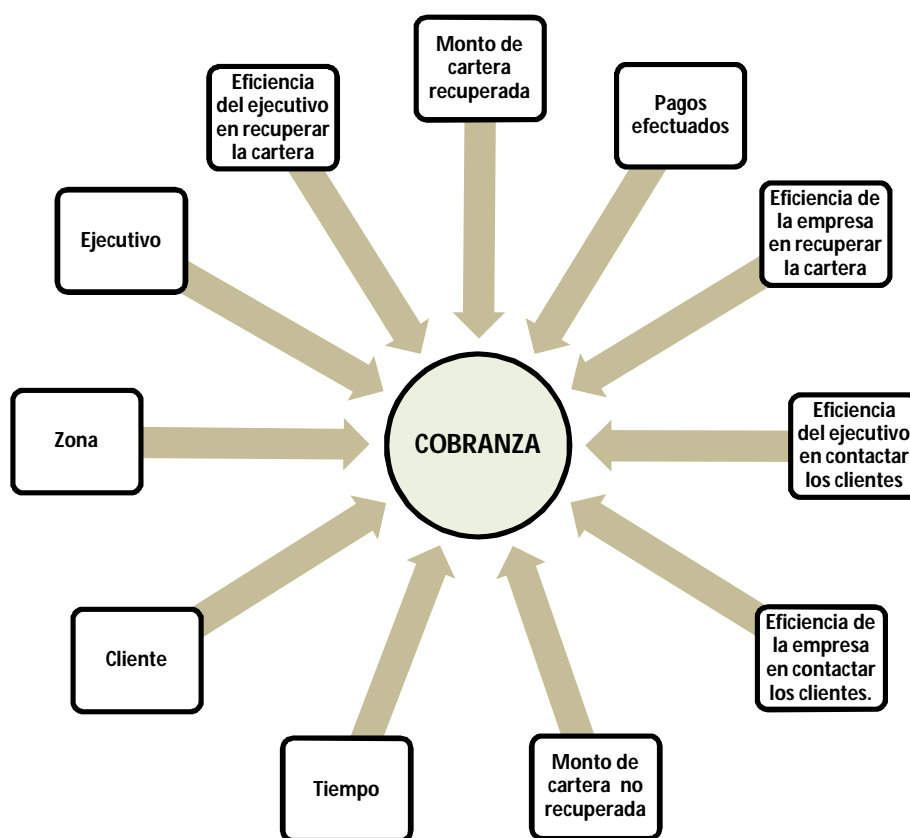


Figura 44. Modelo conceptual alto nivel.

3.1.2. Etapa II). Análisis de OLTP.

Se analiza los OLTP para determinar, cómo se construirán los indicadores, señalar las correspondencias con los datos fuentes y para seleccionar los campos de estudio de cada perspectiva.

3.1.2.1. Construcción de los Indicadores.

3.1.2.1.1. Monto de cartera vencida recuperada.

Este indicador se representa mediante la función de agregación suma; para el caso la sumatoria de cartera recuperada en un periodo de tiempo,

zona y por cada ejecutivo. Para saber el monto recuperado por la empresa se analiza quitando la perspectiva ejecutivo.

Tabla de Hechos: Cartera recuperada

Función de sumarización: SUM

3.1.2.1.2. Monto de cartera vencida no recuperada.

Este indicador representa la sumatoria de cartera no recuperada en un periodo de tiempo, zona y por todos los ejecutivos. Para saber el monto recuperado por la empresa se analiza quitando la perspectiva ejecutivo.

Tabla de Hechos: Cartera no recuperada

Función de sumarización: SUM

3.1.2.1.3. Pagos efectuados.

Este indicador representa la sumatoria de efectuados por cliente en un periodo de tiempo, zona y por ejecutivos.

Tabla de Hechos: Pago

Función de sumarización: SUM

3.1.2.1.4. Eficiencia del ejecutivo en recuperar la cartera.

Este indicador representa la eficiencia del ejecutivo en recuperar cartera vencida en un periodo de tiempo y zona.

Tabla de Hechos: Eficiencia del ejecutivo en recuperar la cartera.

Función:

$$Desempeño = \left[\frac{MontoRecuperado - 100\%}{MontoARecuperar - 100\%} \right] * 100\%$$

Fórmula 3. Cálculo del desempeño de un empleado, para la empresa Top Notch Business.

3.1.2.1.5. Eficiencia del ejecutivo en contactar los clientes.

Este indicador representa la eficiencia del ejecutivo en contactar clientes en un periodo de tiempo y zona.

Tabla de Hechos: Eficiencia del ejecutivo en contactar los clientes

Función:

$$Desempeño = \left[\frac{ClientesContactados - 100\%}{ClientesAContactar - 100\%} \right] * 100\%$$

Fórmula 4. Cálculo del indicador Eficiencia del ejecutivo en contactar a los clientes, para la empresa Top Notch Business.

3.1.2.1.6. Eficiencia de la empresa en recuperar la cartera.

Este indicador representa la eficiencia de la empresa en recuperar cartera vencida en un periodo de tiempo, zona y por todos los ejecutivos.

Tabla de Hechos: Eficiencia de la empresa en recuperar la cartera

Función:

$$Desempeño = \left[\frac{MontoTotalRecuperado - 100\%}{MontoTotalARecuperar - 100\%} \right] * 100\%$$

Fórmula 5. Cálculo del indicador Eficiencia en recuperar la cartera a nivel empresarial, para la empresa Top Notch Business.

3.1.2.1.7. Eficiencia de la empresa en contactar a los clientes.

Tabla de Hechos: Eficiencia de la empresa en contactar los clientes

Función:

$$Desempeño = \left[\frac{TotalClientesContactados - 100\%}{TotalClientesAContactar - 100\%} \right] * 100\%$$

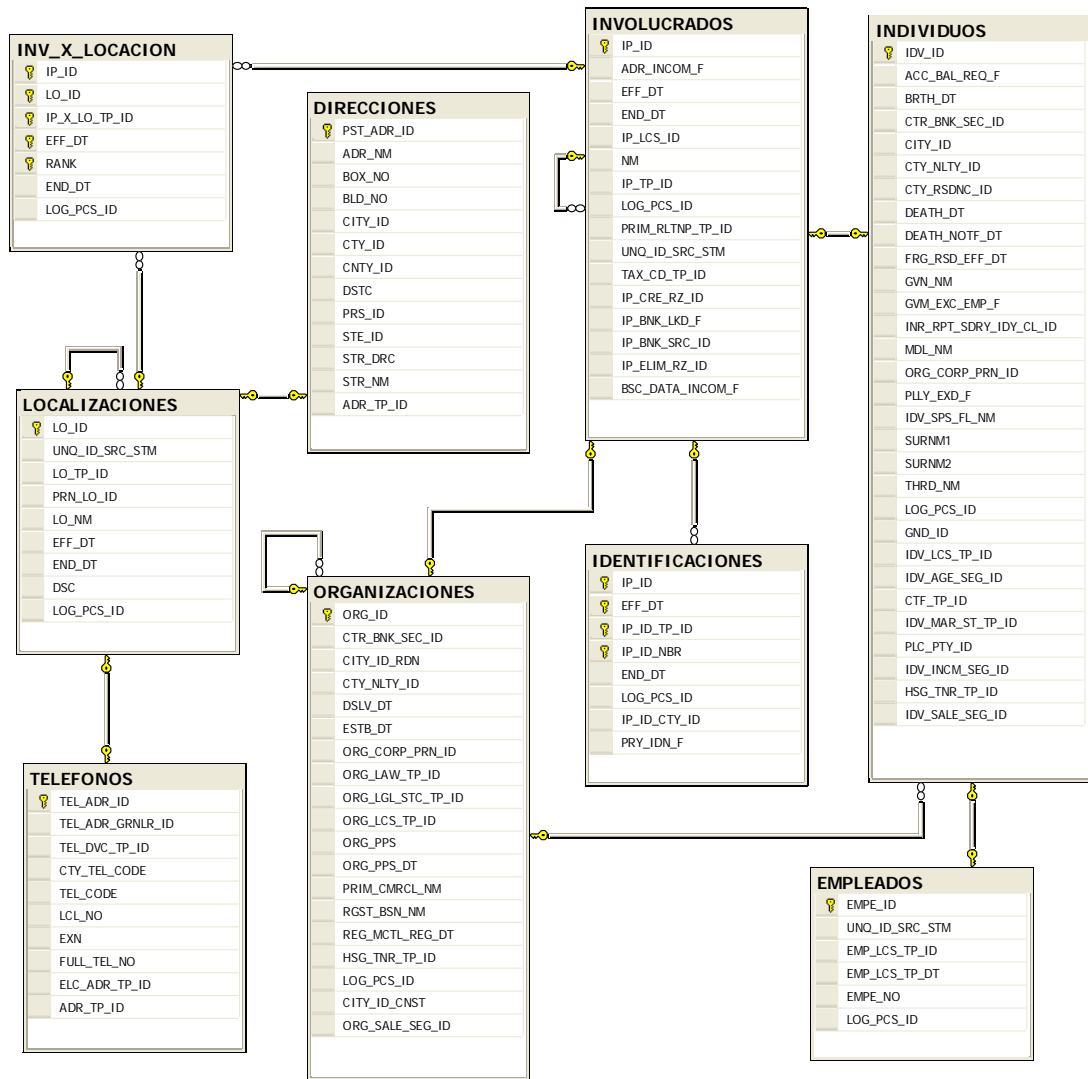
Fórmula 6. Cálculo del indicador Eficiencia en contactar a los clientes, para la empresa Top Notch Business.

Este indicador representa la eficiencia la empresa en contactar clientes en un periodo de tiempo, zona y por todos los ejecutivos.

3.1.2.2. Establecer Correspondencia.

Las correspondencias OLTP para el proyecto se derivan de varias fuentes, por lo cual se analizará cada caso en particular para definir la respectiva correspondencia.

Fuente "BD COMUN TOP".- Base de datos en la cual se encuentran clientes, ejecutivos, direcciones postales y direcciones telefónicas. Los clientes y ejecutivos se encuentran generalizados en un concepto llamado participantes (tabla IP) y especializado en los conceptos individuos (tabla IDV) y organizaciones (tabla ORG); para las direcciones se generaliza en el concepto de localizaciones (tabla LO) y se especializa en direcciones postales (tabla PST_ADR) y direcciones telefónicas (tabla TEL_ADR). En la Figura 45., se describe el diagrama entidad relación de esta fuente.



**Figura 45. Diagrama entidad relación de la base de datos
BD_COMUN_TOP.**

Posterior a identificar la fuente, se procede a realizar la correspondencia entre el modelo conceptual y el diagrama entidad relación.

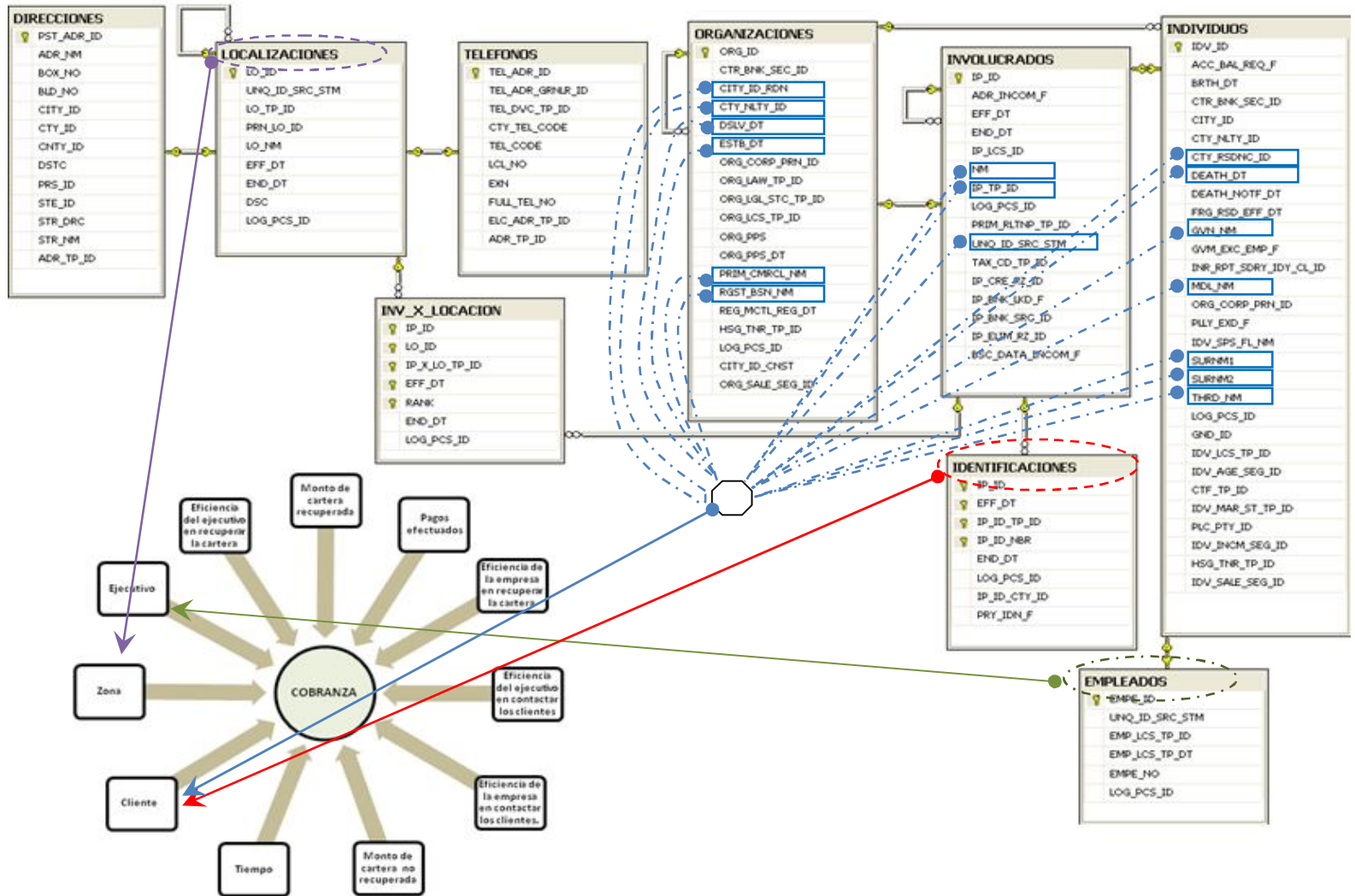


Figura 46. Correspondencia entre el Diagrama ER de la base de datos BD_COMUN_TOP y el modelo conceptual.

Fuente “CARTERA TOP NOTCH”.- Base de datos que proviene de un archivo plano, en la cual se encuentran los montos que se deben recuperar, asociados al cliente, zona, dirección, forma de pago, fecha, días de mora y persona de contacto. En la Figura 47., se describen las columnas que tiene la fuente.

COLUMNAS	
MTX	CIUDAD_COMP
COD_CLIENTE	CEDULA
F_ACT_CUENTA	RUC
STATUS_CUENTA	VIP_CODE
FORMA_PAGO	CICLO
TARJETA	DIAS_ACTUAL
CUENTA_TARJETA	DIAS_0
BANCO	DIAS_30
CUENTA_BANCO	DIAS_60
TIPO_CLIENTE	DIAS_90
BILLGROUP	DIAS_120
COMPANIA	DIAS_150
NOMBRES	DIAS_180
APELLIDOS	DIAS_210
DIRECCION1	DIAS_240
DIRECCION2	DIAS_270
DIRECCION3	DIAS_300
CIUDAD	DIAS_330
TEL_CASA	DIAS_360
TEL_OFI	DIAS_390
FAX	DIAS_TOTAL
CONTACTO1	VENCIMIENTO
TEL_CONTACTO1	PAGOS
CONTACTO2	AJUSTES
TEL_CONTACTO2	DIFERENCIA
DIR_COMP1	ASIGNACION
DIR_COMP2	FECHA ASIGNACION
DIR_COMP3	STATUS

Figura 47. Columnas de la fuente CARTERA TOP NOTCH.

Correspondencia entre el modelo conceptual y las columnas de la fuente proveniente de los archivos planos.

