

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

**Desarrollo de un Data Mart para el sistema de servicios sociales
del CONADIS**

**PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN
SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN**

LUIS EDUARDO MOSQUERA ANDRADE
<aikyu.sama@gmail.com>

DIRECTOR: ING. MARÍA HALLO
<maria.hallo@epn.edu.ec>

Quito, Febrero 2011

DECLARACIÓN

Yo, Luis Eduardo Mosquera A, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normativa institucional vigente.

Luis Eduardo Mosquera A.

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Luis Eduardo Mosquera, bajo mi supervisión.

Ing. María Hallo

DIRECTOR DE PROYECTO

AGRADECIMIENTOS

Agradezco a Dios por la inteligencia y la sabiduría que me ha dado durante estos años de estudio.

Agradezco a mi madre y a mi abuelita, que gracias a su apoyo he podido lograr mis metas y objetivos; aprecio su paciencia y dedicación.

Agradezco a mis amigos por su tiempo y por lo buenos momentos que hemos disfrutado hasta ahora.

Luis Eduardo Mosquera A.

CONTENIDO

ÍNDICE DE FIGURAS.....	VI
ÍNDICE DE TABLAS.....	VIII
RESUMEN.....	XII
1. CAPÍTULO 1 : PLANIFICACIÓN, ANÁLISIS Y DISEÑO LÓGICO DEL DATA MART	1
1.1. PLANIFICACIÓN DEL PROYECTO	1
1.1.1. DEFINICIÓN DEL ALCANCE	1
1.1.2. JUSTIFICACIÓN DEL PROYECTO	1
1.1.3. OBJETIVOS.....	2
1.1.3.1. Objetivo general	2
1.1.3.2. Objetivos específicos	2
1.1.4. ETAPAS DEL PROYECTO	2
1.1.5. RESUMEN DEL CRONOGRAMA DE ACTIVIDADES	3
1.2. DEFINICIÓN DE REQUERIMIENTOS.....	4
1.2.1. TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN.....	4
1.2.1.1. Entrevistas al personal de la organización.....	4
1.2.1.2. Resultados de las entrevistas	5
1.2.1.3. Análisis de documentos y sistemas fuente	5
1.2.2. PREPARACIÓN Y PUBLICACIÓN DE REQUERIMIENTOS	7
1.2.2.1. Requerimientos de análisis	7
1.2.2.2. Requerimientos de información.....	8
1.3. DISEÑO ARQUITECTÓNICO, SELECCIÓN DE PRODUCTOS 9	9
1.3.1. ARQUITECTURA DEL DATA MART.....	9
1.3.1.1. Servicios del Back Room	10
1.3.1.2. Repositorios de datos del Back Room	10
1.3.1.3. Servicios del Front Room	11
1.3.1.4. Repositorios de datos del Front Room.....	12

1.3.2. METADATOS	12
1.3.2.1. Metadatos del sistema fuente	13
1.3.2.2. Metadatos del área de ETL	13
1.3.2.3. Metadatos del Data mart y herramientas de acceso	14
1.3.3. SELECCIÓN DE PRODUCTOS	14
1.3.3.1. Productos para el Back Room	15
1.3.3.2. Productos para el Front Room	15
1.4. MODELADO DIMENSIONAL	17
1.4.1. CONFORMACIÓN DE DIMENSIONES	17
1.4.2. CONFORMACIÓN DE LAS TABLAS DE HECHOS	19
1.4.3. DIAGRAMAS DE TABLAS DE HECHOS	21
1.4.3.1. Diagrama de la tabla de hechos de personas discapacitadas	21
1.4.3.2. Diagrama de la tabla de hechos de autorización de vehículos	22
1.4.3.3. Diagrama de la tabla de hechos de ayudas	22
1.4.4. DETALLE DE TABLAS DE HECHOS	23
1.4.4.1. Tabla de hechos de personas discapacitadas	23
1.4.4.2. Tabla de hechos de autorización de vehículos	24
1.4.4.3. Tabla de hechos de ayudas	25
1.4.5. DETALLE DE LAS TABLAS DE DIMENSIONES	25
1.4.5.1. Dimensión persona	26
1.4.5.2. Dimensión persona – discapacidad	27
1.4.5.3. Dimensión persona – hogar	27
1.4.5.4. Dimensión persona – trabajo	28
1.4.5.5. Dimensión localidad	28
1.4.5.6. Dimensión tiempo	29
1.4.5.7. Dimensión vehículo	29
1.4.5.8. Dimensión unidad autorizadora	30
1.4.5.9. Dimensión ayudas	30
1.4.5.10. Dimensión estado ayuda	31

1.4.5.11. Dimensión entidad subvenciona	31
1.4.5.12. Dimensión entidad provee.....	31
1.4.5.13. Dimensión proyecto.....	32
1.4.6. IDENTIFICACIÓN DE LA FUENTE DE DATOS PARA TABLAS DE HECHOS Y DIMENSIONES.....	33
1.4.6.1. Mapeo detallado de la fuente de datos	34
2. CAPÍTULO 2 : DISEÑO FÍSICO, CONSTRUCCIÓN Y DESPLIEGUE DEL DATA MART	40
2.1. DISEÑO FÍSICO	40
2.1.1. ELABORACIÓN DE ESTÁNDARES	40
2.1.1.1. Estándar para los nombres de los objetos de la base de datos...41	
2.1.1.2. Estándar para los nombres de los objetos del ETL.....42	
2.1.2. CONSTRUCCIÓN DEL MODELO FÍSICO DE DATOS	43
2.1.2.1. Modelo físico de la Tabla de hechos personas discapacitadas ...44	
2.1.2.2. Modelo físico de la Tabla de hechos autorización de vehículos ..45	
2.1.2.3. Modelo físico de la Tabla de hechos ayudas	46
2.1.3. ELABORACIÓN DEL PLAN DE ÍNDICES.....	47
2.1.4. CONSTRUCCIÓN DE LOS CUBOS OLAP	51
2.2. DISEÑO Y DESARROLLO DE LA: EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS (ETL)	54
2.2.1. ELABORACIÓN DEL PLAN ETL.....	54
2.2.1.1. Carga inicial de datos.....	56
2.2.1.2. Carga incremental de datos	56
2.2.1.3. Actualización de datos	57
2.2.2. DEFINICIÓN DE TRANSFORMACIONES Y CÁLCULOS	57
2.2.3. ETL DE LAS TABLAS DE DIMENSIÓN	61
2.2.3.1. ETL de la Dimensión persona	62
2.2.3.2. ETL de la Dimensión persona – discapacidad	63
2.2.3.3. ETL de la Dimensión persona – hogar.....	63
2.2.3.4. ETL de la Dimensión persona – trabajo	64

2.2.3.5. ETL de la Dimensión localidad	65
2.2.3.6. ETL de la Dimensión tiempo	66
2.2.3.7. ETL de la Dimensión vehículo.....	66
2.2.3.8. ETL de la Dimensión unidad autorizadora	67
2.2.3.9. ETL de la Dimensión ayudas	67
2.2.3.10. ETL de la Dimensión estado ayuda	68
2.2.3.11. ETL de la Dimensión entidad subvenciona	69
2.2.3.12. ETL de la Dimensión entidad provee	69
2.2.3.13. ETL de la Dimensión proyecto	70
2.2.4. ETL DE LAS TABLAS DE HECHOS	71
2.2.4.1. ETL de la Tabla de hechos de personas discapacitadas.....	73
2.2.4.2. ETL de la Tabla de hechos de autorización de vehículos.....	74
2.2.4.3. ETL de la Tabla de hechos de ayudas.....	75
2.2.5. CONSTRUCCIÓN DE LOS PROCESOS ETL	76
2.3. CONSTRUCCIÓN DE LA APLICACIÓN DE USUARIO	79
2.3.1. DEFINICIÓN DE LAS PLANTILLAS DE LA APLICACIÓN DE USUARIO	80
2.3.1.1. Determinación del grupo de plantillas inicial	80
2.3.1.2. Especificación detallada de las plantillas	82
2.3.1.3. Diseño del árbol de navegación para los reportes	85
2.3.2. CONSTRUCCIÓN DE LOS REPORTES INICIALES	86
2.3.3. DESARROLLO DE LA APLICACIÓN - MAPSIC	88
2.3.3.1. Especificación de los casos de uso.....	89
2.3.3.2. Arquitectura de la aplicación - Mapsic.....	90
2.3.3.3. Diseño de interfaces de usuario.....	91
2.3.3.4. Diagrama de secuencia de la aplicación - Mapsic	93
2.3.3.5. Construcción de la aplicación - Mapsic	94
2.4. DESPLIEGUE	96
2.4.1. PRUEBAS SOBRE LOS PROCESOS ETL.....	97