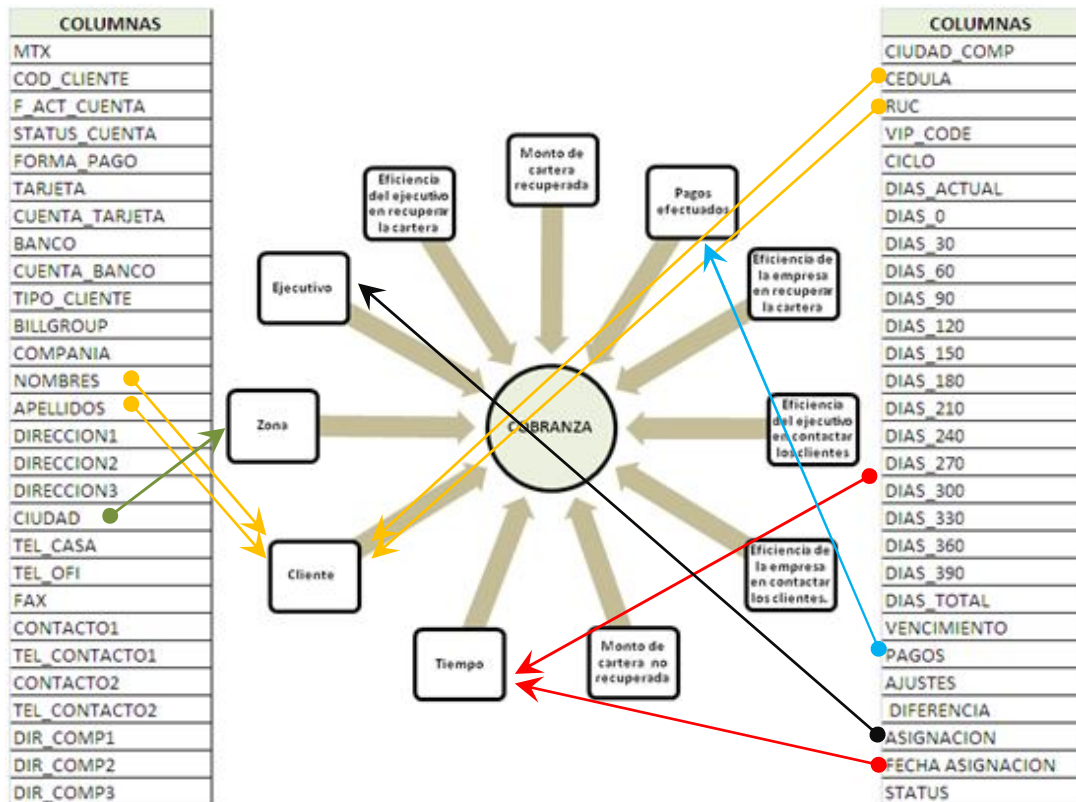


Figura 48. Correspondencia entre las columnas de la fuente CARTERA TOP NOTCH y el modelo conceptual.

Fuente "REPORTE EJECUTIVO".- Base de datos que proviene de un archivo plano, en la cual se encuentran los montos recuperados y el resumen de la gestión realizada por el ejecutivo. En la Figura 49., se describen las columnas que tiene la fuente.

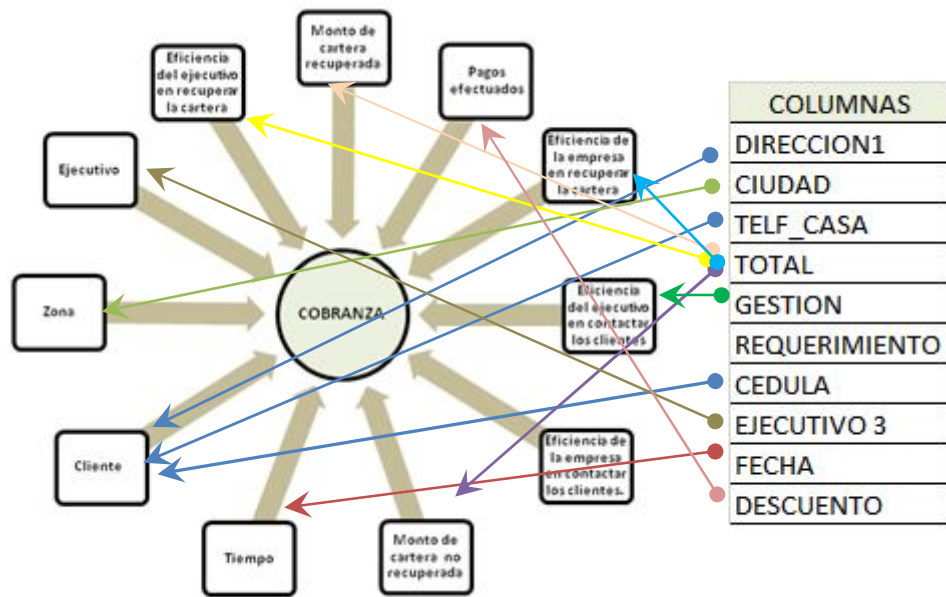


Figura 49. Correspondencia entre las columnas de la fuente REPORTE DE EJECUTIVO y el modelo conceptual.

Fuente “OTROS MONTOS RECUPERADOS”.- Archivo plano en el cual se encuentran montos recuperados por otras empresas de cobranzas asociadas a la empresa TOPNOTCH BUSINESS. En la Figura 50., se describen las columnas que tiene la fuente.

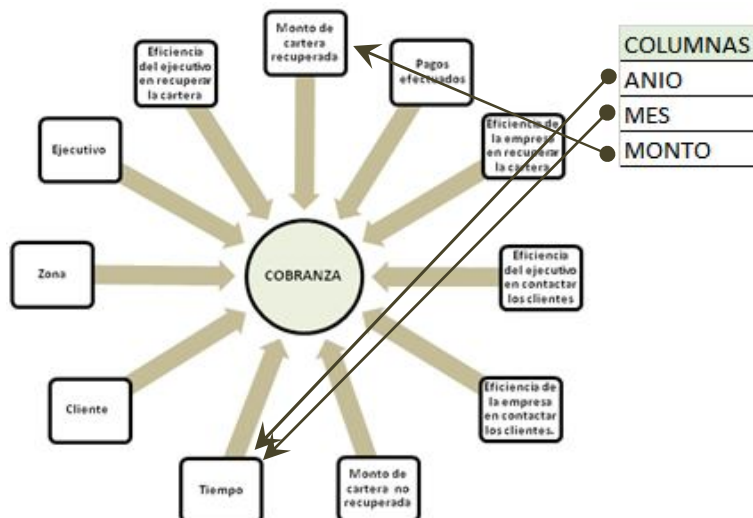


Figura 50. Correspondencia entre las columnas de la fuente “OTROS MONTOS RECUPERADOS” y el modelo conceptual.

Las relaciones identificadas fueron las siguientes:

- La tabla “LOCALIZACIONES” está relacionada con la perspectiva “Zona”.
- La tabla “IDENTIFICACIONES” está relacionada con la perspectiva “Clientes”.
- La tabla “EMPLEADOS” está relacionada con la perspectiva “Ejecutivos”.
- El campo “CITY_ID_RDN de la tabla ORGANIZACIONES” se relaciona con la perspectiva “CLIENTE”.
- El campo “CTY_NLTY_ID de la tabla ORGANIZACIONES” se relaciona con la perspectiva “CLIENTE”.
- El campo “DSL_V_DT de la tabla ORGANIZACIONES” se relaciona con la perspectiva “CLIENTE”.
- El campo “ESTB_DT de la tabla ORGANIZACIONES” se relaciona con la perspectiva “CLIENTE”.
- El campo “PRIM_CMRCL_NL de la tabla ORGANIZACIONES” se relaciona con la perspectiva “CLIENTE”.
- El campo “RGST_BSN_NM de la tabla ORGANIZACIONES” se relaciona con la perspectiva “CLIENTE”.
- El campo “NM de la tabla INVOLUCRADOS” se relaciona con la perspectiva “CLIENTE”.

- El campo "IP_TP_ID de la tabla INVOLUCRADOS" se relaciona con la perspectiva "CLIENTE".
- El campo "UNQ_ID_SRC_STM de la tabla INVOLUCRADOS" se relaciona con la perspectiva "CLIENTE".
- El campo "CTY_RSDNC_ID de la tabla INDIVIDUOS" se relaciona con la perspectiva "CLIENTE".
- El campo "DEATH_DT de la tabla INDIVIDUOS" se relaciona con la perspectiva "CLIENTE".
- El campo "GVN_NM de la tabla INDIVIDUOS" se relaciona con la perspectiva "CLIENTE".
- El campo "MDL_NM de la tabla INDIVIDUOS" se relaciona con la perspectiva "CLIENTE".
- El campo "SURNM1 de la tabla INDIVIDUOS" se relaciona con la perspectiva "CLIENTE".
- El campo "SURNM2 de la tabla INDIVIDUOS" se relaciona con la perspectiva "CLIENTE".
- El campo "THRD_NM de la tabla INDIVIDUOS" se relaciona con la perspectiva "CLIENTE".
- El campo "NOMBRES de la tabla CARTERA" se relaciona con la perspectiva "CLIENTE".

- El campo "APELLIDOS de la tabla CARTERA" se relaciona con la perspectiva "CLIENTE".
- El campo "CÉDULA de la tabla CARTERA" se relaciona con la perspectiva "CLIENTE".
- El campo "RUC de la tabla CARTERA" se relaciona con la perspectiva "CLIENTE".
- El campo "CIUDAD de la tabla CARTERA" se relaciona con la perspectiva "Zona".
- El campo "DIAS_TOTAL de la tabla CARTERA" se relaciona con la perspectiva "TIEMPO".
- El campo "FECHA ASIGNACION de la tabla CARTERA" se relaciona con la perspectiva "TIEMPO".
- El campo "ASIGNACIÓN de la tabla CARTERA" se relaciona con la perspectiva "EJECUTIVO".
- El campo "DIRECCIÓN1 de la tabla GESTION" se relaciona con la perspectiva "CLIENTE".
- El campo "TELF_CASA de la tabla GESTION" se relaciona con la perspectiva "CLIENTE".
- El campo "CÉDULA de la tabla GESTION" se relaciona con la perspectiva "CLIENTE".

- El campo “DESCUENTO de la tabla GESTION” se relaciona con la perspectiva “PAGOS EFECTUADOS”.
- El campo “FECHA de la tabla GESTION” se relaciona con la perspectiva “TIEMPO”.
- El campo “EJECUTIVO3 de la tabla GESTION” se relaciona con la perspectiva “EJECUTIVO”.
- El campo “TOTAL de la tabla GESTION” se relaciona con la perspectiva “EFICIENCIA DE LA EMPRESA EN RECUPERAR CARTERA”.
- El campo “TOTAL de la tabla GESTION” se relaciona con la perspectiva “MONTO DE CARTERA RECUPERADA”.
- El campo “TOTAL de la tabla GESTION” se relaciona con la perspectiva “EFICIENCIA DEL EJECUTIVO EN RECUPERAR CARTERA”.
- El campo “TOTAL de la tabla GESTION” se relaciona con la perspectiva “MONTO DE CARTERA NO RECUPERADA”.
- El campo “GESTION de la tabla GESTION” se relaciona con la perspectiva “EFICIENCIA DEL EJECUTIVO EN CONTACTAR LOS CLIENTES”.
- El campo “CIUDAD de la tabla GESTION” se relaciona con la perspectiva “Zona”.

- El campo “ANIO de la tabla OTROS MONTOS RECUPERADOS” se relaciona con la perspectiva “TIEMPO”.
- El campo “MES de la tabla OTROS MONTOS RECUPERADOS” se relaciona con la perspectiva “TIEMPO”.
- El campo “MONTO de la tabla OTROS MONTOS RECUPERADOS” se relaciona con el indicador “Monto de Cartera Recuperada”.

3.1.2.3. Nivel de granularidad

Una vez establecidas las correspondencias, se analizaron los campos de cada tabla y por cada fuente de información a la que se hace referencia. El proceso que se utilizó fue examinar la base de datos para intuir los significados de cada campo, y luego se consultó con el encargado del sistema el significado de cada uno de los campos, investigación que evita cualquier tipo de mal interpretación.

- Para la perspectiva “Cliente”, los datos disponibles son los siguientes:

FUENTE: BD_COMUN_TOP

TABLA: IDENTIFICACIONES

- IP_ID: identificador de la tabla.
- EFF_DT: Fecha de vigencia del registro.
- IP_ID_TP_ID: Tipo de identificación del cliente puede ser ruc, pasaporte o cédula.

- IP_ID_NBR: Número de identificación.
- END_DT: Fecha de finalización del registro.
- IP_ID_CTY_ID: País de donde proviene la identificación.
- PRY_IDN_F: Muestra si la identificación es primaria o no; un cliente puede tener cédula y pasaporte.

TABLA: ORGANIZACIONES

- CITY_ID_RDN: Ciudad donde se encuentra ubicada la organización.
- CTY_NLTY_ID: País de nacionalidad donde se encuentra la empresa.
- DSLV_DT: Fecha de disolución de la empresa.
- ESTB_DT: Fecha de constitución de la empresa.
- PRIM_CMRCL_NL: Nombre comercial de la empresa.
- RGST_BSN_NM: Nombre de registro de la empresa.

TABLA: INVOLUCRADOS

- NM: Nombre completo de la empresa o persona natural.
- IP_TP_ID: Identifica si es una persona jurídica o natural.
- UNQ_ID_SRC_STM: Es código de original del sistema fuente.

TABLA: INDIVIDUOS

- CTY_RSDNC_ID: Es el país de residencia de las personas naturales.
- DEATH_DT: Fecha de fallecimiento de la persona natural.
- GVN_NM: Primer nombre de la persona natural.
- MDL_NM: Segundo nombre de la persona natural.
- SURNM1: Primer apellido de la persona natural.
- SURNM2: Segundo apellido de la persona natural.
- THRD_NM: Tercer nombre de la persona natural.

FUENTE: CARTERA TOP NOTCH

- NOMBRES: Nombres de los clientes.
- APELLIDOS: Apellidos de los clientes.
- CEDULA: Identificación del cliente.
- RUC: Identificación de personas jurídicas.

FUENTE: REPORTE DE EJECUTIVO

- DIRECCION1: Dirección de residencia o donde labora el cliente.
- TELF_CASA: Teléfono del domicilio del cliente.
- CEDULA: Es la identificación del cliente.

- Con respecto a la perspectiva “Zona”, los datos disponibles son los siguientes:

FUENTE: BD_COMUN_TOP

TABLA: LOCALIZACIONES

- LO_ID: Identificación única de la tabla.
- UNQ_ID_SRC_STM: Código original del sistema fuente.
- EFF_DT: Fecha de vigencia del registro
- LO_TP_ID: Tipo de dirección, puede ser postal, electrónica.
- LO_NM: Nombre de la localización, si es un domicilio teléfono, fax o correo electrónico.
- END_DT: Fecha de finalización de vigencia del registro
- DSC: Detalle de la localización.

FUENTE: CARTERA TOP NOTCH

- CIUDAD: Ciudad donde reside el cliente o empresa.

FUENTE: REPORTE DE EJECUTIVO

- CIUDAD: Ciudad donde reside el cliente o empresa.

- Con respecto a la perspectiva “Ejecutivo”, los datos disponibles son los siguientes:

FUENTE: BD_COMUN_TOP

FUENTE: EMPLEADOS

- EMPE_ID: Código identificador de la tabla.
- UNQ_ID_SRC_STM: Código original del sistema fuente.
- EMP_LCS_TP_IT: Identificador del tipo de empleado.
- EMP_LCS_TP_DT-Fecha de ingreso del empleado.
- EMPE_NO: Número de empleado.

FUENTE: CARTERA TOP NOTCH

- ASIGNACIÓN: Ejecutivo al que está asignado al cliente para su debida gestión.

FUENTE: REPORTE DE EJECUTIVO

- EJECUTIVO3: Persona que realiza la gestión de cobranza.
- Con respecto a la perspectiva "TIEMPO", los datos disponibles son los siguientes:
 - Granularidad para esta perspectiva es:
 - Año.
 - Semestre.
 - Mes.
 - Semana.

- Descripción del campo asociado a la fuente “CARTERA TOP NOTCH” y a la perspectiva.
 - FECHA ASIGNACIÓN: Fecha en que se asignó al ejecutivo la cuenta.

- Descripción del campo asociado a la fuente “REPORTE DE EJECUTIVO” y a la perspectiva.
 - FECHA: Fecha en que se realizó la gestión de cobranza.

Cuando toda la información se recolectó, el siguiente paso fue consultar a los usuarios cuales eran los datos que se consideran de interés para analizar los indicadores ya expuestos. Los resultados obtenidos fueron los siguientes:

- Perspectiva “Cliente”.
 - NOMBRES
 - APELLIDOS
 - COMPANIA

- Perspectiva “Ejecutivo”.
 - ASIGNACION

- Perspectiva “Tiempo”.
 - “Semana”. Referido al número de semana del mes.
 - “Mes”. Referido número de mes.
 - “Trimestre”. Referido al número de trimestre.
 - “Semestre”. Referido al nombre del mes.
 - “Año”. Referido al número de año.

- Perspectiva “Zona”.
 - “CIUDAD” de la fuente “CARTERA TOP NOTCH”. Ya que esta hace referencia a la ciudad donde se encuentra la persona a realizar la cobranza.

3.1.2.4. Modelo Conceptual ampliado.

El modelo conceptual ampliado resultado del análisis del punto 3.1.2.3, se representa en la Figura 51.

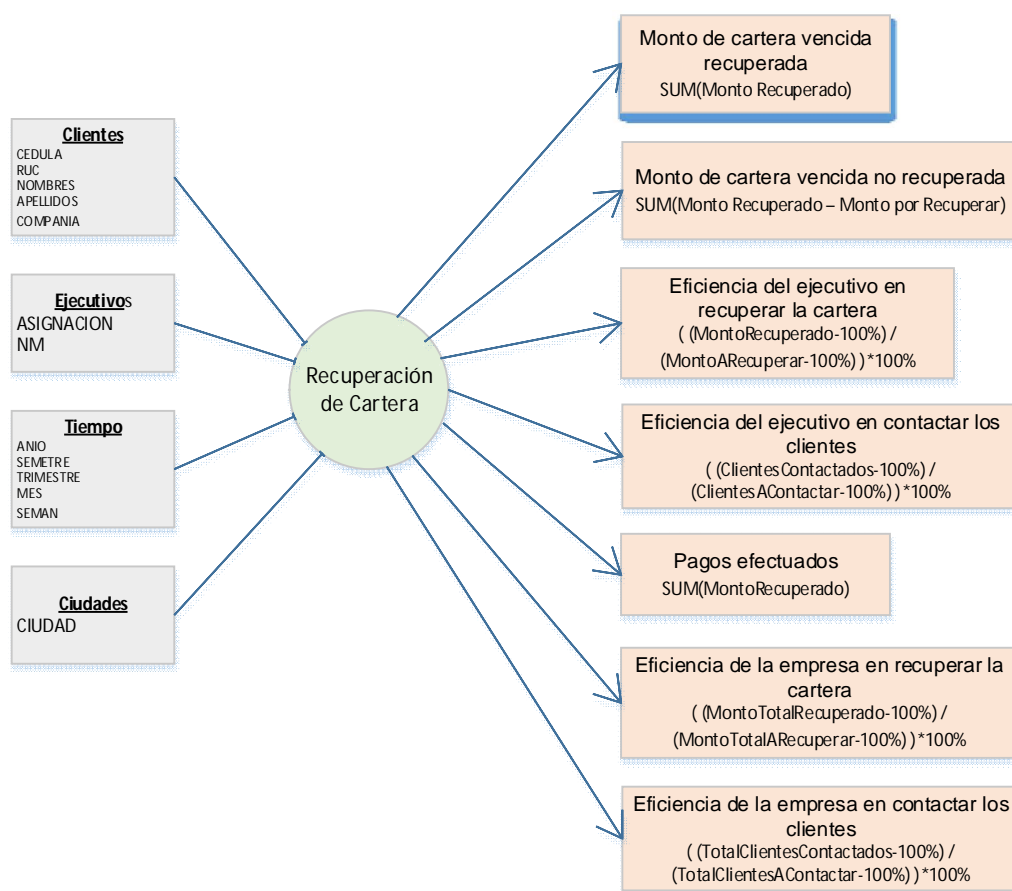


Figura 51. Modelo Conceptual ampliado.

3.1.3. Etapa III). Modelo lógico del Data Warehouse.

3.1.3.1. Tipo de modelo lógico del Data Warehouse.

El esquema que se usará para el modelamiento del Data Warehouse será de tipo estrella.

3.1.3.2. Tablas de dimensión.

Las perspectivas analizadas para el diseño de las tablas de dimensión son: Cliente, Ejecutivo, Tiempo y Zona.

Perspectiva Cliente:

- Nombre de la tabla de dimensión: CLIENTES.
- Clave principal autogenerada: IDENTIFICADOR_CLIENTE.
- Nombre de los campos de la tabla:
 - NUMERO_IDENTIFICACION, proviene de las columnas CEDULA y RUC.
 - TIPO_IDENTIFICACION, campo calculado en base a si existe información en las columnas CEDULA y RUC.
 - NOMBRE_CLIENTE, proviene de las columnas NOMBRES, APELLIDOS Y COMPANIA.
- Gráficamente:

CLIENTES			
<u>IDENTIFICADOR_CLIENTE</u>	<pi>	<u>DMN ID PK</u>	<M>
TIPO_IDENTIFICACION		Variable characters (15)	
NUMERO_IDENTIFICACION		Variable characters (13)	
NOMBRE_CLIENTE		Variable multibyte (200)	
IDR_CLI	<pi>		

Figura 52. Tabla dimensión Clientes.**Perspectiva Ejecutivo:**

- Nombre de la tabla de dimensión: EJECUTIVOS.
- Clave principal autogenerada: IDENTIFICADOR_EJECUTIVO.
- Nombre de los campos de la tabla:

- CODIGO_ORIGINAL_EJECUTIVO, proviene de la columna ASIGNACION.
- NOMBRE_EJECUTIVO, proviene de la columna NM.
- Gráficamente:

EJECUTIVOS	
<u>IDENTIFICADOR EJECUTIVO</u> <pi>	<u>DMN_ID_PK</u> <M>
CODIGO_ORIGINAL_EJECUTIVO	Variable characters (15)
NOMBRE_EJECUTIVO	Variable characters (200)
IDR_EJC <pi>	

Figura 53. Tabla dimensión Ejecutivos.

Perspectiva Tiempo:

- Nombre de la tabla de dimensión: FECHA.
- Clave principal autogenerada: IDENTIFICADOR_FECHA.
- El nombre de los campos se mantiene excepto el de año que cambia por ANIO.
- Gráficamente:

FECHA	
<u>IDENTIFICADOR FECHA</u> <pi>	<u>DMN_ID_PK</u> <M>
ANIO	Integer
SEMESTRE	Integer
TRIMESTRE	Integer
MES	Integer
SEMANA	Integer
IDR_FEC <pi>	

Figura 54. Tabla dimensión Fecha.

Perspectiva Zona:

- Nombre de la tabla de dimensión: CIUDADES.
- Clave principal autogenerada: IDENTIFICADOR_CIUADAD.
- Nombre de los campos de la tabla:
 - NOMBRE_CIUADAD, proviene de la columna CIUADAD.
- Gráficamente:

CIUDADES			
<u>IDENTIFICADOR_CIUADAD</u>	<pi>	<u>DMN_ID_PK</u>	<M>
NOMBRE_CIUADAD		Variable characters (100)	
IDR_CIU	<pi>		

Figura 55. Tabla dimensión Ciudades.**3.1.3.3. Tablas de hechos.**

La tabla de hechos resultado del análisis representa el proceso principal de la empresa que es la recuperación cartera; a continuación se describe su diseño:

- Nombre de la tabla de hechos: RECUPERACION_CARTERA.
- Clave principal autogenerada: IDENTIFICADOR_CIUADAD.
- Nombre de los campos de la tabla:
 - NOMBRE_CIUADAD, proviene de la columna CIUADAD.

- Gráficamente:

RECUPERACION_CARTERA	
MONTO CARTERA VENCIDA_RECUPERADA	DMN_MONTO
MONTO CARTERA VENCIDA NO RECUPERADA	DMN_MONTO
EFICIENCIA_RECUPERACION_CARTERA	DMN_MONTO
EFICIENCIA_CONTACTO_CLIENTES	DMN_MONTO
PAGOS_EFEKTUADOS	DMN_MONTO
IDR_RCP_CTR	<pi>

Figura 56. Tabla de Hechos Recuperación Cartera.

3.1.3.4. Uniones.

En este paso se realizan las uniones de las dimensiones Tiempo, Clientes, Ejecutivos y Ciudades con la tabla de hechos Recuperacion_Cartera. Gráficamente las uniones se describen en la Figura 57.

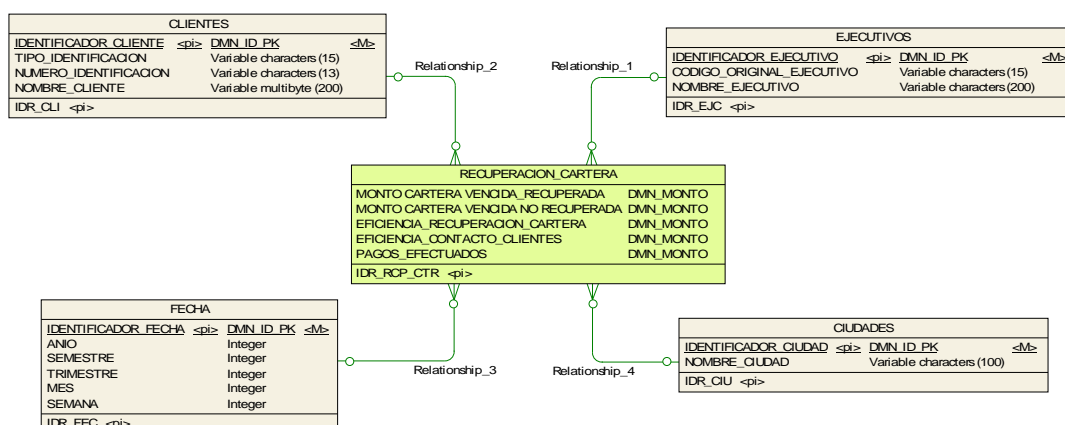


Figura 57. Uniones de la tabla de dimensiones y la tabla de hechos.

3.2. Análisis KPI.

3.2.1. Definición del Objetivo del Estudio.

Mediante consulta a los directivos de la empresa, se determinó que se quiere abarcar estrategias que involucren la parte financiera y estratégica. Por lo cual, se han definido los siguientes objetivos estratégicos:

Tabla 8. Descripción de los KPI y su perspectiva.

No.	PERSPECTIVA	OBJETIVO
1	Financiera	Incrementar la recuperación de cartera de la empresa.
2	Estratégica	Incrementar la eficiencia de la empresa en contactar a los clientes.

ESPACIO EN BLANCO

INTENCIONAL

3.2.2. Matriz de KPI.

Tabla 9. Descripción de los KPI y su perspectiva.

N°	DEFINIR	ACLARAR	CONCEPTUALIZAR	FORMULAR
1	Incrementar la recuperación de cartera de la empresa.	Aumentar la recuperación de cartera en un 2% anualmente	Que la recuperación de la cartera de la empresa, en dólares, del año actual sea mayor que el año anterior	Incremento recuperación de cartera en (USD\$) = $\frac{\sum_{i=E}^f \text{USD\$ Ventas 2013}}{\sum_{i=E}^f \text{USD\$ Ventas 2012}}; i = E, \dots, D \text{ (semanas)}$
2	Incrementar la eficiencia de la empresa en contactar a los clientes.	Aumentar la eficiencia en contactar a los clientes en un 3% semanalmente.	Comparar el número de clientes contactos versus el número total de clientes asignados.	% Devolución por Producto Defectuoso = $\frac{\# \text{Clientes Contactados}}{\text{Total de Clientes A Gestionar}}$

CAPÍTULO 4

DESARROLLO Y CONSTRUCCIÓN

4.1. Aplicación Metodología Hefesto.

4.1.1. Etapa IV). Integración de datos.

Una vez construido el modelo lógico, este tiene que ser probado con los datos que vienen de las distintas fuentes, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc.; posteriormente se definirán las reglas y políticas para su respectiva actualización, como parte del gobierno de la información. Para la carga se ha implementado dos áreas un área temporal llamada Desembarco y el área del DWH. En la Figura 58., se especifica el flujo entre las dos.

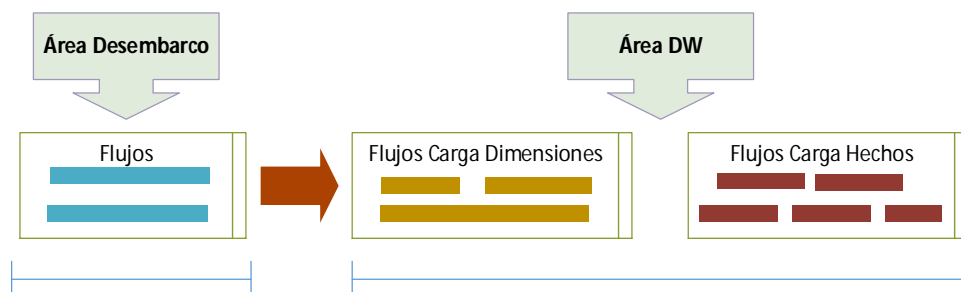


Figura 58. Flujo de las áreas de Desembarco.

4.1.1.1. Carga Inicial.

El esquema de procesos para la construcción de cada ETL que permite la Carga Inicial de todo el ambiente del DWH involucra tres fases principales, que son:

- ETL para la carga de las fuentes externas al área de Desembarco.
- Procesos ETL para la carga de las tablas dimensión.
- Procesos ETL para la carga de la tabla de hechos.

4.1.1.1.1 Proceso 1. Carga fuentes externas al área de Desembarco.

Este proceso se lleva a cabo con el siguiente esquema de ETL:

Tabla 10. Esquema de procesos ETL de carga fuente externa.

No.	Nombre ETL
1	1_CRG_CFG_FEC_PCS.dtsx
2	2_CRG_DCO_CARTERA.dtsx
3	3_CRG_DCO_CLIENES.dtsx
4	4_CRG_DCO_GESTION.dtsx
5	5_CRG_DCO_CIUADAD.dtsx
6	6_CRG_DCO_RECUPERACION_ADICIONAL.dtsx
7	7_CRG_DCO_EJECUTIVOS.dtsx

- ETL Asignación parámetros iniciales (1_CRG_CFG_FEC_PCS).

El proceso contiene la tarea de asignación de la fecha de proceso para todo el esquema que carga las fuentes externas al área de desembarco y estableciendo "PROCESANDO" a la bandera "ESTADO" de la tabla de configuración de Fechas procesadas "BDWH_CFG_PRC". En la Figura 59., se muestra la tarea del ETL.

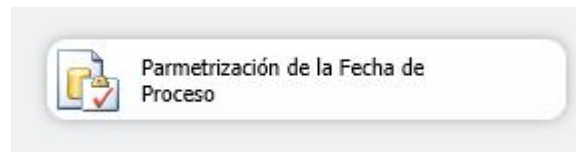


Figura 59. Tarea proceso ETL Asignación parámetros iniciales.

- ETL Desembarco Cartera (2_CRG_DCO_CARTERA).

El proceso contiene las tareas para desembarcar la fuente externa Cartera; el archivo del que se lee tiene el siguiente formato: CARTERA TOP NOTCH AL dd-mm-yyyy.txt. El flujo de tareas que llevan a cabo este proceso se muestra en la Figura 60.

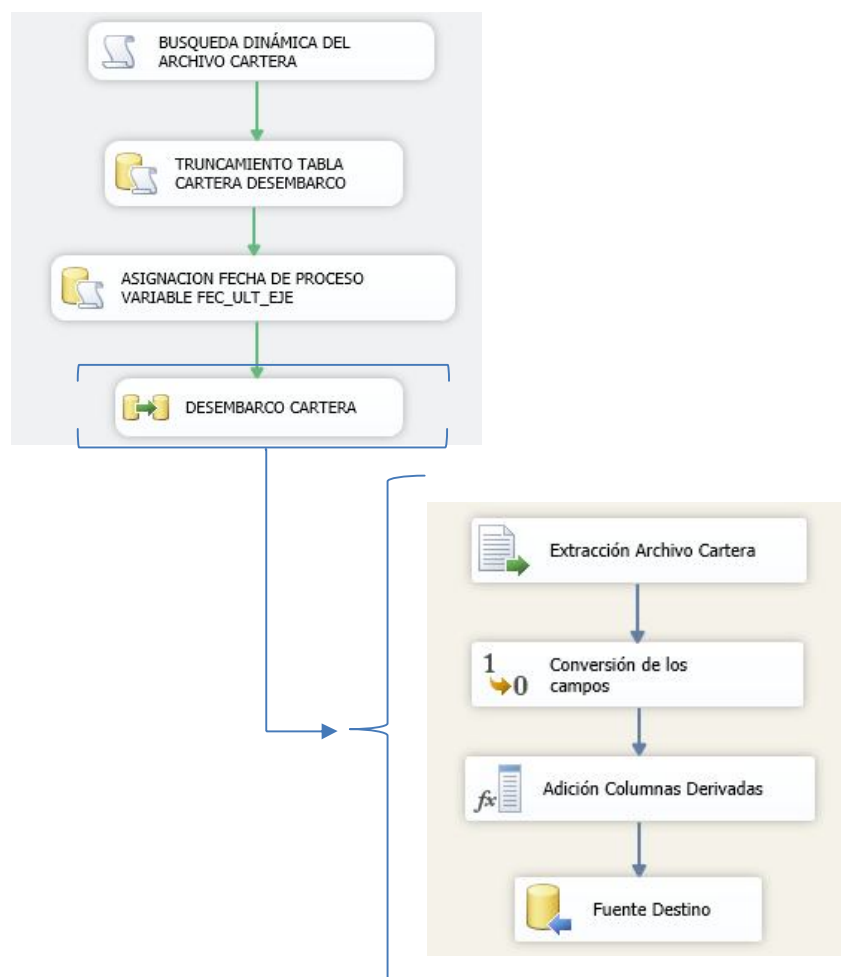


Figura 60. Tareas proceso ETL Desembarco Cartera.