2.4.1.1. Casos de prueba para la Carga inicial	97
2.4.1.2. Casos de prueba para la Carga Incremental	98
2.4.1.3. Casos de prueba para la Actualización	99
2.4.1.4. Casos de prueba para la validación de dependencias.....	101
2.4.2. PRUEBAS SOBRE LOS REPORTES INICIALES.....	102
2.4.2.1. Caso de prueba para la medida: Persona.....	103
2.4.2.2. Caso de prueba para la medida: Vehículos	103
2.4.2.3. Caso de prueba para la medida: Ayudas	104
2.4.2.4. Caso de prueba para las medidas: Monto, Subvención.....	105
2.4.2.5. Caso de prueba para las medidas de tiempo.....	106
2.4.2.6. Caso de prueba para la medida Porcentaje	107
2.4.3. PRUEBAS SOBRE LA APLICACIÓN MAPSIC	108
2.4.3.1. Casos de prueba para los reportes	108
2.4.3.2. Casos de prueba para la distribución de datos en el mapa	109
3. CAPITULO 3 : CONCLUSIONES Y RECOMENDACIONES	111
3.1. CONCLUSIONES	111
3.2. RECOMENDACIONES	112
BIBLIOGRAFÍA.....	113
GLOSARIO	114
ANEXOS	

ÍNDICE DE FIGURAS

Figura 1. Ciclo de vida del Data warehouse propuesto por Kimball	2
Figura 2. Arquitectura del Data mart.....	9
Figura 3. Diagrama de la tabla de hechos de personas discapacitadas	21
Figura 4. Diagrama de la tabla de hechos de autorización de vehículos	22
Figura 5. Diagrama de la tabla de hechos de ayudas	22
Figura 6. Modelo físico de la Tabla de hechos personas discapacitadas	44
Figura 7. Modelo físico de la Tabla de hechos autorización de vehículos	45
Figura 8. Modelo físico de la Tabla de hechos ayudas	46
Figura 9. Estructura básica del esquema Mondrian	51
Figura 10. Escenario del área ETL	54
Figura 11. Diagrama del Trabajo para dimensiones.....	55
Figura 12. Diagrama del Trabajo para tablas de hechos.....	56
Figura 13. Diagrama del flujo de datos base para dimensiones.....	61
Figura 14. ETL de la Dimensión persona	62
Figura 15. ETL de la Dimensión persona – discapacidad	63
Figura 16. ETL de la Dimensión persona – hogar	63
Figura 17. ETL de la Dimensión persona – trabajo	64
Figura 18. ETL de la Dimensión localidad	65
Figura 19. ETL de la Dimensión tiempo	66
Figura 20. ETL de la Dimensión vehículo.....	66
Figura 21. ETL de la Dimensión unidad autorizadora	67
Figura 22. ETL de la Dimensión ayudas.....	67
Figura 23. ETL de la Dimensión estado ayuda.....	68
Figura 24. ETL de la Dimensión entidad subvencionada	69
Figura 25. ETL de la Dimensión entidad provee	69
Figura 26. ETL de la Dimensión proyecto	70
Figura 27. Diagrama del flujo de datos base para tablas de hechos.....	72

Figura 28. ETL de la Tabla de hechos de personas discapacitadas	73
Figura 29. ETL de la Tabla de hechos de autorización de vehículos	74
Figura 30. ETL de la Tabla de hechos de ayudas	75
Figura 31. Trabajo para las Dimensiones	76
Figura 32. Trabajo para las Tablas de hechos	76
Figura 33. Diagrama de componentes para el proceso ETL	77
Figura 34. Escenario del Front Room	80
Figura 35. Árbol de navegación para los reportes	85
Figura 36. Árbol de navegación de reportes en Openi	86
Figura 37. Reporte - Causa de deficiencia por edades en Openi	86
Figura 38. Reporte – Porcentaje de PCD por genero, vista de impresión	87
Figura 39. Reporte – Distribución de PCD por su situación laboral, gráfico	87
Figura 40. Diseño de la pagina web de la aplicación - Mapsic	91
Figura 41. Contenido de la pestaña Mapa	92
Figura 42. Contenido de la pestaña Reportes	92
Figura 43. Diagrama de secuencia para la aplicación - Mapsic	93
Figura 44. Diagrama de componentes de la aplicación - Mapsic	94

ÍNDICE DE TABLAS

Tabla 1. Cronograma de actividades.....	3
Tabla 2. Procesos del sistema web SIC asociados a los indicadores.....	6
Tabla 3. Fuentes de datos para el Data mart	11
Tabla 4. Especificaciones de la fuente de datos	13
Tabla 5. Políticas de cambio para Dimensiones.....	13
Tabla 6. Especificaciones de la bdd del Data mart.....	14
Tabla 7. Producto seleccionado para la DBMS	15
Tabla 8. Producto seleccionado para el área ETL.....	15
Tabla 9. Producto seleccionado como motor OLAP	16
Tabla 10. Producto seleccionado como herramienta de acceso a datos	16
Tabla 11. Otros servidores	16
Tabla 12. Matriz de Dimensiones y Tablas de hechos	20
Tabla 13. Detalle de la Tabla de hechos de personas discapacitadas.....	23
Tabla 14. Detalle de la Tabla de hechos de autorización de vehículos.....	24
Tabla 15. Detalle de la Tabla de hechos de ayudas.....	25
Tabla 16. Detalle de la tabla de Dimensión persona	26
Tabla 17. Categorías de la Dimensión persona	26
Tabla 18. Detalle de la tabla de Dimensión persona – discapacidad	27
Tabla 19. Categorías de la Dimensión persona – discapacidad	27
Tabla 20. Detalle de la tabla de Dimensión persona – hogar.....	27
Tabla 21. Categorías de la Dimensión persona – hogar	27
Tabla 22. Detalle de la tabla de Dimensiones persona – trabajo	28
Tabla 23. Categorías de la Dimensión persona – trabajo	28
Tabla 24. Detalle de la tabla de Dimensión localidad.....	28
Tabla 25. Categorías de la Dimensión localidad	29
Tabla 26. Detalle de la tabla de Dimensión tiempo	29
Tabla 27. Categorías de la Dimensión tiempo.....	29

Tabla 28. Detalle de la tabla de Dimensión vehículo.....	29
Tabla 29. Categorías de la Dimensión vehículo.....	29
Tabla 30. Detalle de la tabla de Dimensión unidad autorizadora	30
Tabla 31. Categorías de la Dimensión unidad autorizada.....	30
Tabla 32. Detalle de la tabla de dimensión ayudas	30
Tabla 33. Categorías de la dimensión ayudas	30
Tabla 34. Detalle de la tabla de Dimensión estado ayuda	31
Tabla 35. Categorías de la Dimensión estado ayuda.....	31
Tabla 36. Detalle de la tabla de Dimensión entidad subvenciona.....	31
Tabla 37. Categorías de la dimensión entidad subvenciona	31
Tabla 38. Detalle de la tabla de dimensión entidad provee.....	31
Tabla 39. Categorías de la Dimensión entidad provee.....	32
Tabla 40. Detalle de la tabla de Dimensión proyecto	32
Tabla 41. Categorías de la Dimensión proyecto.....	32
Tabla 42. Identificación de fuentes de datos para las tablas de Data mart.....	33
Tabla 43. Mapeo detallado con la fuente de datos.....	34
Tabla 44. Estándar para los nombres de las tablas de la bdd.....	41
Tabla 45. Nombres de las tablas de la bdd del Data mart.....	41
Tabla 46. Estándar para los nombres de las columnas de la bdd.....	41
Tabla 47. Ejemplos de nombres para las columnas de la bdd del Data mart	42
Tabla 48. Estándar para los nombres de los índices de la bdd.....	42
Tabla 49. Estándar para los nombres de los pasos del ETL	42
Tabla 50. Estándar para los nombres de los archivos del ETL	43
Tabla 51. Resumen de índices para cada Tabla de hechos y Dimensiones.....	48
Tabla 52. Transformaciones y Cálculos en Dimensiones.....	57
Tabla 53. Transformaciones y Cálculos en Tablas de hechos	60
Tabla 54. Conexiones a base de datos	76
Tabla 55. Reporte plantilla - Distribución de PCD por deficiencia principal.....	82
Tabla 56. Reporte plantilla - Porcentaje de PCD por genero	82