A continuación se describe con un mayor detalle las tareas que involucra este proceso:

- **Búsqueda dinámica del archivo cartera.-** Esta tarea consiste en asignar el valor de la fecha de actual, usando la consulta que se indica en la Figura 61., a la variable general del proceso “FEC_ACT”, para armar la cadena dinámica de conexión al archivo externo.

```
SELECT [BDWH_CFG_IDR]
      ,[BDWH_CFG_FEC_PRC]
      ,[BDWH_CFG_EST]
FROM [TOP_NOTCH_CFG].[dbo].[BDWH_CFG_PRC]
WHERE BDWH_CFG_EST = 'PROCESANDO'
```

Figura 61. Consulta fecha en la que se está procesando.

- **Truncamiento tabla cartera desembarco.-** Tarea que trunca la tabla de cartera “TAB_CTR”.
- **Asignación fecha de proceso variable “FEC_ULT_EJE”.-** Se asigna la fecha actual a la variable “FEC_ULT_EJE”, para derivar la columna Fecha de proceso al flujo de ejecución.
- **Desembarco Cartera.-** Esta tarea contiene las siguientes subtareas:
 - Extracción Archivo Cartera.- Extrae el archivo de Cartera (CARTERA TOP NOTCH AL dd-mm-yyyy.txt), según la cadena de conexión armada para la fecha de proceso.

- Conversión de los campos.- Todos los campos de la fuente son convertidos al tipo y tamaño de la tabla destino “TAB_CTR”.
 - Adición Columnas Derivadas.- En la información convertida se añade el campo “FEC_PCS”, con la fecha de procesamiento.
 - Fuente Destino.- Finalmente se mapean las columnas con la información transformada y las derivadas para ser enviadas a cargar a la tabla “TAB_CTR” del área de desembarco.
- ETL Desembarco Clientes (3_CRG_DCO_CLIENTES).

El proceso contiene las tareas para desembarcar la información de clientes usando la tabla anteriormente cargada de cartera “TAB_CTR” que se encuentra en el área de desembarco. El flujo de tareas que llevan a cabo este proceso se muestra en la Figura 62.

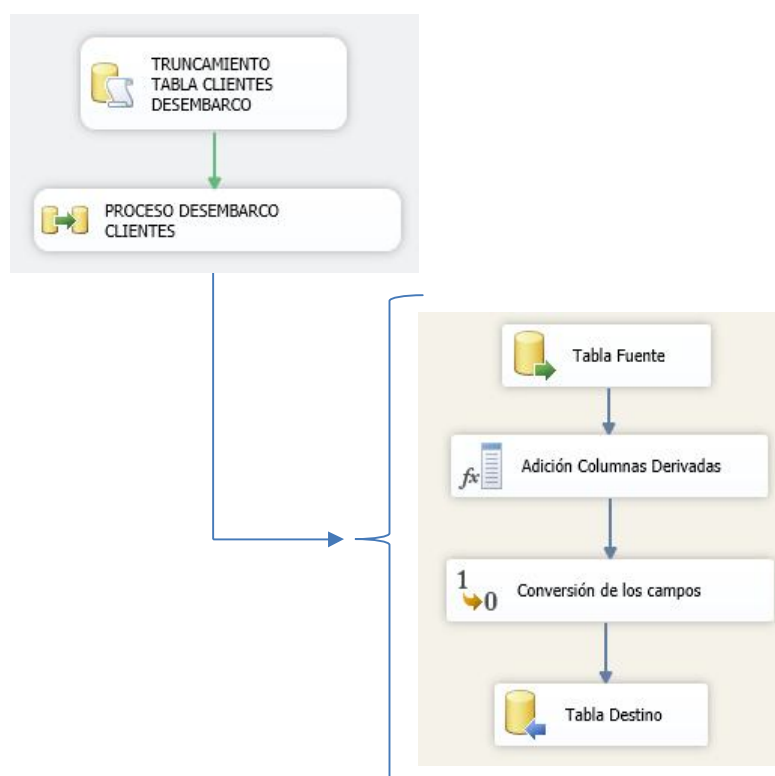


Figura 62. Tareas proceso ETL Desembarco Clientes.

- **Truncamiento tabla clientes.-** Tarea que trunca la tabla de clientes “TAB_CLI”.
- **Desembarco Clientes.-** Esta tarea contiene las siguientes subtareas:
 - Extracción tabla Cartera.- Extrae la información, de la fecha de proceso, de la tabla Cartera “TAB_CTR” para extraer los clientes.
 - Adición Columnas Derivadas.- En la información convertida se reemplaza el campo “NUM_IDE” quitando todos los espacios encontrados.
 - Conversión de los campos.- Todos los campos de la fuente son convertidos al tipo y tamaño de la tabla destino “TAB_CLI”.

- Fuente Destino.- Finalmente se mapean las columnas con la información transformada y las derivadas para ser enviadas a cargar a la tabla “TAB_CLI” del área de desembarco.

- ETL Desembarco Clientes (4_CRG_DCO_GESTION).

El proceso contiene las tareas para desembarcar la información de la gestión de recuperación de cartera realizada por cada empleado; el archivo del que se lee tiene el siguiente formato: REPORTE_GESTION_001_dd-mm-yyyy.txt. El flujo de tareas que llevan a cabo este proceso se muestra en la Figura 63.

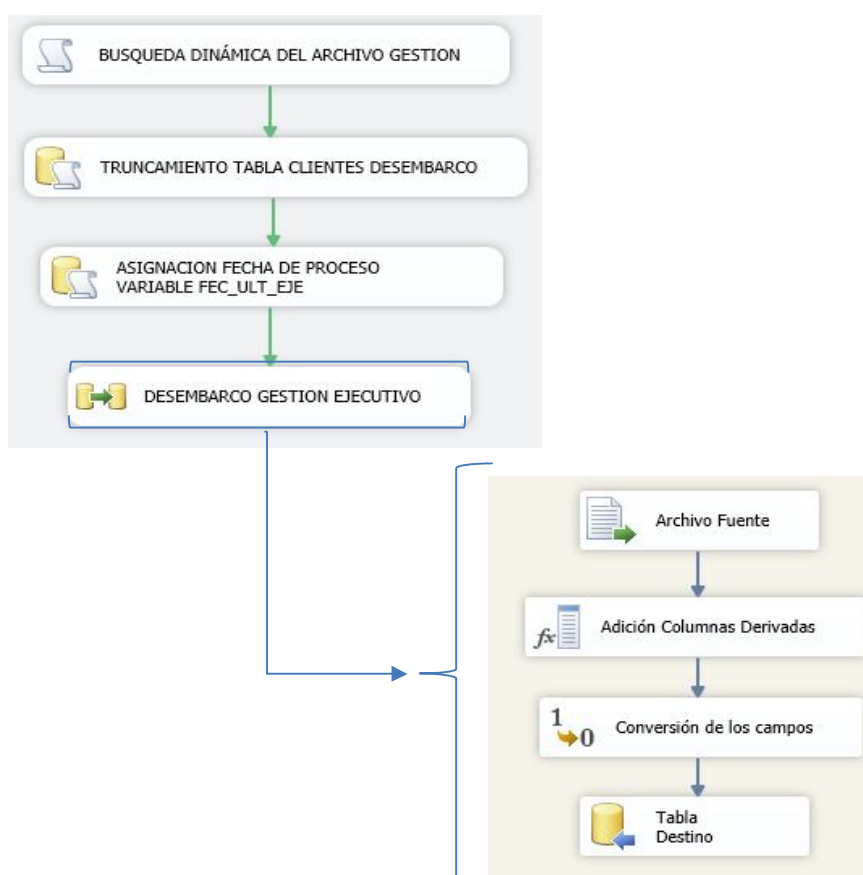


Figura 63. Tareas proceso ETL Desembarco Gestión.

- **Búsqueda dinámica del archivo gestión.-** Esta tarea consiste en asignar el valor de la fecha de actual, usando la consulta que se indica en la Figura 61., a la variable general del proceso “FEC_ACT”, para armar la cadena dinámica de conexión al archivo externo.
- **Truncamiento tabla cartera desembarco.-** Tarea que trunca la tabla de cartera “TAB_GES_EJC”.
- **Asignación fecha de proceso variable “FEC_ULT_EJE”.-** Se asigna la fecha actual a la variable “FEC_ULT_EJE”, para derivar la columna Fecha de proceso al flujo de ejecución.
- **Desembarco Gestión Ejecutivo.-** Esta tarea contiene las siguientes subtareas:
 - Extracción Archivo Gestión.- Extrae el archivo de Gestión (REPORTE_GESTION_001_dd-mm-yyyy.txt), según la cadena de conexión armada para la fecha de proceso.
 - Adición Columnas Derivadas.- En la información convertida se añade el campo “FEC_PCS_GES”, con la fecha de procesamiento.
 - Conversión de los campos.- Todos los campos de la fuente son convertidos al tipo y tamaño de la tabla destino “TAB_GES_EJC”.

- Tabla Destino.- Finalmente se mapean las columnas con la información transformada y las derivadas para ser enviadas a cargar a la tabla “TAB_GES_EJC” del área de desembarco.

- ETL Desembarco Clientes (5_CRG_DCO_CIUIDAD).

El proceso contiene las tareas para desembarcar la información de las ciudades usando la tabla anteriormente cargada de cartera “TAB_CTR” que se encuentra en el área de desembarco. El flujo de tareas que llevan a cabo este proceso se muestra en la Figura 64.

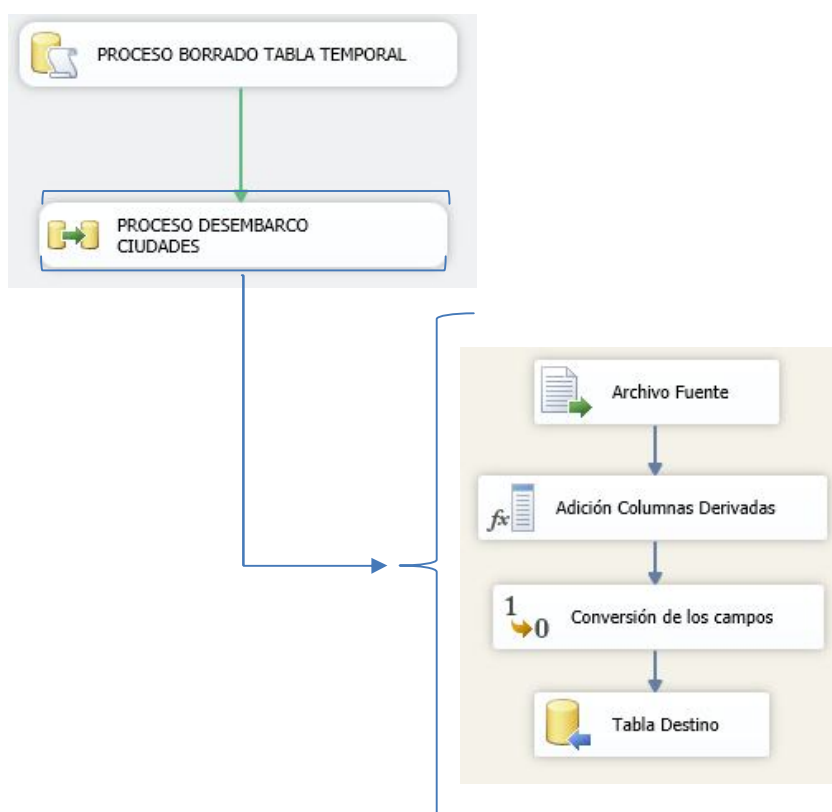


Figura 64. Tareas proceso ETL Desembarco Ciudades.

- **Truncamiento tabla ciudades.-** Tarea que trunca la tabla de clientes “TAB_CIU”.

- **Desembarco Clientes.-** Esta tarea contiene las siguientes subtareas:

- Extracción tabla Cartera.- Extrae la información, de la fecha de proceso, de la tabla Cartera “TAB_CTR” para extraer las ciudades, usando la consulta de la Figura 65.

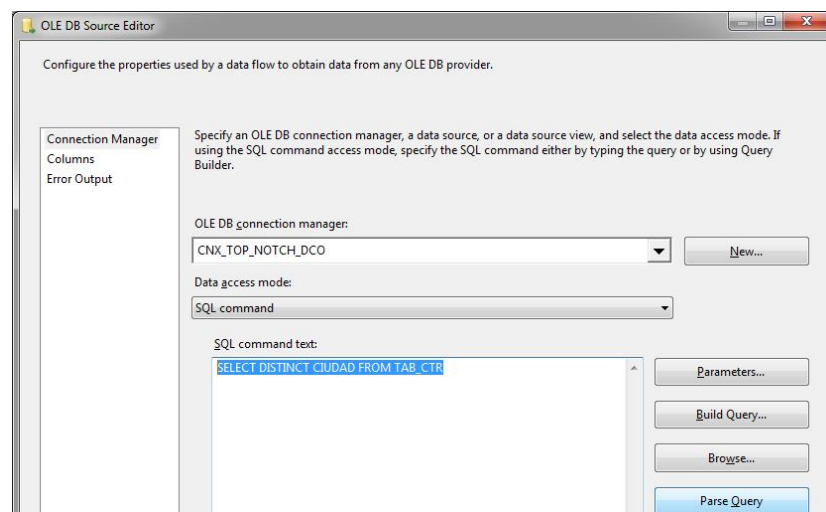


Figura 65. Consulta extracción Ciudades.

- Conversión de los campos.- Todos los campos de la fuente son convertidos al tipo y tamaño de la tabla destino “TAB_CIU”.
 - Fuente Destino.- Finalmente se mapean las columnas con la información transformada y las derivadas para ser enviadas a cargar a la tabla “TAB_CIU” del área de desembarco.
- ETL Desembarco Recuperación Adicional (6_CRG_DCO_RECUPERACION_ADICIONAL).

El proceso contiene las tareas para desembarcar la información de los montos adicionales recuperados usando el archivo con el siguiente formato: RecuperacionNoAsignadaTopNotch_dd-mm-yyyy.txt. El flujo de tareas que llevan a cabo este proceso se muestra en la Figura 66.



Figura 66. Tareas proceso ETL Desembarco Recuperación Adicional.

- **Lectura Configuración.-** Tarea que lee la configuración y asigna a la variable general del proceso, para saber la fecha de lectura del archivo.
- **Borrado Tabla Temporal.-** Tarea que trunca la tabla de recuperación adicional "TAB_RCP_CTR_NO_ASG".
- **Desembarco Monto Recuperación Adicional.-** Esta tarea contiene las siguientes sub-tareas:

- Extracción tabla Cartera.- Extrae la información, de la fecha de proceso, del archivo fuente “RecuperacionNoAsignadaTopNotch_dd-mm-yyyy.txt”.
 - Conversión de los campos.- Todos los campos de la fuente son convertidos al tipo y tamaño de la tabla destino “TAB_RCP_CTR_NO_ASG”.
 - Tabla Destino.- Finalmente se mapean las columnas con la información transformada y las derivadas para ser enviadas a cargar a la tabla “TAB_RCP_CTR_NO_ASG” del área de desembarco.
- ETL Desembarco Ejecutivos (7_CRG_DCO_EJECUTIVOS).

El proceso contiene las tareas para desembarcar la información de los ejecutivos, usando la base de datos fuente “” y cruzando las tablas “IP” y “EMPE”. El flujo de tareas que llevan a cabo este proceso se muestra en la Figura 67.



Figura 67. Tareas proceso ETL Desembarco Ejecutivos.

- **Borrado Tabla Temporal.-** Tarea que trunca la tabla de recuperación adicional “TAB_EMP”.
- **Desembarco Ejecutivo.-** Esta tarea contiene las siguientes subtareas:
 - Extracción tablas fuentes.- Extrae la información, de la fecha de proceso, de la tabla “IP” y “EMPE” usando la consulta que se indica en la Figura 68.

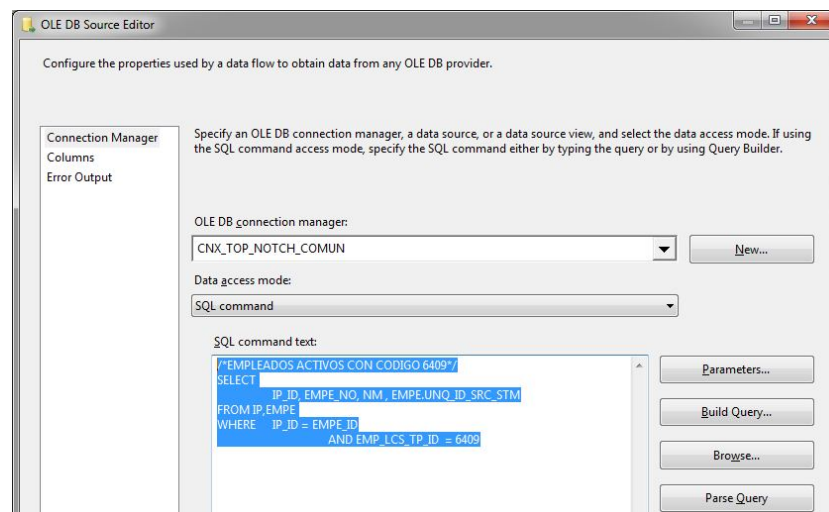


Figura 68. Consulta extracción Ejecutivos.

- Tabla Destino.- Finalmente se mapean las columnas para cargar la tabla “TAB_EMP” del área de desembarco.

4.1.1.1.2 Proceso 2. Carga de las tablas de dimensión al área del DWH.

Este proceso se lleva a cabo con el siguiente esquema de ETL:

Tabla 11. Esquema de procesos ETL de carga tablas dimensiones.

No.	Nombre ETL
1	8_CRG_DWH_DIMENSIONES.dtsx

- ETL Carga Dimensiones (8_CRG_DWH_DIMENSIONES).

El proceso contiene las tareas para cargar las tablas que van a servir como insumo para la construcción de las dimensiones del cubo, la fuente que usa son las tablas del área de desembarco. El flujo de tareas que llevan a cabo este proceso se muestra en la Figura 69.



Figura 69. Flujo de tareas carga tablas de dimensión al DWH.

- **Mantenimiento Fechas.-** Tarea que carga la fecha de procesamiento a la tabla “FEC” del área de desembarco hacia el área del DWH.
- **Mantenimiento Dimensión Ciudades.-** Tarea que carga las ciudades a la tabla “CIU” del área de desembarco hacia el área del DWH.
- **Mantenimiento Dimensión Clientes.-** Tarea que carga los clientes a la tabla “CLI” del área de desembarco hacia el área del DWH.
- **Mantenimiento Dimensión Ejecutivos.-** Tarea que carga los ejecutivos a la tabla “EJC” del área de desembarco hacia el área del DWH.

4.1.1.1.3 Proceso 3. Carga de la tabla de hechos al área del DWH.

Este proceso se lleva a cabo con el siguiente esquema de ETL:

Tabla 12. Esquema de procesos ETL de carga tabla de hechos.

No.	Nombre ETL
1	9_CRG_DWH_HECHOS.dtsx

- ETL Carga Hechos (9_CRG_DWH_HECHOS).

El proceso contiene las tareas para cargar la tabla que va a servir como insumo para la construcción de los hechos del cubo, la fuente que usa son

las tablas del área de desembarco. El flujo de tareas que llevan a cabo este proceso se muestra en la Figura 70.



Figura 70. Flujo de tareas carga tabla de hechos al DWH.

- **Asignación Fecha a procesar.-** La tarea asigna, al proceso en general, la fecha en la que se debe realizar cada acción.
- **Eliminación data fecha a procesar.-** Tarea para borra la información de la tabla de hechos “RCP_CTR” según la fecha en la que se va a procesar.
- **Carga tabla Hechos.-** Tarea que carga la tabla de hechos “RCP_CTR”, del área de desembarco hacia el área del DWH.

4.1.1.2. Actualización.

El proceso ETL para la actualización del DWH es muy similar al de la Carga Inicial pero definidas las siguientes políticas de actualización, que han sido elaboradas a través de las reuniones con los usuarios; y son las siguientes:

- La información se actualizará el último día de cada mes.
- Para hacer un reproceso se necesita establecer la fecha de reproceso en la tabla de configuraciones, tener los archivos con la fecha de reproceso en la ruta compartida y finalmente correr el flujo.
- Estas acciones se realizarán durante un periodo de prueba, posterior a este tiempo se analizará la manera más eficiente de generar las actualizaciones; decisión que se basará en el estudio de los cambios que se producen en los OLTP y que afectan al contenido del Data Warehouse.

4.1.1.3. Construcción del Cubo de Recuperación de Cartera.

A continuación se creará el cubo multidimensional para la recuperación de cartera, que será llamado "CARTERA" y que estará basado en el modelo lógico diseñado en el punto 3.1.3.4 basados en la metodología Hefesto. La figura 71., presenta el diseño de las tablas dentro de la herramienta Microsoft SQL Manager 2012.

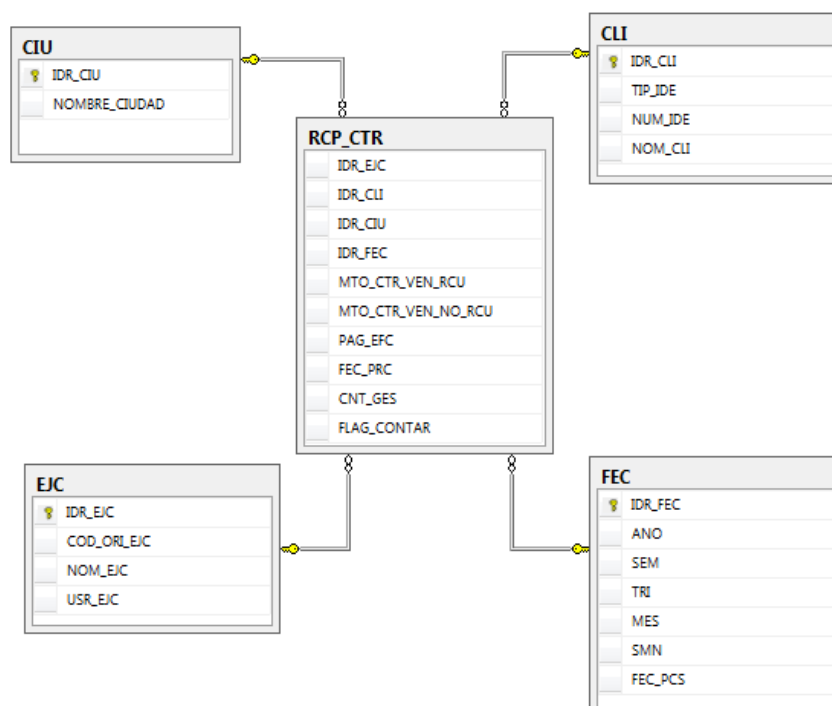


Figura 71. Diseño del modelo del cubo multidimensional Recuperación Cartera.

Cabe denotar que todas las tareas de construcción del cubo multidimensional, que se describirán en los pasos a continuación, serán enfocadas a las herramientas de Microsoft como son: SQL Manager 2012 y Visual Studio 2010.

4.1.1.3.1 Creación del proyecto Analysis Services Multidimensional.

El primer paso a realizar es crear un proyecto de tipo Business Intelligence – Analysis Services. La figura 72., describe visualmente el tipo de proyecto a crear dentro Visual Studio.

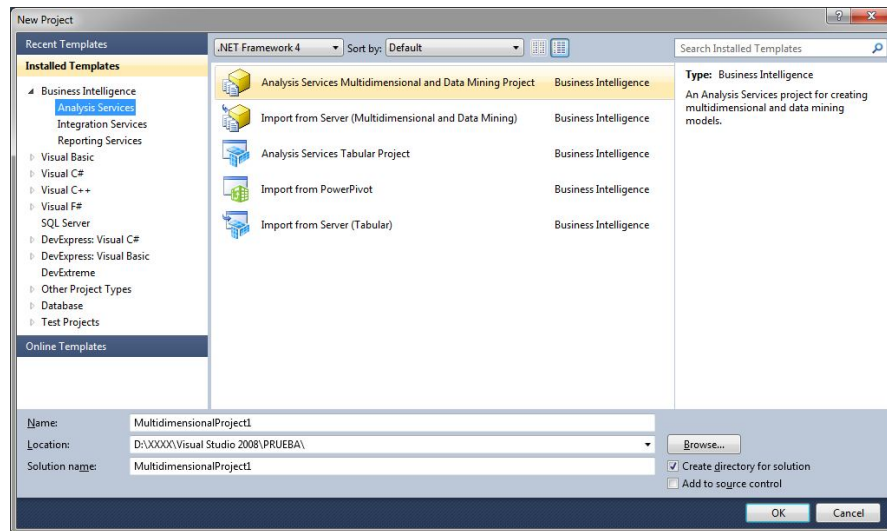


Figura 72. Selección del proyecto para creación del Cubo Multidimensional.

4.1.1.3.2 Configuración de la conexión a la base del DWH.

Situados en el área de trabajo del proyecto, se observa cuatro principales carpetas en el explorador de Soluciones, la primera carpeta con el nombre “Data Sources” servirá para crear las conexiones del cubo multidimensional. En la Figura 73, se visualizan las carpetas mencionadas.

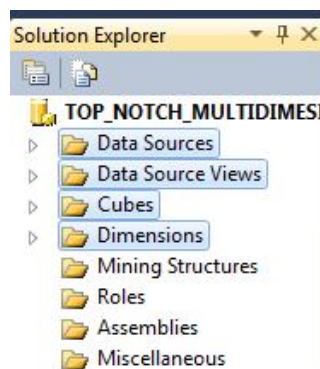


Figura 73. Carpetas de trabajo Explorador de Soluciones.

Para crear el objeto de conexión a la base de datos se da un clic derecho sobre la carpeta “Data Sources” y se elige la opción “New Data Source”,

posteriormente se procede a establecer los parámetros de conexión como se muestra en la Figura 74.

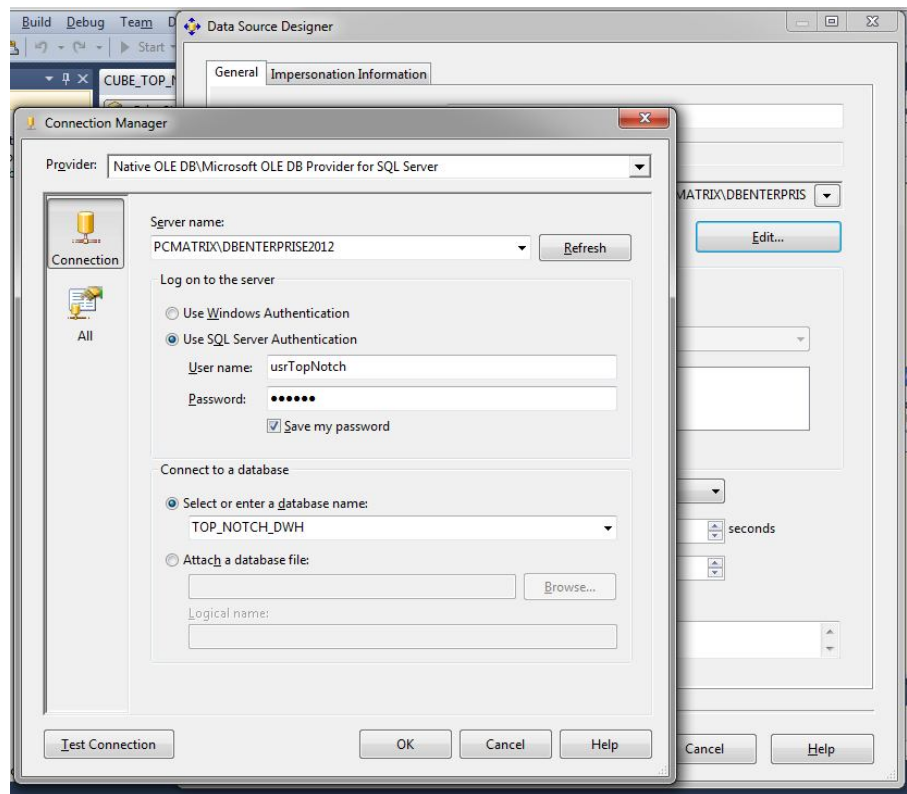


Figura 74. Parámetros de conexión a la base del DWH.

4.1.1.3.3 Configuración de la vista base del Cubo Multidimensional.

En la carpeta “Data Sources Views” se da clic derecho, se selecciona “New Data Sources View”, se escoge la conexión y se seleccionan las tablas de dimensión y hechos a diseñar. En la Figura 75., se visualiza la selección de las tablas.

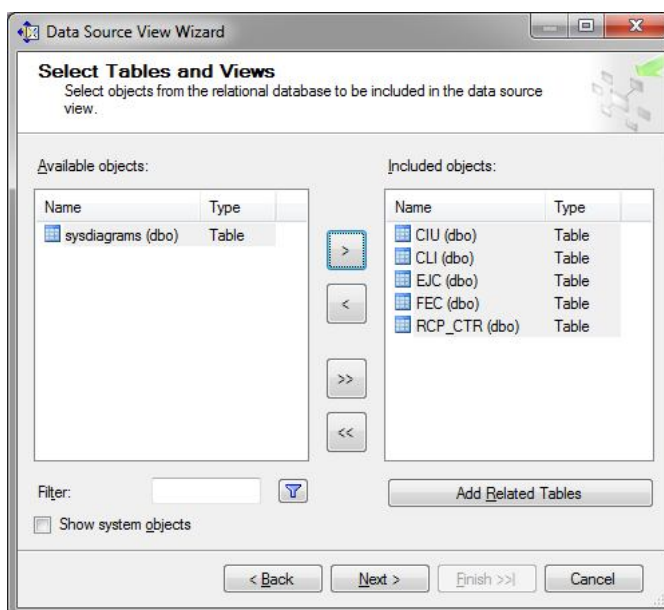


Figura 75. Selección de tablas de dimensión y hechos.

4.1.1.3.4 Creación de las dimensiones del Cubo Multidimensional.

En la carpeta “Dimensions” se da clic derecho, se selecciona “New Dimension” y se procede con las siguientes acciones de creación de dimensión:

1. **Selección del método de creación.-** Esta opción permite seleccionar el método de generación de las tablas fuente para crear la dimensión. Para la construcción de todas las dimensiones del Cubo de Recuperación de Cartera se selecciona la primera opción “Use an existing table”, como se visualiza en la Figura 76.



Figura 76. Selección método generación dimensiones.

2. **Selección de la tabla base de la dimensión.-** En esta opción se selecciona la vista fuente, la tabla base de la dimensión, la columna clave de la tabla y la columna nombre de la dimensión. Las figuras a continuación describen la creación de las dimensiones: ciudades, fechas, ejecutivos y clientes.

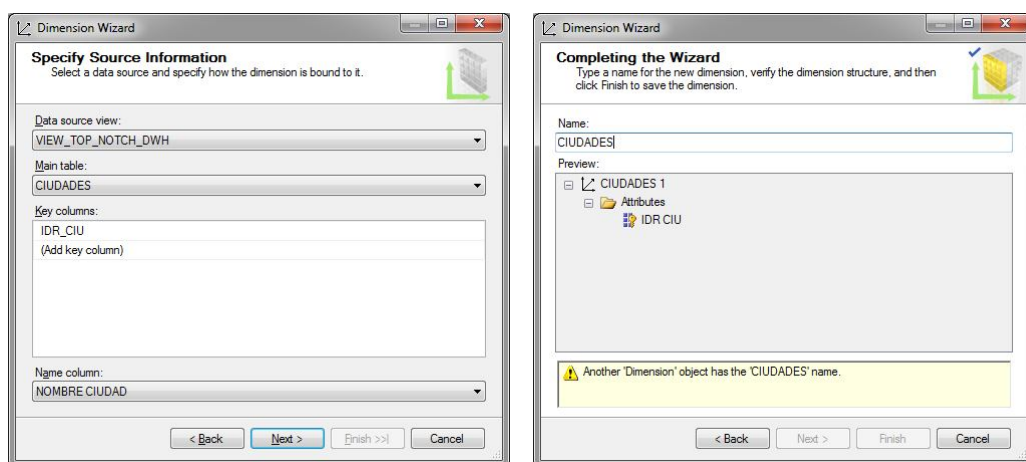


Figura 77. Selección fuente y nombre para la dimensión CIUDADES.