Tabla 57. Reporte plantilla - Distribución local de PCD por situación laboral.....	82
Tabla 58. Reporte plantilla - Causa de deficiencia mas frecuente por edades	83
Tabla 59. Reporte plantilla - PCD registradas por año.....	83
Tabla 60. Reporte plantilla - Distribución local de ayudas técnicas.....	83
Tabla 61. Reporte plantilla - Monto total en ayudas entregadas al año	84
Tabla 62. Reporte plantilla - Vehículos autorizados al año	84
Tabla 63. Reporte plantilla - Tiempo promedio para autorizar un vehículo	84
Tabla 64. Especificación del caso de uso: Realizar consultar interactiva	89
Tabla 65. Especificación del caso de uso: Cambiar la distribución del mapa	89
Tabla 66. Especificación del caso de uso: Ver reporte.....	90
Tabla 67. Capas de la aplicación - Mapsic.....	90
Tabla 68. Procedimiento de prueba para la Carga inicial.....	97
Tabla 69. Caso de prueba 1 - Carga Inicial.....	97
Tabla 70. Caso de prueba 2 - Carga Inicial.....	98
Tabla 71. Procedimiento de prueba para la Carga inicial.....	98
Tabla 72. Caso de prueba 1 - Carga Incremental	98
Tabla 73. Procedimiento de prueba para la Actualización	99
Tabla 74. Caso de prueba 2 – Actualización.....	99
Tabla 75. Caso de prueba 2 – Actualización.....	99
Tabla 76. Caso de prueba 3 – Actualización.....	100
Tabla 77. Procedimiento de prueba para la validación de dependencias	101
Tabla 78. Caso de prueba – Validación de dependencias	101
Tabla 79. Procedimiento de prueba para verificación de medidas.....	102
Tabla 80. Caso de prueba para la medida: Persona, agregación: sumatoria	103
Tabla 81. Caso de prueba para la medida: Vehículos, agregación: sumatoria...	103
Tabla 82. Caso de prueba para la medida: Ayudas, agregación: sumatoria.....	104
Tabla 83. Caso de prueba para las medidas: Monto, Subvención, agregación: sumatoria.....	105
Tabla 84. Caso de prueba para las medidas de tiempo, agregación: promedio .	106

Tabla 85. Caso de prueba para la medida Porcentaje	107
Tabla 86. Procedimiento de prueba para los Reportes de la aplicación Mapsic.	108
Tabla 87. Caso de prueba 1 - Reportes de la aplicación Mapsic	108
Tabla 88. Caso de prueba 2 - Reportes de la aplicación Mapsic	109
Tabla 89. Procedimiento de prueba para la distribución de datos en el mapa....	109
Tabla 90. Caso de prueba para la distribución de datos en el mapa	110

RESUMEN

Este documento explica todo el proceso de desarrollo del Data mart para el sistema de servicios sociales del CONADIS.

El primer capítulo explica la planificación del proyecto y la metodología bajo la cual se realizó el desarrollo, la etapa de recolección de requerimientos y análisis. Se plantea la arquitectura del Data mart y se seleccionan los productos de software mediante los cuales se implementó la solución

También se realizó el diseño lógico de dimensiones y tablas de hechos.

En el segundo capítulo se definió los estándares bajo los cuales se realizó el diseño físico del Data mart, se explica la construcción de los cubos multidimensionales y todo el proceso de extracción, transformación y carga de datos desde el sistema fuente hacia el Data mart; detallando para cada dimensión y tabla de hechos el respectivo proceso.

Se documentó el desarrollo de las aplicaciones de usuario, es decir la creación de reportes y la construcción de la aplicación Mapsic.

Finalmente se contemplan las pruebas realizadas sobre los procesos ETL, sobre los reportes y sobre la aplicación Mapsic; para su posterior despliegue.

El tercer capítulo contiene las conclusiones y recomendaciones de este trabajo.

1. CAPÍTULO 1 : PLANIFICACIÓN, ANÁLISIS Y DISEÑO LÓGICO DEL DATA MART

1.1. PLANIFICACIÓN DEL PROYECTO

1.1.1. DEFINICIÓN DEL ALCANCE

El Data mart apoya los procesos de la Dirección técnica y del Centro de información del CONADIS¹, procesos relacionados al sistema web SIC², tales como: registro nacional de discapacidades, solicitud y entrega de ayudas técnicas, solicitud y entrega de medicamentos o insumos, solicitud y autorización de vehículos para personas discapacitadas.

El desarrollo se realizó siguiendo una serie de etapas que cubren desde la planeación hasta el despliegue del producto terminado, en la fase de planeación se elaboró un plan inicial, el cual contempla brevemente objetivos, alcance; la fase de requerimientos permitió detectar las necesidades de información que tiene el CONADIS, una vez recopilada toda la información se procedió al análisis y posterior diseño de la solución convirtiendo los requerimientos en especificaciones, reconocimiento de fuentes de datos e indicadores claves.

La construcción del Data mart y procesos ETL se llevó a cabo en herramientas de libre difusión, los productos fueron expuestos a un ambiente de pruebas para detectar errores y asegurar la calidad, para su posterior implantación.

El principal medio de acceso a los datos es a través de un sistema de reportes web, además, una aplicación GIS permite presentar la información que entregue el data mart en un mapa para apoyar la presentación de estadísticas del sitio web.

1.1.2. JUSTIFICACIÓN DEL PROYECTO

El sistema permite realizar consultas interactivas y distribuir en el mapa indicadores tales como, el número de personas con discapacidades, de tal forma que se puedan realizar programas convenientes de apoyo social de acuerdo a diversos factores.

¹ Consejo Nacional de Discapacidades

² Sistema de Información CONADIS

La EPN y el CONADIS tienen un convenio de cooperación para el desarrollo de proyectos de apoyo al área de discapacidades.

1.1.3. OBJETIVOS

1.1.3.1. Objetivo general

Desarrollar un Data mart que apoye de manera eficiente la toma de decisiones en el área de servicios sociales del CONADIS.

1.1.3.2. Objetivos específicos

- Integrar el Data mart con el Sistema de servicios sociales actual del CONADIS, el cual es la fuente de información histórica.
- Ofrecer opciones para consultas interactivas.
- Utilizar una plataforma de libre difusión con la finalidad de reducir costo y mantener el lineamiento de la organización.

1.1.4. ETAPAS DEL PROYECTO¹

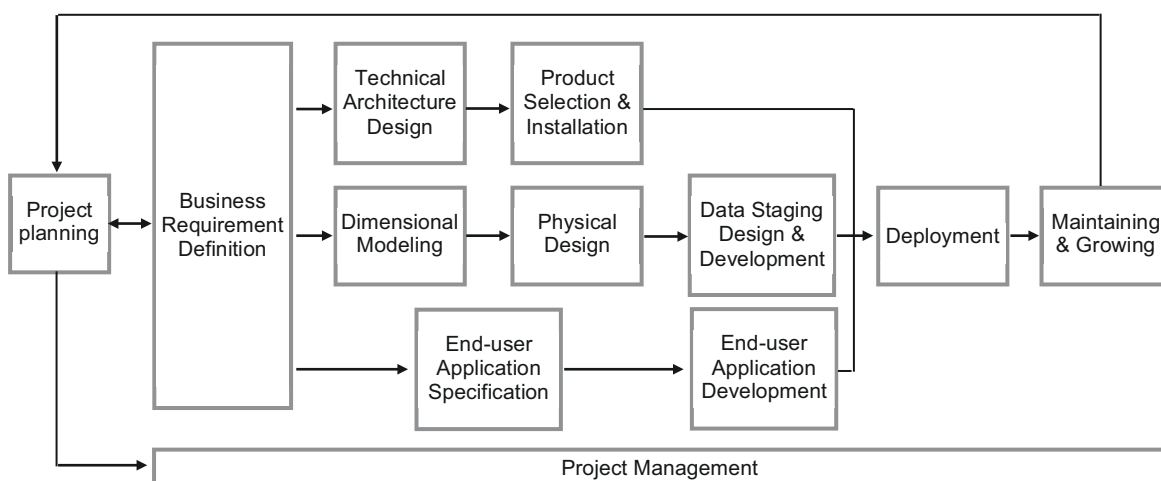


Figura 1. Ciclo de vida del Data warehouse propuesto por Kimball

La Figura 1, muestra el ciclo de vida para Data Warehouse propuesto por Ralf Kimball. Corresponde a una metodología Bottom-up.

¹ The Data Warehouse Lifecycle Toolkit, Chapter 2 – The Business Dimensional Lifecycle

El ciclo de vida de Kimball ilustra el flujo general para la implementación de un Data warehouse, identifica tareas secuenciales y actividades que ocurren en paralelo, es flexible, no necesariamente se debe realizar todas las tareas en detalle para cada proyecto.

Esta metodología se utilizó para llevar a cabo el desarrollo del presente Data mart.

1.1.5. RESUMEN DEL CRONOGRAMA DE ACTIVIDADES

Estimación de los tiempos empleados en cada etapa del ciclo de vida propuesta para el Data mart.

Tabla 1. Cronograma de actividades			
Actividad	Fecha Inicio	Fecha Fin	Tiempo (días)
Planificación del Proyecto	Abril 20	Mayo 01	10
Recolección de Información	Mayo 04	Junio 05	25
Diseño Arquitectónico y Selección de Productos	Mayo 18	Junio 01	11
Modelado Dimensional	Mayo 25	Junio 15	16
Diseño Físico	Junio 16	Julio 14	21
Diseño y Desarrollo ETL	Julio 15	Septiembre 04	38
Construcción de la Aplicación de Usuario Final	Mayo 28	Agosto 05	49
Despliegue	Septiembre 07	Septiembre 18	10

1.2. DEFINICIÓN DE REQUERIMIENTOS¹

Esta etapa define los requerimientos que soportan todas las demás etapas del ciclo de vida del Data mart, determina requisitos para la arquitectura y selección de productos, determina requerimientos de análisis e información para el modelado dimensional.

También determina especificaciones para las aplicaciones de usuario final.

1.2.1. TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN

Para llevar acabo la definición de requerimientos se recolectó información mediante las siguientes técnicas:

- Entrevistas al personal de la organización
- Análisis de documentos y sistemas fuente

1.2.1.1. Entrevistas al personal de la organización²

Se efectuó una entrevista al jefe del área de sistemas, el propósito de esta entrevista es conocer el entorno de hardware y software sobre el cual se encuentran implantados los sistemas operacionales de la organización, saber sobre posibles cuellos de botella y entender como es el proceso de difusión de la información, además, conocer cuales son los reportes y peticiones de información mas comúnmente solicitados.

También se efectuó una entrevista al Director del CONADIS, con el propósito de conocer los objetivos y funciones de la organización y sobre todo las necesidades de información.

¹ The Data Warehouse Lifecycle Toolkit, Chapter 4 – Collecting the Requirements

² Preguntas de la entrevista, ver Anexo A1

1.2.1.2. Resultados de las entrevistas

- Información organizacional del CONADIS, ver Anexo A2.
- La organización se maneja sobre un entorno de software libre, por lo tanto se mantuvo este principio al momento de realizar la selección de productos.
- La organización cuenta con el sistema web SIC¹, la información generada por este sistema fue analizada posteriormente como fuente de alimentación para el Data mart.
- La difusión de información se la realiza mediante el sitio web del CONADIS, cuentan con una sección de estadísticas, se tuvo esta información en cuenta al momento de desarrollar las aplicaciones de usuario final.

1.2.1.3. Análisis de documentos y sistemas fuente

Se analizó una serie de consultas SQL² que normalmente se utilizan para obtener información de la base de datos del sistema web SIC, estas consultas generan reportes totalizados de personas discapacitadas, ayudas técnicas, medicamentos.

En base a estas consultas, se detectaron como indicadores claves³:

- Número de personas discapacitadas
- Cantidad de ayudas técnicas solicitadas y entregadas
- Cantidad de medicamentos solicitados y entregados
- Cantidad de vehículos autorizados

Se detecto además datos descriptivos bajo los cuales se categoriza y totaliza la información, tales como: provincia, cantón, género, discapacidad, edad, tipo de ayuda técnica, tipo de medicamento, entre otros.

¹ Información sobre el sistema web SIC, ver Anexo A3

² Structured query language, ver Glosario

³ KPI, ver Glosario

Se analizó la base de datos del sistema web SIC para identificar el origen de la información anteriormente descrita y buscar otros posibles indicadores y datos descriptivos, como resultado se obtuvo los siguientes indicadores adicionales:

- Porcentaje de discapacidad de una persona
- Edad de la discapacidad de una persona
- Monto y subvención de ayudas técnicas
- Monto y subvención de medicamentos
- Cantidad de tiempo que lleva autorizar un vehículo

La Tabla 2 resume los indicadores claves asociados a los procesos del sistema web SIC.

Tabla 2. Procesos del sistema web SIC asociados a los indicadores	
Proceso: Registro Nacional de Discapacidades	
Indicador	Número de personas discapacitadas Porcentaje de la discapacidad Edad de la discapacidad
Categorías	Provincia, cantón, parroquia, discapacidad, instrucción, trabaja, genero, edad, grado de discapacidad, causa de la deficiencia.
Proceso: Solicitud y entrega de Ayudas Técnicas	
Indicador	Cantidad de ayudas técnicas solicitadas y entregadas Monto y subvención de ayudas técnicas
Categorías	Provincia, cantón, parroquia, discapacidad, instrucción, trabaja, genero, edad, grado de discapacidad, causa de la deficiencia, tipo de ayuda técnica, entidad que provee, entidad que subvenciona, proyecto.
Proceso: Solicitud y entrega de Medicamentos e insumos	
Indicador	Número de medicamentos solicitados y entregados Monto y subvención de medicamentos
Categorías	Provincia, cantón, parroquia, discapacidad, instrucción, trabaja, genero, edad, grado de discapacidad, causa de la deficiencia, tipo de medicamento, entidad que provee, entidad que subvenciona, proyecto.
Proceso: Solicitud y autorización de vehículos	
Indicador	Número de vehículos autorizados Cantidad de Tiempo que lleva autorizar un vehículo
Categorías	Provincia, cantón, parroquia, discapacidad, instrucción, trabaja, genero, edad, grado de discapacidad, causa de la deficiencia, unidad autorizadora, tipo de vehículo.

Se revisó los reportes presentes en la sección de estadísticas del sitio web¹ del CONADIS, estos reportes se tuvieron en cuenta como plantillas para la elaboración de reportes en las aplicaciones de usuario final.

1.2.2. PREPARACIÓN Y PUBLICACIÓN DE REQUERIMIENTOS

En base a la información recopilada y analizada anteriormente se elaboró los requerimientos de análisis e información.

1.2.2.1. Requerimientos de análisis

Información cuantitativa que se desea analizar.

- Conocer cual es el número de personas discapacitadas distribuidas por provincia, cantón, parroquia.
- Conocer cual es el número de personas discapacitadas distribuidas por provincia y cantón en base a factores como: edad, género, discapacidad, causa de la deficiencia, instrucción académica.
- Conocer cuales son las deficiencias y causantes mas frecuentes dependiendo de la edad de la persona.
- Conocer como esta la situación laboral, y familiar de las personas discapacitadas en función del tipo de discapacidad.
- Conocer como esta la situación académica en niños, adolescentes y jóvenes.
- Conocer cual es el porcentaje promedio de discapacidad de las personas de acuerdo a la: discapacidad, grado de discapacidad, causa de la discapacidad.
- Conocer cuales son las cantidades totales de ayudas técnicas entregadas por provincia, cantón y tipo de ayuda técnica.
- Conocer cuales son las cantidades totales de medicamentos entregadas por provincia, cantón y tipo de medicamento.