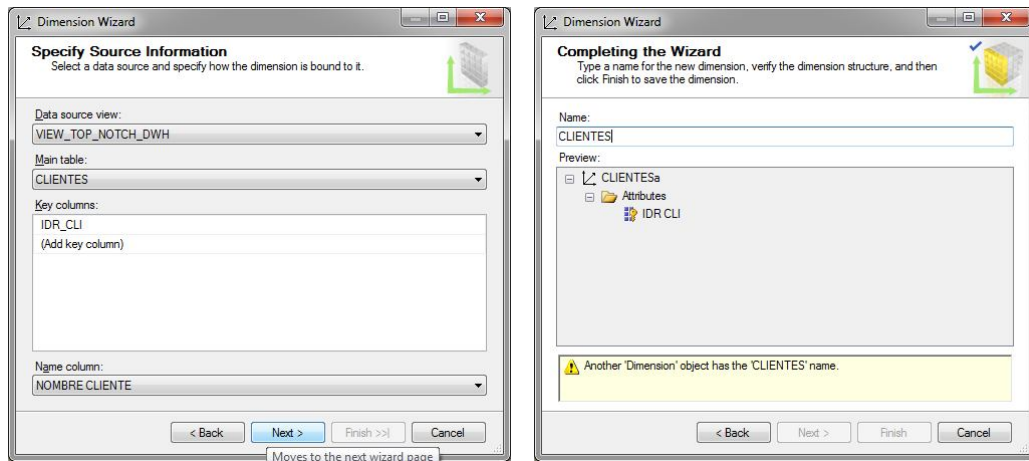


Figura 78. Selección fuente y nombre para la dimensión CLIENTES.

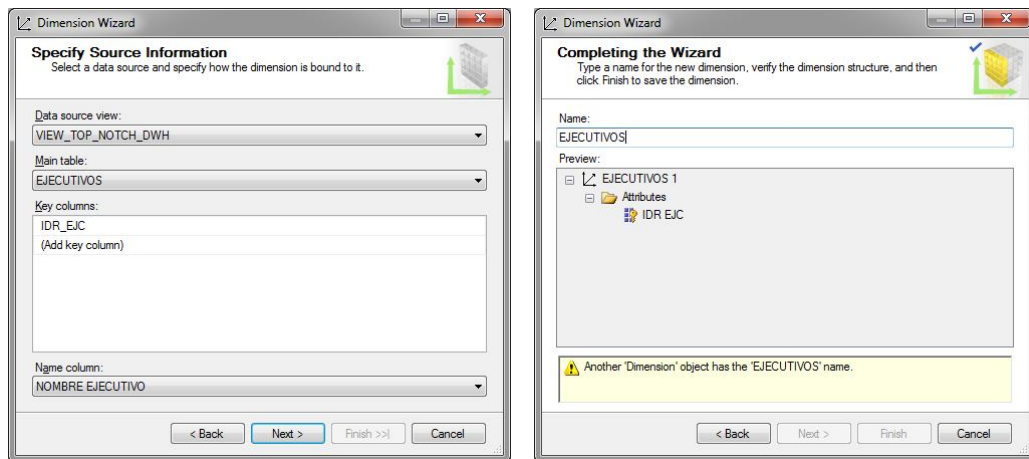


Figura 79. Selección fuente y nombre dimensión EJECUTIVOS.

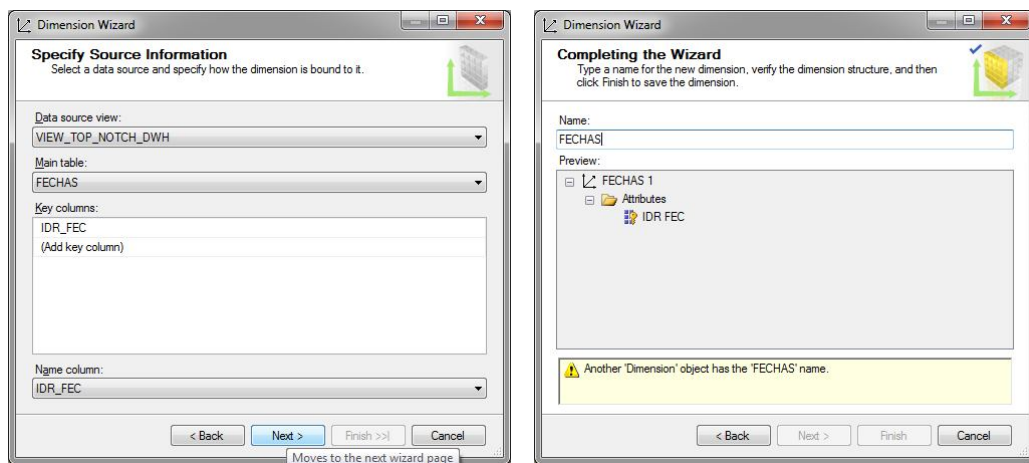


Figura 80. Selección fuente y nombre dimensión FECHAS.

4.1.1.3.5 Creación del Cubo Multidimensional.

Para crear el cubo de Recuperación de Cartera se da clic sobre la carpeta “Cubes” y se selecciona la opción “New Cube”. Posteriormente se realizan las siguientes acciones:

1. **Selección del método de creación del cubo.**- En este paso seleccionar la primera opción “Use existing tables” como se muestra en la Figura 81.

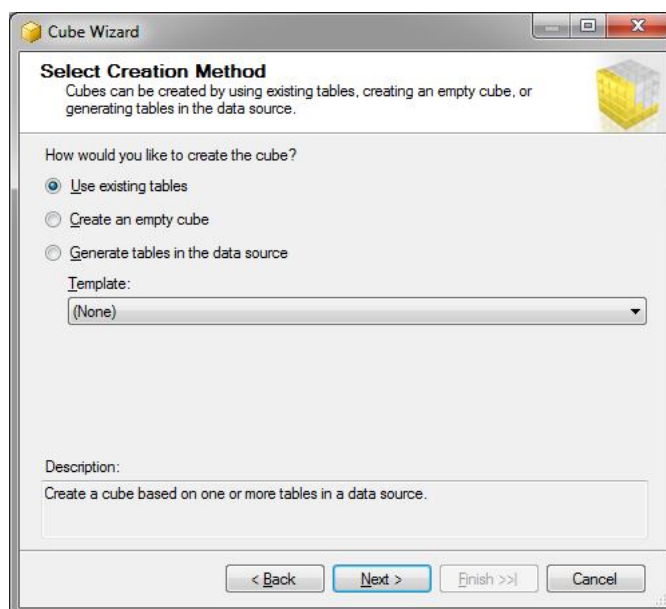


Figura 81. Selección método de creación Cubo Multidimensional.

2. **Selección de la tabla del grupo de medidas.** Este paso permite elegir la tabla en donde está el grupo de métricas. Para el cubo de Recuperación de Cartera se tiene en la tabla “CARTERA”.

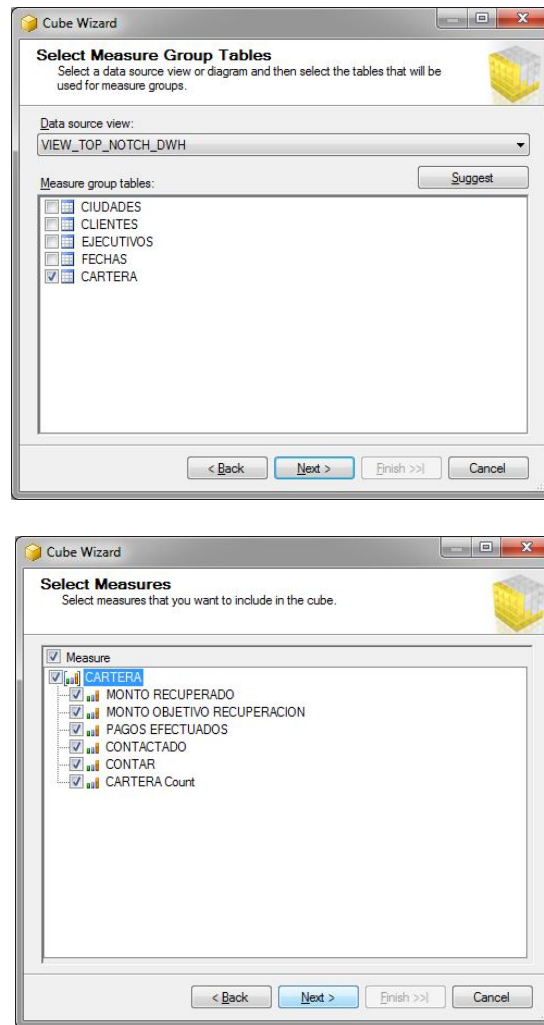


Figura 82. Selección de la tabla fuente del grupo de métricas.

3. **Selección de las dimensiones.**- En este paso se podrá elegir las dimensiones que formarán parte de la construcción del cubo, como se visualiza en la Figura 83.

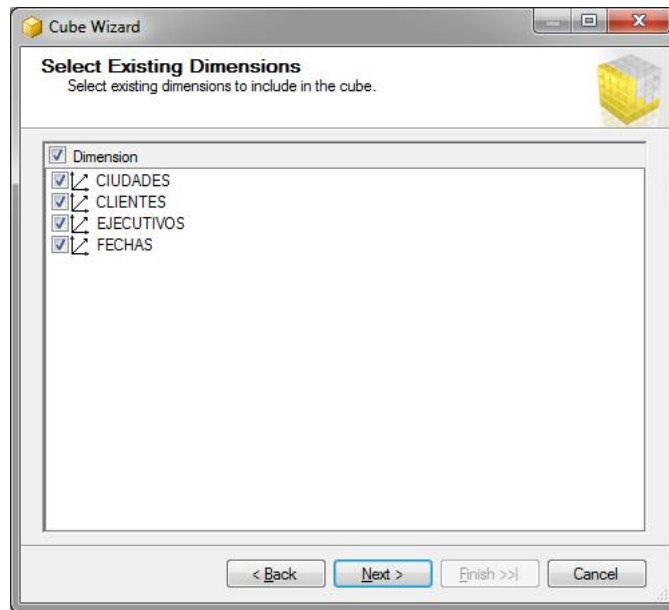


Figura 83. Selección dimensiones del Cubo de Recuperación de Cartera.

4. Asignación del nombre del Cubo.

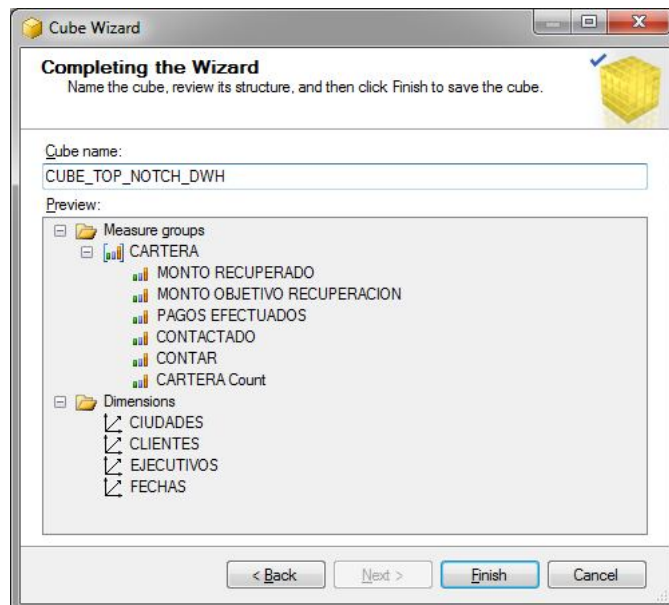


Figura 84. Asignación del nombre del Cubo para la Recuperación de Cartera.

Como paso final de la construcción del cubo, este se muestra en la Figura 85.

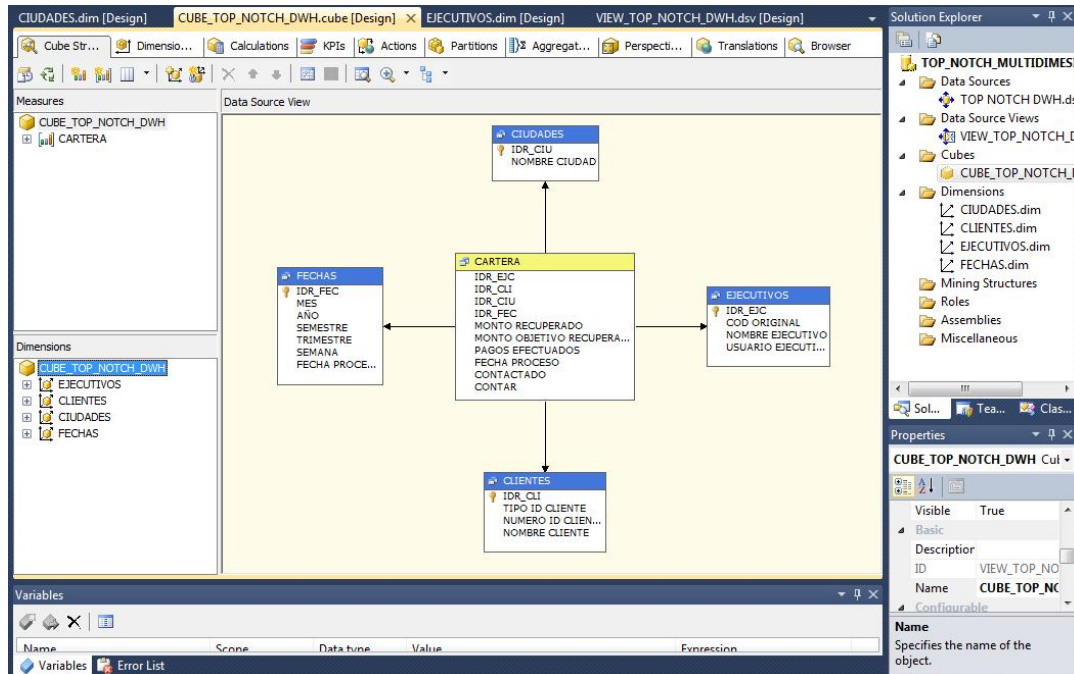


Figura 85. Asignación del nombre del Cubo para la Recuperación de Cartera.

4.1.1.3.6 Creación de Jerarquías.

Existe una sola jerarquía que se creará y pertenece a la dimensión Fecha, la cual será visualizada en la Figura 86.

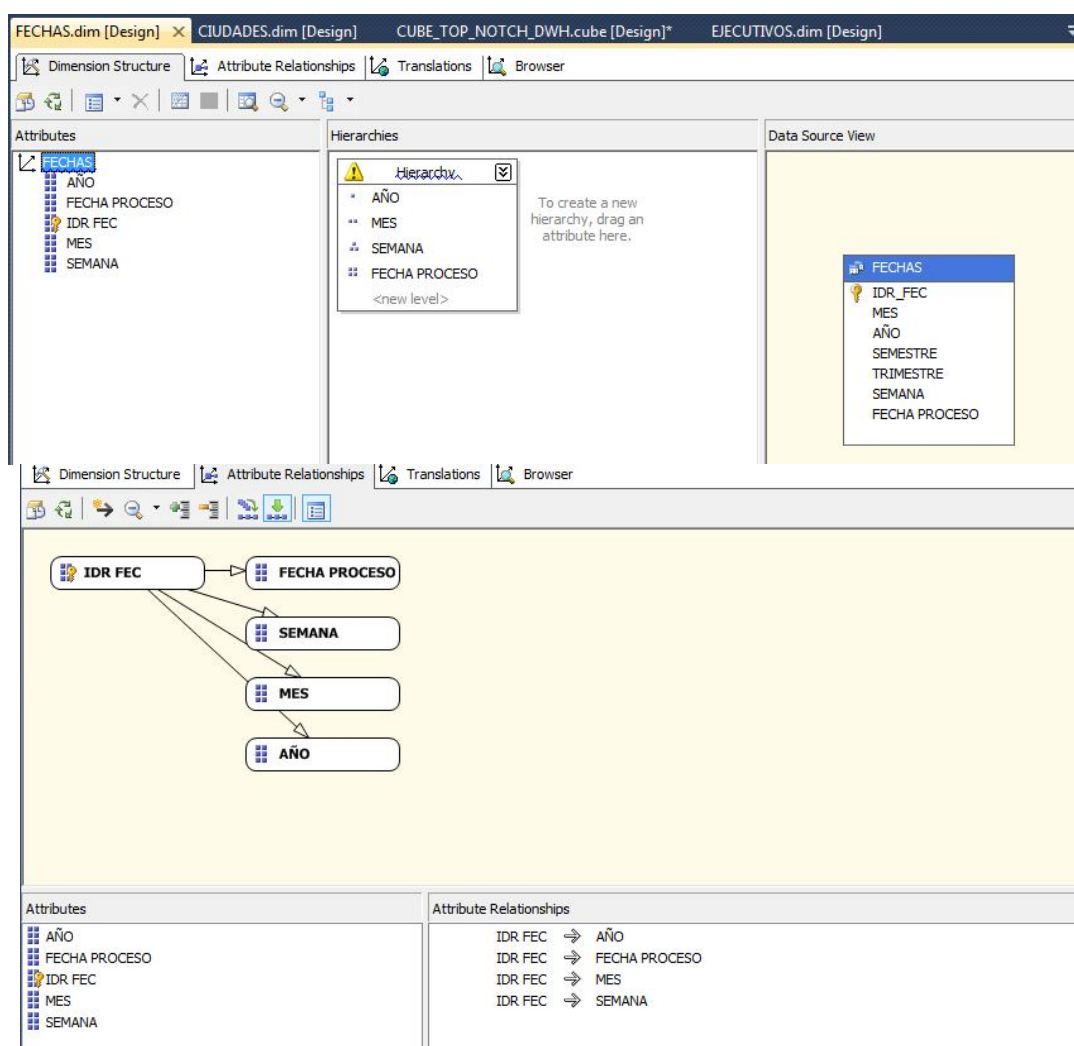


Figura 86. Jerarquización dimensión FECHAS.