¹ <http://www.conadis.gov.ec/estadisticas.htm>

- Conocer cuales son los cantidades totales de subvenciones y montos en medicamentos y ayudas técnicas.
- Conocer cual es el tiempo promedio de atención de las solicitudes de vehículos para personas discapacitadas.

1.2.2.2. Requerimientos de información

Información descriptiva bajo la cual se categoriza y totaliza la información cuantitativa.

Personas discapacitadas

Datos personales: nombres, apellidos, edad, género, estado civil, con quien vive

Situación laboral y académica: nivel de instrucción académica, tipo de trabajo o razón por la cual no trabaja.

Residencia: provincia, cantón, parroquia, tipo de vivienda

Discapacidad: grado de discapacidad, causa de la discapacidad, deficiencia principal, porcentaje de discapacidad, edad de la discapacidad.

Ayudas técnicas

Tipos de ayudas técnicas, entidades que proveen, entidades que subvencionan, fechas de entrega, estado de la ayuda.

Medicamentos e insumos

Nombre del medicamento, fechas de entrega, entidades que proveen, entidades que subvencionan, estado.

Localidades

Nombres y códigos de provincias, cantones y parroquias.

Autorización de vehículos

Fechas de solicitud, fecha de autorización, tipo de vehículo, unidad autorizadora.

1.3. DISEÑO ARQUITECTÓNICO, SELECCIÓN DE PRODUCTOS

1.3.1. ARQUITECTURA DEL DATA MART¹

En esta etapa del ciclo de vida se busca consolidar tres factores:

- Los requerimientos de la organización.
- El ambiente actual de hardware y software de la organización
- Los componentes requeridos para la implementación del Data mart

Para esta tarea se consideró dos escenarios:

Back Room.- Encapsula todos los procedimientos y recursos necesarios para llevar a cabo el tratamiento de los datos, básicamente los procesos de ETL².

Front Room.- Encapsula todos los procedimientos y recursos necesarios para presentar los datos ya tratados al usuario final.

En medio de estos dos escenarios se ubica el servidor de presentación, y los metadatos.

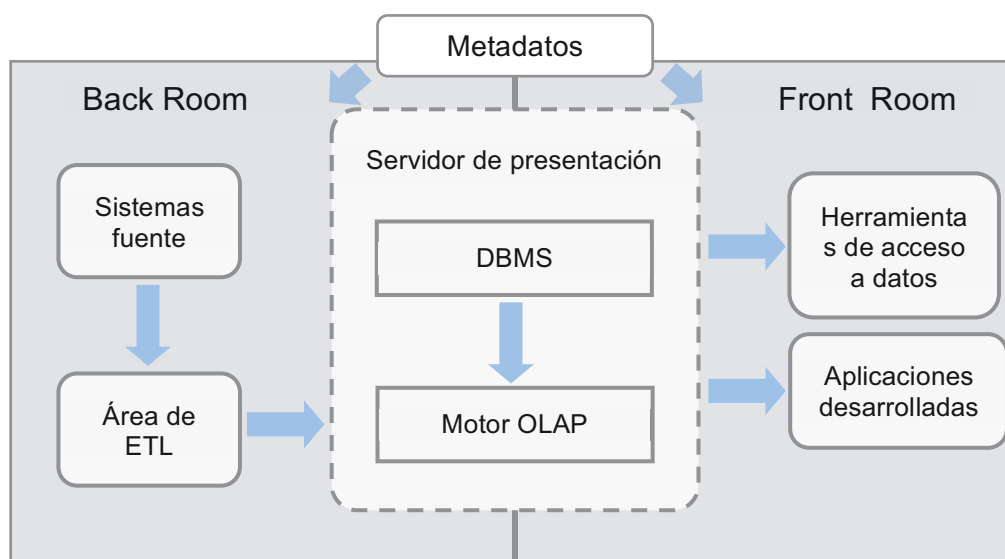


Figura 2. Arquitectura del Data mart

¹ The Data Warehouse Lifecycle Toolkit, chapter 8-10

² Extract, Transform and Load, ver Glosario

La arquitectura presenta dos tipos de componentes fundamentales, servicios y repositorios de datos.

Servicios.- Son las funciones necesarias para llevar a cabo las tareas requeridas en el Data mart.

Repositorios de datos.- Son lugares temporales o permanentes en los cuales se almacenan los datos.

A continuación se detalla los servicios y repositorios de datos tanto para el Back Room como para el Front Room.

1.3.1.1. Servicios del Back Room

Actividades empleadas en el proceso de extracción, transformación y carga (ETL) de los datos, los servicios son los siguientes:

Servicio de extracción de datos.- Es el encargado de obtener los datos de los sistemas fuente.

Servicio de transformación de datos.- Cambios que sufren los datos necesarios para el rendimiento del Data mart, entre estas actividades están: generación de claves subrogadas, mapeo de claves de un sistema a otro, denormalización¹ de tablas, conversión de los tipos de datos, cálculo de valores.

Servicio de carga de datos.- Es el encargado de almacenar los datos en los sistemas destino.

Servicio de control de tareas.- Ayuda a controlar los errores y a manejar las excepciones mediante la definición y programación de tareas, monitoreo y registro de eventos.

1.3.1.2. Repositorios de datos del Back Room

Lugares temporales o permanentes en los cuales se depositan los datos que se generan y se consumen en el Back Room.

¹ Definición, ver Glosario

Sistemas fuentes.- Fuente de datos que alimenta el Data mart, sobre este repositorio de datos ocurre el servicio de extracción de datos.

Tabla 3. Fuentes de datos para el Data mart	
Sistema web SIC	
DBMS	PostgreSQL
BDD	sic
Nro. De tablas	64

Área de ETL.- Repositorio temporal de datos, aquí tiene lugar el proceso de transformación de datos y el servicio de control de tareas, el cual consiste en la calendarización y monitoreo de los procesos ETL.

DBMS¹.- Repositorio permanente de los datos resultantes del proceso ETL, sobre este repositorio ocurre el servicio de carga de datos.

1.3.1.3. Servicios del Front Room

Actividades que deben ofrecer las Herramientas y sistemas que utilizan y presentan los datos almacenados en el Data mart.

Servicio de navegación sobre el Data mart.- El proceso de navegación a través de la información es uno de los procesos más importantes del Front Room, permite al usuario mediante una herramienta del sistema buscar algún dato que le sea importante o necesite saber.

Servicio de acceso y seguridad.- Estos servicios también pueden ser llamados de autorización y autenticación, y ambos se refieren a las conexiones que los usuarios necesitan para acceder a las bases de datos. La autenticación se refiere a alguna técnica para verificar que la persona es quien dice ser, este servicio puede ser controlado por otro sistema.

Servicio de administración de consultas.- La administración de consultas le permite al usuario una interacción con las bases de datos.

Servicio de reportes estándar.- El servicio de reportes estándar le permite al usuario generar reportes con formatos estándar.

¹ Database management system, ver Glosario

1.3.1.4. Repositorios de datos del Front Room

Motor OLAP¹.- Genera los cubos y dimensiones con la información almacenada en la base de datos del Data mart y procesa las consultas analíticas, responde al Servicio de administración de consultas MDX².

Herramientas de acceso a datos.- Herramientas que permiten a los usuarios interactuar con los datos almacenados en el Data mart, mediante consultas MDX y SQL, además deben proporcionar los servicios de navegación sobre el Data mart y servicio de reportes estándar.

Aplicaciones desarrolladas.- Aplicaciones desarrolladas por la organización, para este caso se planteó la construcción de una pequeña aplicación web que permite al usuario interactuar con los datos almacenados en el Data mart mediante consultas SQL, valiéndose de un componente GIS³ para representar la información de manera gráfica en un mapa.

1.3.2. METADATOS⁴

Los metadatos consisten en los datos acerca de los datos, un concepto algo difuso, por lo tanto actualmente se habla acerca de los metadatos del Back Room, y los metadatos del Front Room.

Los metadatos del Back Room están relacionados a los procesos ETL, guían la extracción, limpieza y carga de datos; por lo tanto entre estos metadatos esta la información acerca de los sistemas fuente y el área ETL.

Los metadatos del Front Room son mas descriptivos, ayudan a las herramientas de consulta, nos permite entender de donde vienen los datos; estos metadatos cubren información acerca del Data mart y herramientas de acceso a datos.

La mayor parte de estos metadatos se generan en la etapa correspondiente del ciclo de vida del Data mart, a continuación se describe el tipo de información que a de contemplarse en cada etapa para los metadatos.

¹ On-Line Analytical Processing, ver Glosario

² Multidimensional Expressions, ver Glosario

³ geographic information system, ver Glosario

⁴ The Data Warehouse Lifecycle Toolkit, chapter 11 – Infrastructure and Metadata

1.3.2.1. Metadatos del sistema fuente

Especificaciones de la fuente de datos.- Información de la base de datos del sistema web SIC.

Tabla 4. Especificaciones de la fuente de datos		
Plataforma		PostgreSQL
Nombre		sic
Nro. de tablas		64
Owner	usuario	sic
	password	sic
	privilegios	nosuperuser, createdb, createrole

Diccionario de Datos.- Descripción de tablas y columnas de la bdd, nombre de columnas, tipo de dato, descripción.

1.3.2.2. Metadatos del área de ETL

Políticas de cambio para Dimensiones¹: Determina la forma en como se administra los cambios que ocurran en la Dimensiones.

Tabla 5. Políticas de cambio para Dimensiones	
Tipo de cambio	Descripción
Tipo 0	Sin cambio
Tipo 1	Sobre escritura
Tipo 2	Creación de un nuevo registro en la dimensión

Conexiones a base de datos: Detalla el nombre de la conexión, descripción, datos de la conexión (nombre del servidor o dirección IP, DBMS, nombre de la base de datos, usuario, clave).

Transformaciones de datos: Identificar nombre de la columna sometida a transformaciones y cálculos, la tabla a la que pertenece, descripción de la transformación a llevarse a cabo.

Flujo de Datos de la fuente al destino: Diagramas para los flujos de datos, información detallada para cada paso del flujo de datos (nombre del paso, objeto provisto por la herramienta ETL que realiza la acción, descripción).

¹ The Data Warehouse Toolkit, chapter 4 – Slowly changing Dimensions

Archivos que intervienen en el proceso ETL: Archivos generados por la herramienta ETL que representan físicamente los flujos de datos.

1.3.2.3. Metadatos del Data mart y herramientas de acceso

Especificaciones de la bdd del Data mart: Información de la base de datos del Data mart.

Tabla 6. Especificaciones de la bdd del Data mart		
Plataforma		PostgreSQL
Nombre		sic
Nro. de tablas		13
Owner	usuario	sic_analisis
	password	Sicanalsisis
	privilegios	nosuperuser, createdb, createrole

Definición de dimensiones y tablas de hechos: Información detallada de los atributos de las dimensiones (nombre, descripción, política de cambio, ejemplos). Información detallada de las tablas de hechos (claves dimensionales, medidas, formula para calcular la medida, agregaciones).

Índices: Índices por tabla, nombre del índice, tipo de índice, columnas que conforman el índice.

Diccionario de Datos: Descripción de tablas y columnas de la bdd del Data mart, nombre de columnas, tipo de dato, descripción.

Capacitación a usuarios: Manual de despliegue, manual de usuario.

1.3.3. SELECCIÓN DE PRODUCTOS¹

En esta etapa del ciclo de vida, se busca elegir herramientas y aplicaciones de software apropiadas para la implementación de la arquitectura propuesta.

Los productos seleccionados son de software libre, para seguir los lineamientos de la organización y además satisfacen los servicios tanto para el Back Room como para el Front Room.

¹ The Data Warehouse Lifecycle Toolkit, chapter 13 – Selecting Products

1.3.3.1. Productos para el Back Room

Producto seleccionado para la DBMS

Actualmente la organización utiliza PostgreSQL para la persistencia de datos de sus sistemas, por lo tanto el repositorio del Data mart se mantuvo bajo esta plataforma.

Tabla 7. Producto seleccionado para la DBMS	
Nombre	PostgreSQL
Versión	8.4

Producto seleccionado para el área ETL

Para llevar a cabo los servicios primordiales del Back Room se eligió como herramienta ETL a: Pentaho Data Integration - kettle.

La herramienta permite crear flujos de datos denominados Jobs y Transformaciones, cuenta con una gran variedad de componentes para llevar a cabo actividades de extracción, transformación, cálculos y carga de datos entre una variedad de bases de datos y archivos estructurados.

Además, por su larga trayectoria como herramienta ETL de software libre y entorno intuitivo y fácil de usar.

Tabla 8. Producto seleccionado para el área ETL	
Nombre	Pentaho data integration - Kettle
Versión	3.2

1.3.3.2. Productos para el Front Room

Producto seleccionado para el motor OLAP

Como motor de análisis se eligió a Mondrian¹, actualmente soportado por la comunidad de Pentaho; y base de la solución comercial Pentaho BI.

Escrito en java, es un motor completamente multiplataforma, resuelve consultas MDX, compatible con almacenamiento multidimensional y relacional.

¹ <http://mondrian.pentaho.com/>

La definición de cubos, dimensiones, agregaciones, permisos, roles, entre otros objetos, los maneja mediante un esquema de etiquetas XML muy bien definido.

Además, por su naturaleza de software libre y estabilidad como motor de análisis y gran soporte comunitario.

Tabla 9. Producto seleccionado como motor OLAP	
Nombre	Mondrian
Versión	3.1.6

Producto seleccionado como herramienta de acceso a datos

Para la generación y consulta de reportes se eligió a Openi¹, aplicación web 2.0, compatible con Mondrian y que tolera acceso a datos mediante XMLA².

Satisface todos los servicios del Front Room, permitiendo la administración de usuarios, creación de reportes y gráficos, administración de fuentes de datos OLAP y relacionales, permite estructurar consultas utilizando sintaxis MDX y SQL.

Posee una interfaz fácil de usar sobre todo al momento de armar reportes, facilita la asignación de columnas, filas y filtros, los reportes se presentan en tablas pivot, por lo tanto permiten realizar operaciones de drilldown y ordenamientos.

Tabla 10. Producto seleccionado como herramienta de acceso a datos	
Nombre	Openi
Versión	2.0 – RC2

Tabla 11. Otros servidores	
Servidor de aplicaciones java	
Plataforma	Tomcat
Versión	6.0
Servidor web php	
Plataforma	Apache
Versión	2.2
Servidor gis	
Plataforma	Mapserver
Versión	3.0

¹ <http://openi.org/>

² XML for Analysis, ver Glosario

1.4. MODELADO DIMENSIONAL¹

Disciplina específica para el modelado de datos que es una alternativa al modelo entidad – relación. El modelo dimensional contiene la misma información que un modelo de entidad – relación, pero empaqueta los datos en un formato simétrico cuyos objetivos son, comprensión por parte de los usuarios, rendimiento de las consultas y resistencia al cambio.

Cada modelo dimensional está compuesto por una tabla cuya clave es compuesta, llamada tabla de hechos, y un grupo de tablas pequeñas llamadas tablas de dimensiones.

Cada tabla de dimensión tiene una clave primaria singular que corresponde exactamente con uno de los componentes de la clave compuesta de la tabla de hechos, esta estructura adopta el nombre de modelo en estrella.

Tabla de hechos.- Esta tabla dado que presenta una clave compuesta por dos o más claves foráneas, expresa una relación de varios a varios, las tablas más útiles son aquellas que contienen una o más medidas, valores numéricos y aditivos que resultan de la combinación única de claves.

Tabla de dimensiones.- Estas tablas en contraste con las anteriores contienen información textual descriptiva, estos atributos son los filtros en las consultas realizadas sobre el Data mart y generalmente conforman las cabeceras de las filas en los reportes.

1.4.1. CONFORMACIÓN DE DIMENSIONES²

En base a los requerimientos de información y a la información descriptiva de las consultas SQL analizadas se definió las siguientes agrupaciones de atributos descriptivos como dimensiones:

Dimensión persona.- Información descriptiva de las personas discapacitadas, datos como: nombres, apellidos, estado civil, edad, estado del registro.