4.1.1.3.7 Creación de Indicadores.

En este momento se crearán los indicadores que serán incluidos en el cubo “Recuperación de Cartera”:

- De la tabla de hechos “CARTERA”, se procederá a crear el indicador “EFICIENCIA RECUPERACION CARTERA EJECUTIVO X FECHA”, basado en el siguiente código MDX:

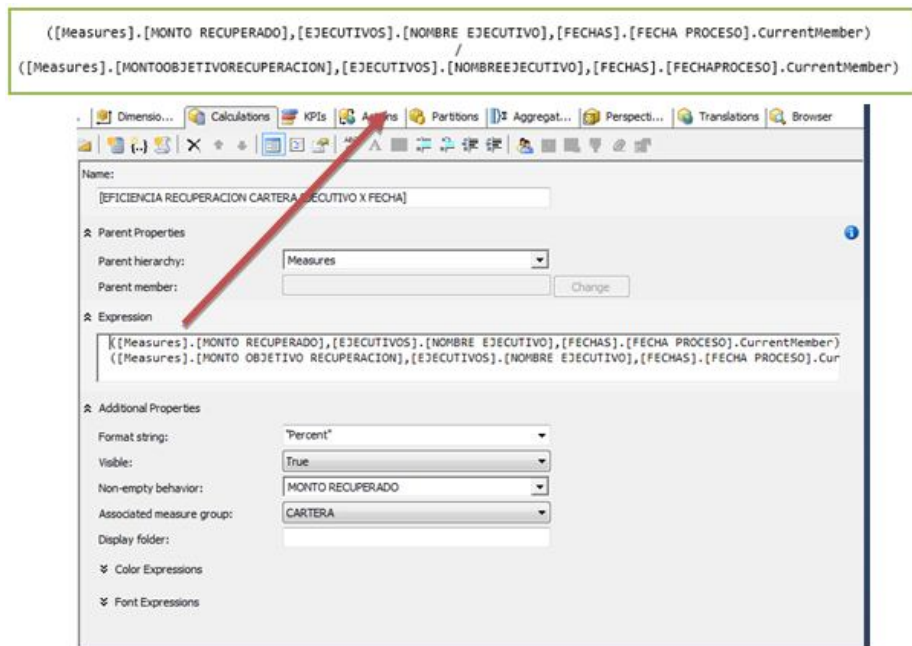


Figura 87. Código MDX para construir el indicador “EFICIENCIA RECUPERACIÓN CARTERA EJECUTIVO X FECHA”.

- De la tabla de hechos “CARTERA”, se procederá a crear el indicador “EFICIENCIA CONTACTAR CLIENTES”, basado en el siguiente código MDX:

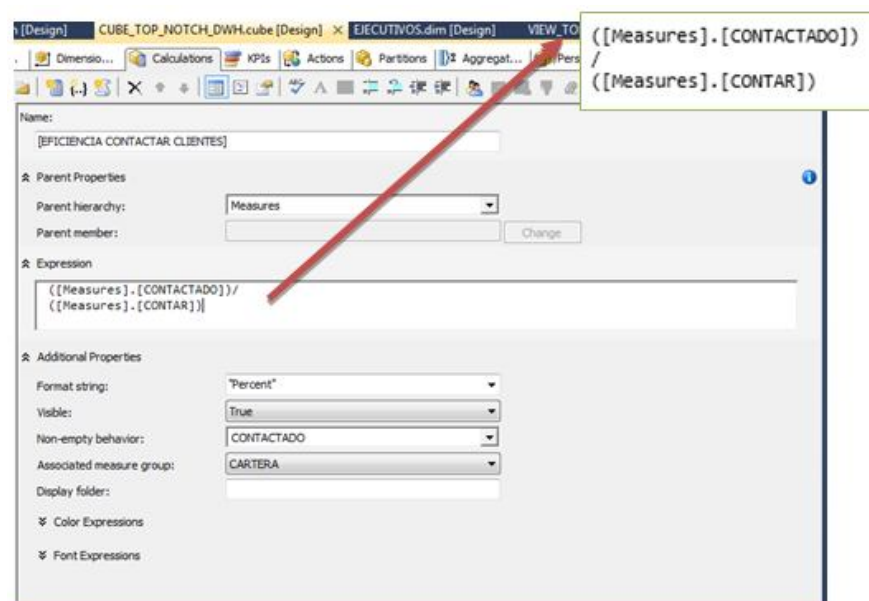


Figura 88. Código MDX para construir el indicador “EFICIENCIA CONTACTAR CLIENTES”.

- De la tabla de hechos “CARTERA”, se procederá a crear el indicador “EFICIENCIA CONTACTAR CLIENTES”, basado en el siguiente código MDX:

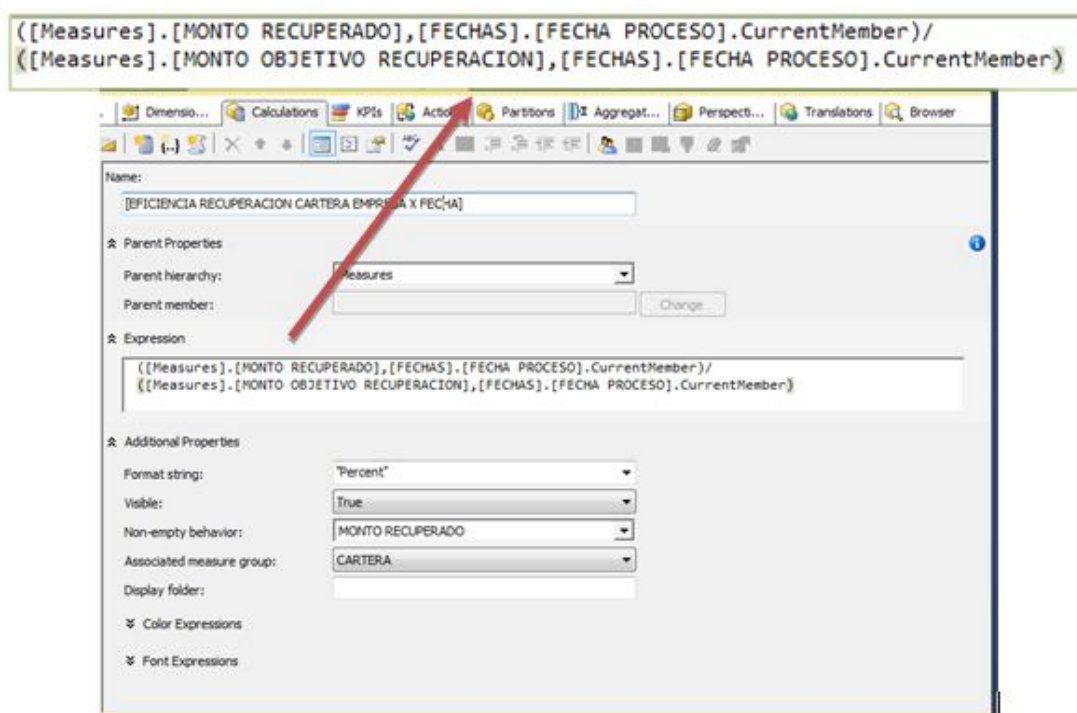


Figura 89. Código MDX para construir el indicador “EFICIENCIA RECUPERACIÓN CARTERA EMPRESA X FECHA”.

4.1.1.3.8 Creación de KPI.

Los KPI que serán construidos en el cubo de Recuperación de Cartera se mencionan a continuación:

- De la tabla de hechos “CARTERA”, se procederá a crear el KPI “KPI INCREMENTO RECUPERACIÓN CARTERA”, que analiza el incremento en la recuperación de cartera entre los dos últimos periodos cargados. El KPI está basado en el siguiente código MDX:

```

(
    SUM([Measures].[MONTO RECUPERADO],[FECHAS].[Hierarchy].[AÑO].&[2013].&[11].&[46].&[2013-11-10])
) / SUM([Measures].[MONTO RECUPERADO],[FECHAS].[Hierarchy].[AÑO].&[2013].&[11].&[44].&[2013-11-02])

```

The screenshot shows the configuration for a KPI named "KPI INCREMENTO RECUPERACION CARTERA". The "Associated measure group" is "CARTERA". The "Value Expression" field contains the MDX code: `(SUM([Measures].[MONTO RECUPERADO],[FECHAS].[Hierarchy].[AÑO].&[2013].&[11].&[46].&[2013-11-10])) / SUM([Measures].[MONTO RECUPERADO],[FECHAS].[Hierarchy].[AÑO].&[2013].&[11].&[44].&[2013-11-02])`. The "Goal Expression" is "0.02". The "Status indicator" is "Traffic light" and the "Status expression" is a case statement: `Case When KpiValue("KPI INCREMENTO RECUPERACION CARTERA")>=0.02 Then 1 When KpiValue("KPI INCREMENTO RECUPERACION CARTERA")>=0.01 Then 0 When KpiValue("KPI INCREMENTO RECUPERACION CARTERA")<0.01 Then -1 End`. The "Trend indicator" is "Standard arrow".

Figura 90. Código MDX para construir el KPI “KPI INCREMENTO RECUPERACIÓN CARTERA”.

- De la tabla de hechos “CARTERA”, se procederá a crear el KPI “KPI INCREMENTO CONTACTAR A LOS CLIENTES”, que analiza el incremento en contactar a los clientes entre los dos últimos periodos cargados. El KPI está basado en el siguiente código MDX:

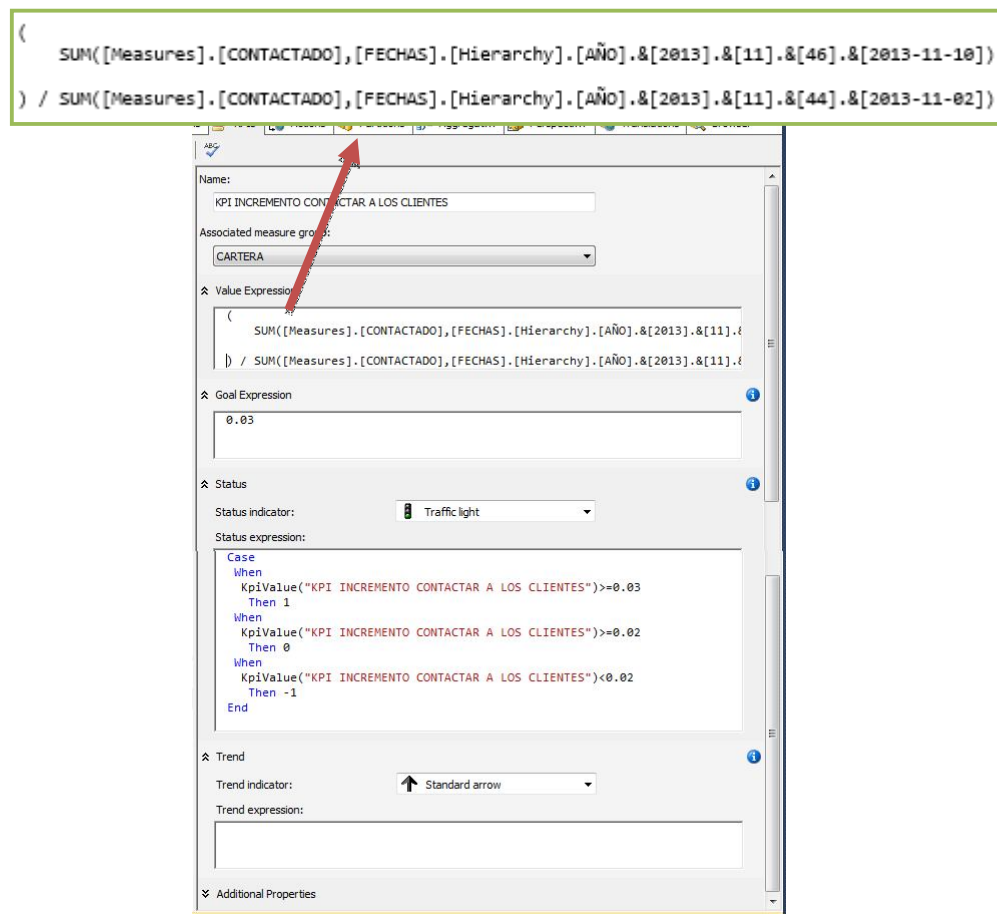


Figura 91. Código MDX para construir el KPI “KPI INCREMENTO CONTACTAR A LOS CLIENTES”.

4.2. Aplicaciones para Usuarios Finales.

4.2.1. Tablero de Control.

El cubo se ha explotado usando la herramienta de Microsoft Visual Studio 2010 creando una aplicación de Windows para construir el Tablero de control con conexión al cubo multidimensional.

El tablero de control es una herramienta gráfica para crear, editar y visualizar cuadros de mando; ofrece las siguientes perspectivas:

- Fichero: Permite crear, guardar o cargar un tablero de control desde cero o desde una plantilla.



Figura 92. Grupo de botones para el manejo de Ficheros.

- Fuentes de Datos: Sirve para definir las conexiones a las fuentes de datos utilizadas por los componentes.




Figura 93. Grupo de botones para el manejo de la Fuente de Datos.

- Componentes: Para agregar y configurar los distintos componentes que conforman el tablero de control; componentes visuales como cuadro de texto, tablas, tablas cruzadas, gráficas de barras, gráficas de pastel, indicador o fichas.



Figura 94. Grupo de botones para el manejo de la Fuente de Datos.

Para empezar a diseñar un tablero de control es necesario iniciar el servicio de SQL Server Analysis Services y ejecutar la aplicación de Windows ( TOP_NOTCH_DASHBOARD.exe).

La creación de tableros de control se encuentra en mayor detalle en el Manual de Usuario del Tablero de Control en el Anexo D.

A continuación, en la Figura 95., se visualiza el cuadro de mandos aplicado a graficar los indicadores de Recuperación Cartera por Ejecutivo y empresa.

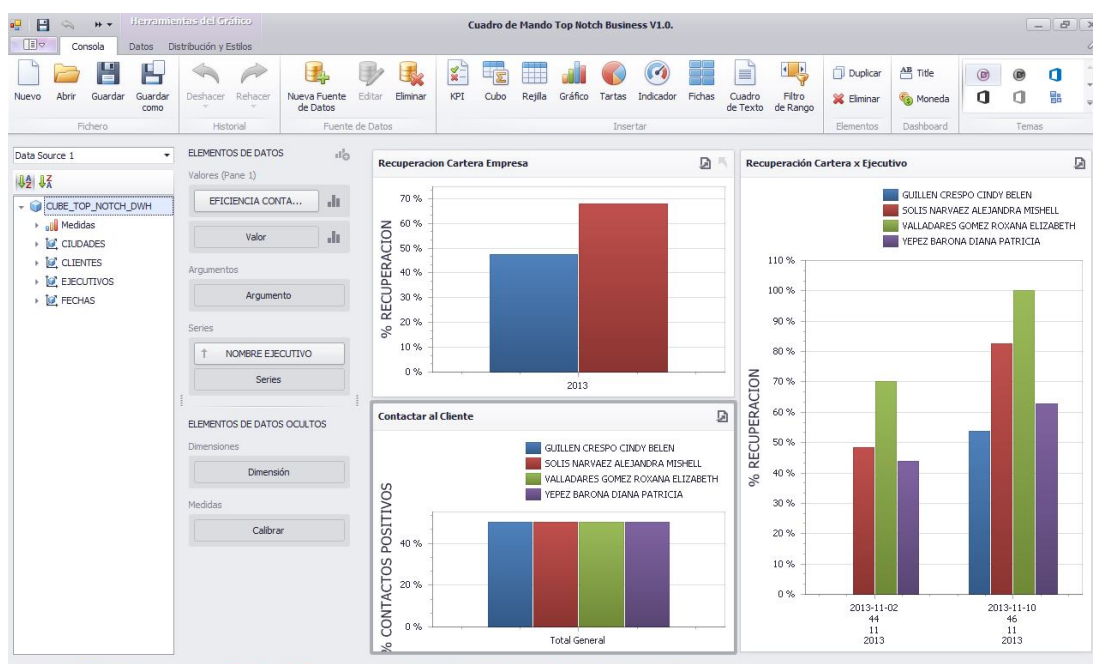


Figura 95. Grupo de botones para el manejo de la Fuente de Datos.