¹ The Data Warehouse Lifecycle Toolkit, chapter 5 – Dimensional Modeling

² The Data Warehouse Toolkit, chapter 6

Dimensión persona – discapacidad.- Información descriptiva concerniente a la discapacidad de las personas, datos como: deficiencias principales, grados de discapacidad, causa de las deficiencias.

Dimensión persona – hogar.- Información descriptiva sobre la situación familiar y de vivienda de la persona, datos como: con quien vive, tipo de vivienda.

Dimensión persona – trabajo.- Información descriptiva sobre la situación laboral e instrucción académica de la persona, datos como: nivel de instrucción, si trabaja o no, tipo de trabajo, razón por la cual no trabaja.

Dimensión localidad.- Información descriptiva sobre la ubicación de la persona, datos como: provincia, cantón, parroquia.

Dimensión Tiempo.- Agrupa años, meses y días, relacionados a fechas como: fecha de carnetización de la persona, fecha de solicitud de ayudas técnicas y medicamentos, fechas relacionadas a la autorización de vehículos.

Dimensión vehículo.- Información descriptiva sobre los vehículos para personas discapacitadas, datos como: tipo de vehículo.

Dimensión unidad autorizadora.- Información descriptiva sobre las unidades autorizadoras, las cuales intervienen en el proceso de autorización de vehículos, datos como: nombre, ubicación, director.

Dimensión ayudas.- Información sobre ayudas técnicas y medicamentos.

Dimensión estado ayuda.- Información sobre los estados de la solicitud de una ayuda técnica o un medicamento.

Dimensión entidad subvenciona.- Información descriptiva sobre las entidades que subvencionan ayudas técnicas y medicamentos, datos como: nombre de la entidad, siglas.

Dimensión entidad provee.- Información descriptiva sobre las entidades que proveen ayudas técnicas y medicamentos, datos como: nombre de la entidad, siglas.

Dimensión proyecto.- Información descriptiva sobre proyectos que intervienen en los procesos de ayudas técnicas y medicamentos, datos como: nombre del proyecto, tipo de proyecto, responsable, ubicación.

1.4.2. CONFORMACIÓN DE LAS TABLAS DE HECHOS

La primera tarea en realizarse fue definir las tablas de hechos agrupando las medidas asociadas a los indicadores claves detectados en la fase de análisis, en la Tabla 2 se agrupó dichos indicadores por el proceso asociado al sistema web SIC, conforme a esto se determinó las siguientes tablas de hechos:

- Tabla de hechos de personas discapacitadas
- Tabla de hechos de ayudas (agrupa ayudas técnicas y medicamentos)
- Tabla de hechos de autorización de vehículos

Para realizar el diseño de estas tablas de hechos se dividió la tarea en cuatro pasos y son:

Selección del Data mart

Generalmente este paso se lo lleva acabo cuando se establece una solución de Data warehouse en el cual existen múltiples Data mart, sin embargo para este caso existe un único Data mart y sobre este residirán las tablas de hechos a conformar.

Selección de las dimensiones

La Tabla 12 muestra el cruce entre dimensiones y tablas de hechos, la selección de dimensiones para cada tabla de hechos se realizó en función de los datos descriptivos anotados en la Tabla 2 para cada proceso del sistema web SIC.

Tabla 12. Matriz de Dimensiones y Tablas de hechos			
Tabla de Hechos	Personas discapacitadas	Ayudas	Autorización de vehículos
Dimensiones			
Persona	X	X	X
Persona - Discapacidad	X	X	X
Persona - Hogar	X	X	X
Persona - Trabajo	X	X	X
Localidad	X	X	X
Tiempo	X	X	X
Vehículo		X	
Unidad autorizadora		X	
Ayudas			X
Estado - Ayuda			X
Entidad subvenciona			X
Entidad provee			X
Proyecto			X

Determinación de la granularidad

Se identificó el nivel mas atómico de los procesos en el cual se puede apreciar los indicadores y se puede cuantificar su valor.

Tabla de hechos de personas discapacitadas: Cada registró de persona discapacitada del sistema web SIC es un registro de la tabla de hechos.

Tabla de hechos para ayudas: Cada solicitud de ayuda (medicamento o ayuda técnica) del sistema web SIC es un registro de la tabla de hechos.

Tabla de hechos de autorización de vehículos: Cada solicitud de vehículo con su respectiva autorización del sistema web SIC es un registro de la tabla de hechos.

Selección de medidas

La selección de medidas se la realizó en base a los indicadores claves y a la granularidad determinada para cada tabla de hechos.

Tabla de hechos de personas discapacitadas: Captura el evento de registro de personas discapacitadas, las medidas para esta tabla son: el número de personas discapacitadas, el porcentaje de discapacidad y edad de la discapacidad.

Tabla de hechos de ayudas: Captura el evento de solicitar ayudas técnicas y medicamentos; las medidas para esta tabla son: el número de ayudas técnicas y medicamentos, el monto y la subvención correspondiente a la solicitud registrada.

Tabla de hechos de autorización de vehículos: Captura el evento de autorización de vehículos; las medidas para esta tabla son: el número de vehículos autorizados, los intervalos de tiempo entre las fechas que se registran en el proceso de autorización de vehículos y un tiempo total de atención.