4.2.2. Aplicación WEB.

El cubo también se ha explotado usando la herramienta de Microsoft Visual Studio 2010 creando una aplicación Web para la intranet de la empresa TopNotch Business.

La aplicación Web ofrece las siguientes perspectivas:

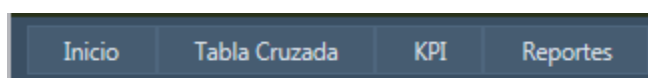


Figura 96. Grupo de botones de la aplicación Web.

- Inicio: Permite describir el objetivo de la aplicación Web y presenta algunas gráficas relacionadas a la empresa.
- Tabla Cruzada: Componente para explotar el cubo basadas en los montos de recuperación como métricas y las dimensiones de visualización.
- KPI: Contiene una tabla cruzada que presenta a los dos KPI construidos.
- Reportes: Esta sección cuenta con reportes diseñados para validar la última data que se cargó de la fuente versus la bodega de información DWH.

El detalle de como interactuar con la aplicación Web se encuentra en mayor detalle en el Manual de Usuario aplicación Web en el Anexo E.

A continuación se visualizan los perfiles de la aplicación Web.

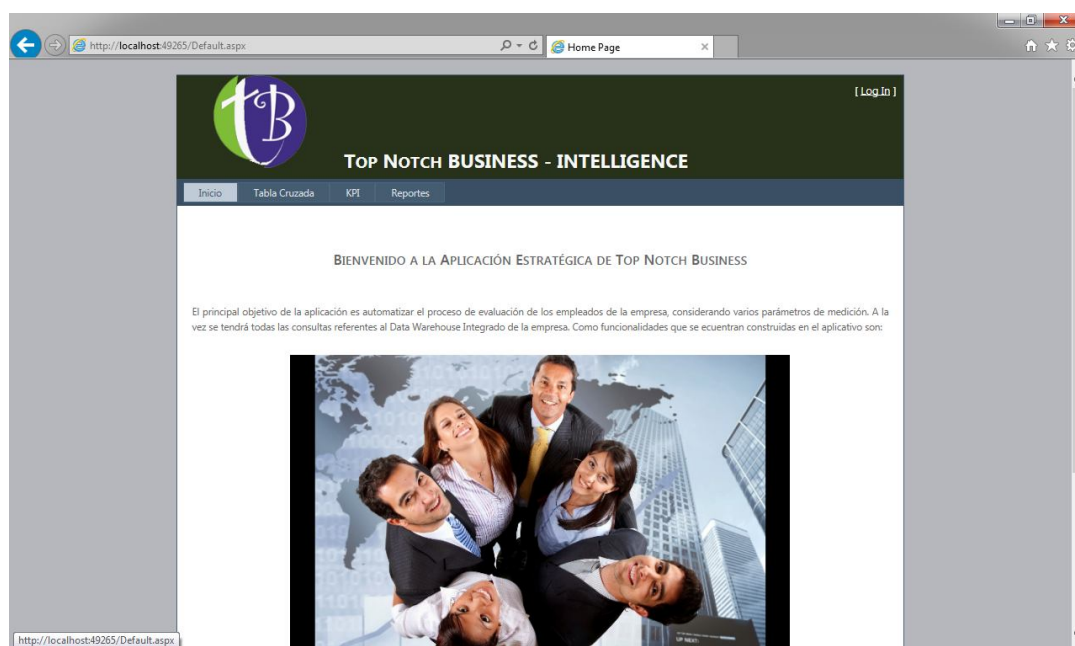


Figura 97. Perfil Inicio de la aplicación Web.

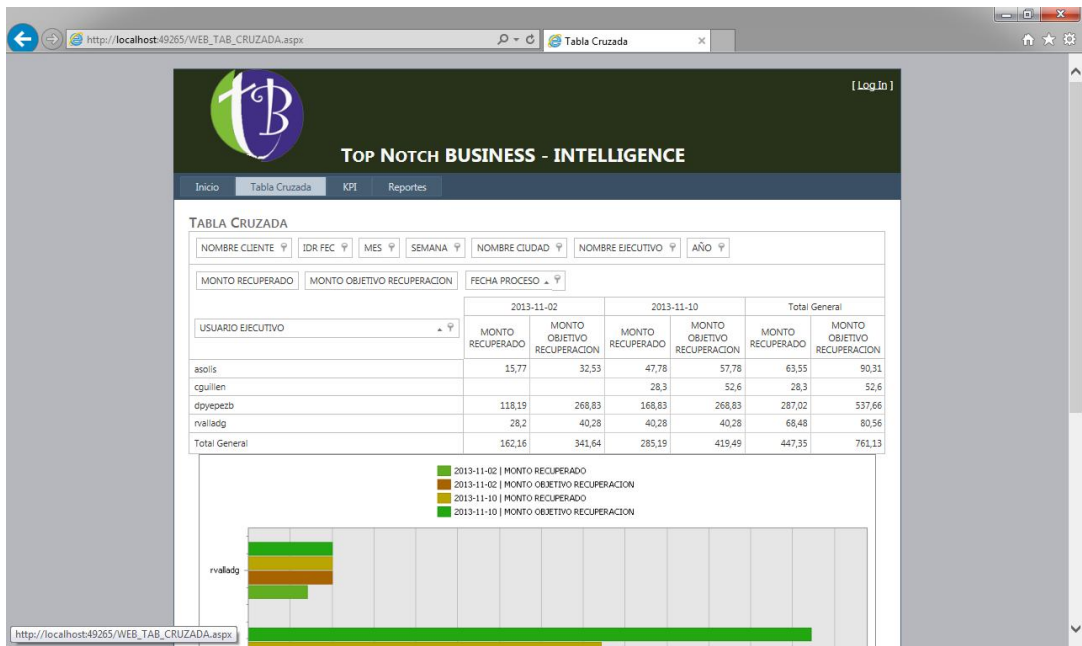


Figura 98. Perfil Tabla Cruzada de la aplicación Web.

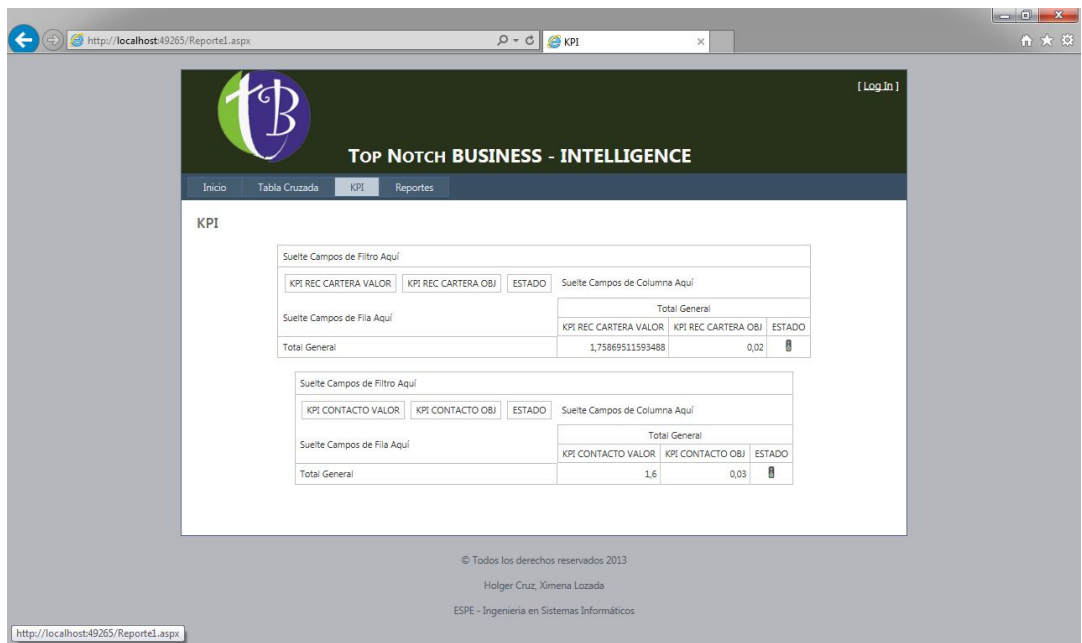


Figura 99. Perfil KPI de la aplicación Web.

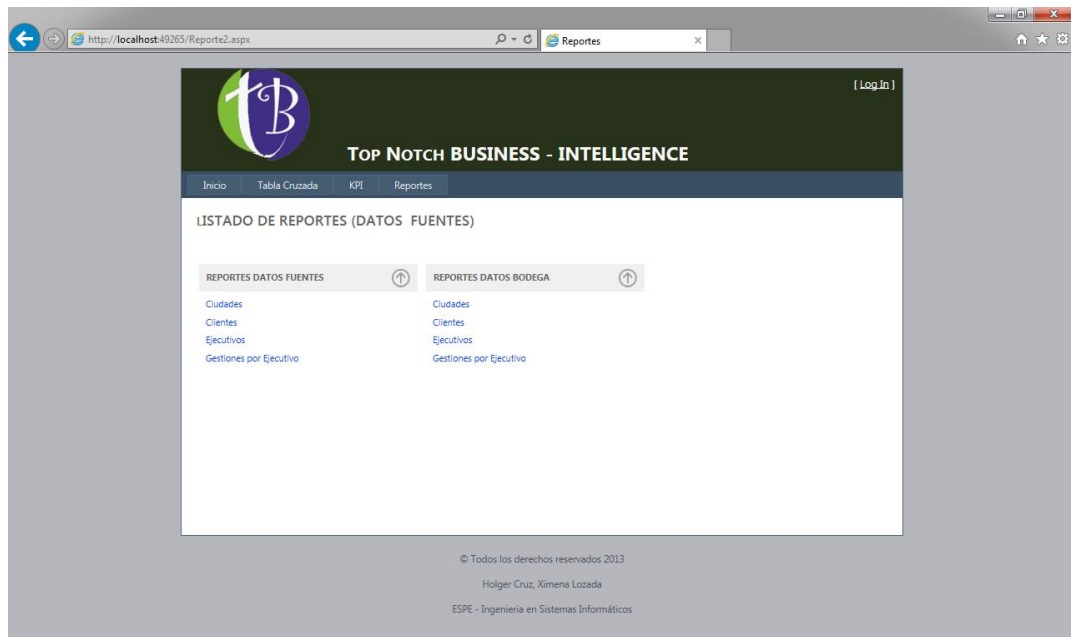


Figura 100. Perfil Reportes de fuente y bodega aplicación Web.

CAPÍTULO 5

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

La teoría del Data Warehousing, Business Intelligence y la metodología Hefesto permite la implementación correcta de cada uno de los pasos para la construcción del Data Warehouse, de forma estructurada y organizada.

Por otra parte, un punto importante a destacar, es el costo en tiempo tomado en la fase de análisis de requerimientos y OLTP de la metodología Hefesto, por la carencia de documentación y procesos claros sobre el negocio; lo cual implica un mayor esfuerzo en la construcción del proyecto y un desfase en la planificación.

SATB logró muy buenos resultados al realizar la carga de la información y el posterior análisis de la información multidimensional; este referente le permitirá, a la empresa TopNotch Business, tomar las decisiones estratégicas oportunas para el cumplimiento de sus objetivos empresariales.

5.2. Recomendaciones

Realizar un plan de mantenimiento y actualización, tanto para el software como para la infraestructura; procedimiento importante para mantener en correcto funcionamiento el Data Warehouse, debido a su demandante crecimiento en espacio físico.

Crear y administrar los niveles de acceso a los usuarios, bajo un listado de acceso controlado a los servidores, aplicación, reportes y cubo multidimensional, protegiendo a la información de la empresa.

Definir claramente los lineamientos y roles de cada una de las áreas involucradas (e-Governance), para que de esta manera exista una buena integración y manejo del proyecto tecnológico.

Incrementar la funcionalidad de la aplicación Web del proyecto SATB, operativizando todos los elementos de análisis multidimensional existente en la aplicación de escritorio.

BIBLIOGRAFÍA

- Bernabeu, D. (6 de Mayo de 2009). *Arquitectura Data Warehouse*. Obtenido de HEFESTO: Metodología para la Construcción de un Data Warehouse: <http://www.dataprix.com/32-oltp>
- Bernabeu, D. (13 de Mayo de 2009). *Estructura Data Warehouse*. Obtenido de HEFESTO: Metodología para la Construcción de un Data Warehouse: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-conceptos-8>
- Bernabeu, D. (6 de Mayo de 2009). *Herramientas de consulta y análisis, Arquitectura Data Warehouse*. Obtenido de HEFESTO: Metodología para la Construcción de un Data Warehouse: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-concepto-17>
- Bernabeu, D. (6 de Mayo de 2009). *OLTP*. Obtenido de HEFESTO: Metodología para la Construcción de un Data Warehouse: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto>
- Bernabeu, D. (12 de Mayo de 2009). *Proceso de BI*. Obtenido de DATAPRIX: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/1-business-intelligence/13-proceso-bi>
- Bernabeu, D. (6 de Mayo de 2009). *Query Manager, Arquitectura Data Warehouse*. Obtenido de HEFESTO: Metodología para la Construcción de un Data Warehouse: <http://www.dataprix.com/blogs/bernabeu-dario/hefesto-v20>
- Bernabeu, D. (19 de Julio de 2010). *HEFESTO: Metodología para la Construcción de un Data Warehouse*. Obtenido de <http://tgx-hefesto.blogspot.com/2010/07/hefesto-v20.html>
- Bernabeu, D. (17 de 04 de 2013). *Clasificación KPI*. Obtenido de HEFESTO: Metodología para la Construcción de un Data Warehouse: <http://www.dataprix.com/category/tags/indicadores>
- Curto, J. (7 de Octubre de 2007). *Information Management: Wordpress*. Obtenido de Information Management: <http://informationmanagement.wordpress.com/2007/10/07/data-warehousing-data-warehouse-y-datamart/>
- Dario, B. (12 de Mayo de 2009). *Dataprix Knowledge is the Goal*. Obtenido de <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-conceptos-8>
- Dario, B. (6 de 5 de 2009). *Load Manager*. Obtenido de Load Manager: <http://www.dataprix.com/33-load-manager>

- NextGenerationCenter. (s.f.). *Diálogo TI*. Obtenido de NextGenerationCenter:
<http://www.google.com.ec/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&cad=rja&ved=0CDYQFjAA&url=http%3A%2F%2Fwww.dialogoti.com%2Far%2FscriptServices%2FcourseToPdf.ashx%3FcourseId%3D7d6f058a-2ecd-42a4-8d0b-ff610aa97548&ei=sz1aUoOxCNHpkAe40oCoBw&usg=AFQjC>
- Ralph, k., & Ross, M. (2002). *The data warehouse toolkit : the complete guide to dimensional modeling* (Second ed.). Toronto, United States of America: Wiley Computer Publishing. Recuperado el 19 de 10 de 2013
- Ricardo, I. B. (19 de Julio de 2010). HEFESTO, Metodología para la Construcción de un. Córdoba, Argentina.
- SIXTINA CONSULTING GROUP. (2012). *Biblioteca de Indicadores*. Obtenido de <http://www.sixtinagroup.com/herramientas-y-recursos/biblioteca-de-indicadores/>
- Sixtinagroup. (s.f.). *Biblioteca de Indicadores (KPIs)*. Obtenido de <http://www.sixtinagroup.com/herramientas-y-recursos/biblioteca-de-indicadores/>
- Wikipedia. (8 de Marzo de 2013). *Almacén operacional de los datos*. Obtenido de http://es.wikipedia.org/wiki/Almac%C3%A9n_operacional_de_los_datos
- Wikipedia. (27 de Mayo de 2013). *OLTP*. Obtenido de <http://es.wikipedia.org/wiki/OLTP>