1.4.3. DIAGRAMAS DE TABLAS DE HECHOS

1.4.3.1. Diagrama de la tabla de hechos de personas discapacitadas

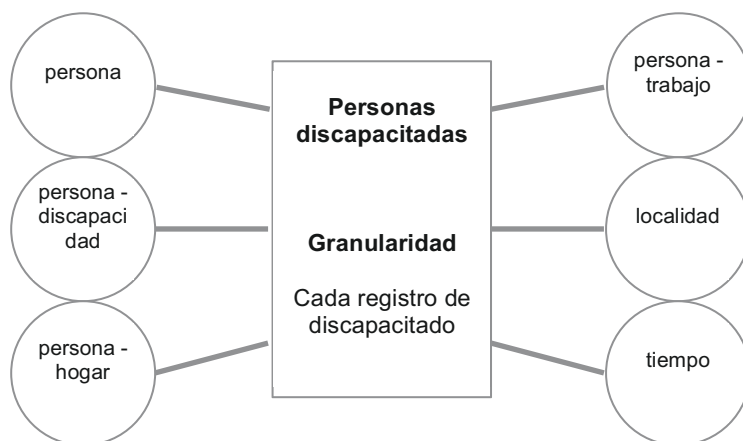


Figura 3. Diagrama de la tabla de hechos de personas discapacitadas

1.4.3.2. Diagrama de la tabla de hechos de autorización de vehículos

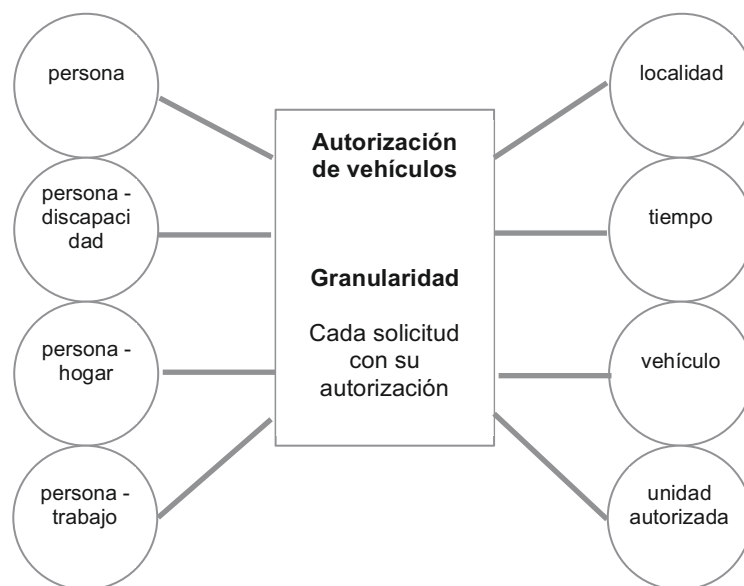


Figura 4. Diagrama de la tabla de hechos de autorización de vehículos

1.4.3.3. Diagrama de la tabla de hechos de ayudas

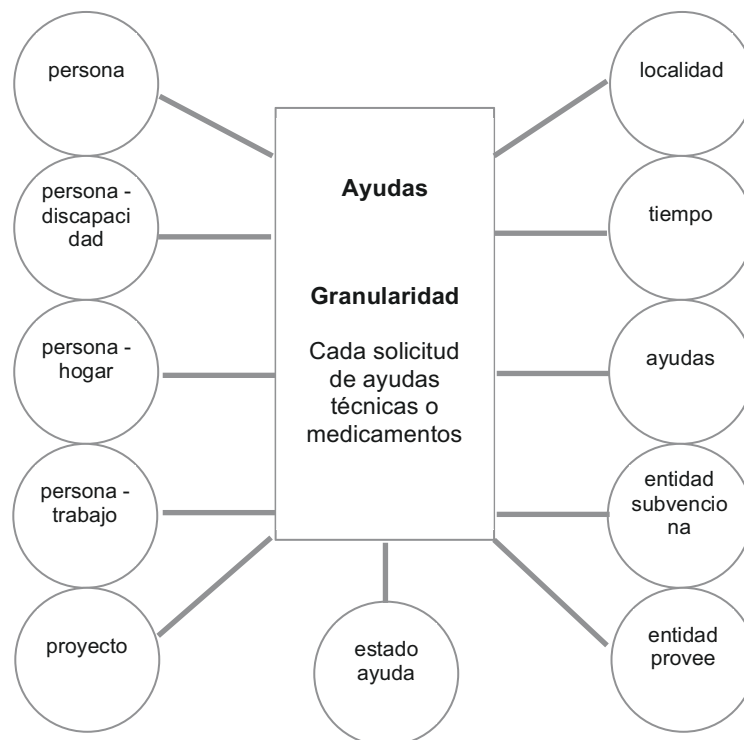


Figura 5. Diagrama de la tabla de hechos de ayudas

1.4.4. DETALLE DE TABLAS DE HECHOS

El detalle de las tablas de hechos contiene información sobre las claves dimensionales, es decir las claves subrogadas¹ de las dimensiones asociadas a la tabla de hechos, información sobre las medidas, el tipo de medida, la regla de agregación y la formula mediante la cual se obtiene.

Los tipos de medidas son los siguientes:

Columna.- La medida esta contenida en una columna de la fuente de datos, y se pasa a la tabla de hechos sin mayor inconveniente.

Transformación.- Para obtener la medida debe haber previamente un tratamiento de los datos.

Calculo.- Medida calculada en base a otras medidas previamente definidas

1.4.4.1. Tabla de hechos de personas discapacitadas

Tabla 13. Detalle de la Tabla de hechos de personas discapacitadas		
Claves	Descripción	
Key: persona	Clave de la dimensión persona	
Key: persona - discapacidad	Clave de la dimensión persona - discapacidad	
Key: persona - hogar	Clave de la dimensión persona - hogar	
Key: persona - trabajo	Clave de la dimensión persona - trabajo	
Key: localidad	Clave de la dimensión localidad	
Key: tiempo - fecha carnetización	Clave de la dimensión tiempo con respecto a las fecha de carnetización	
Medidas	Descripción	
Fact: Personas	Número de personas discapacitadas en base a la granularidad de las dimensiones.	
	Tipo	Transformación
	Agregación	Suma
Fact: Porcentaje discapacidad	Porcentaje de discapacidad que presenta la persona.	
	Tipo	Columna
	Agregación	Promedio
Fact: Edad discapacidad	Edad a la que la persona adquirió la discapacidad	
	Tipo	Columna
	Agregación	Promedio

¹ Definición, ver Glosario

1.4.4.2. Tabla de hechos de autorización de vehículos

Tabla 14. Detalle de la Tabla de hechos de autorización de vehículos		
Claves	Descripción	
Key: persona	Clave de la dimensión persona	
Key: persona - discapacidad	Clave de la dimensión persona - discapacidad	
Key: persona - hogar	Clave de la dimensión persona - hogar	
Key: persona - trabajo	Clave de la dimensión persona - trabajo	
Key: localidad	Clave de la dimensión localidad	
Key: vehículo	Clave de la dimensión vehículo	
Key: unidad autorizadora	Clave de la dimensión unidad autorizadora	
Key: tiempo - fecha solicitud	Clave de la dimensión tiempo con respecto a las fecha de solicitud de vehículo	
Key: tiempo - fecha recepción	Clave de la dimensión tiempo con respecto a las fecha de recepción de la solicitud	
Key: tiempo - fecha reunión	Clave de la dimensión tiempo con respecto a las fecha de reunión de la junta que autoriza	
Key: tiempo - fecha autorización	Clave de la dimensión tiempo con respecto a las fecha de autorización de la solicitud	
Medidas	Descripción	
Fact: Vehículos	Número de solicitudes de vehículos autorizadas en base a la granularidad de las dimensiones.	
	Tipo	Transformación
	Agregación	Suma
Fact: Recepción(días)	Número de días transcurridos entre la fecha de solicitud (FS) y la fecha de recepción (FR)	
	Tipo	Transformación
	Formula	TR = day(FR) – day(FS)
	Agregación	Promedio
Fact: Atención(días)	Número de días transcurridos entre la fecha de recepción (FR) y la fecha de reunión (FU)	
	Tipo	Transformación
	Formula	TA = day(FU) – day(FR)
	Agregación	Promedio
Fact: Resolución(días)	Número de días transcurridos entre la fecha de reunión (FU) y la fecha de autorización (FA)	
	Tipo	Transformación
	Formula	TU = day(FA) – day(FU)
	Agregación	Promedio
Fact: Total(días)	Número de días total transcurridos	
	Tipo	Calculo
	Formula	TT = TR + TA + TU
	Agregación	Promedio

1.4.4.3. Tabla de hechos de ayudas

Tabla 15. Detalle de la Tabla de hechos de ayudas		
Claves	Descripción	
Key: persona	Clave de la dimensión persona	
Key: persona - discapacidad	Clave de la dimensión persona - discapacidad	
Key: persona - hogar	Clave de la dimensión persona - hogar	
Key: persona - trabajo	Clave de la dimensión persona - trabajo	
Key: localidad	Clave de la dimensión localidad	
Key: ayudas	Clave de la dimensión ayudas	
Key: estado ayuda	Clave de la dimensión estado ayuda	
Key: entidad subvenciona	Clave de la dimensión entidad subvenciona	
Key: entidad provee	Clave de la dimensión entidad provee	
Key: proyecto	Clave de la dimensión proyecto	
Key: tiempo - fecha entrega	Clave de la dimensión tiempo con respecto a las fecha de entrega de la ayuda	
Medidas	Descripción	
Fact: Ayudas	Número de ayudas en base a la granularidad de las dimensiones.	
	Tipo	Transformación
	Agregación	Suma
Fact: Monto	Valor monetario del monto total de la ayuda	
	Tipo	Columna
	Agregación	Suma
Fact: Subvención	Valor monetario del monto subvencionado de la ayuda	
	Tipo	Columna
	Agregación	Suma

1.4.5. DETALLE DE LAS TABLAS DE DIMENSIONES

El detalle de las tablas de dimensiones contiene información sobre los atributos, descripción, política de cambio ver Tabla 5, y un ejemplo.

Los atributos pueden estar precedidos por los siguientes identificadores, los cuales determinan su función:

Key: El atributo es la clave subrogada de la dimensión.

Pk: El atributo es parte de las clave operacionales mediante las cuales se identifica como único cada registro de la dimensión y se asocia a una clave

subrogada también única, corresponden a claves primarias obtenidas de la bdd fuente.

Id: Son claves primarias obtenidas de la bdd fuente o claves generadas en el proceso ETL asociadas a atributos descriptivos, pero que no forman parte de las claves operacionales.

Los atributos sin ningún identificador son atributos descriptivos.

Además se presenta información sobre las categorías planteadas para la dimensión, es decir la relación jerárquica existente entre los atributos.

1.4.5.1. Dimensión persona

Tabla 16. Detalle de la tabla de Dimensión persona			
Atributo	Descripción	Cambio	Ejemplo
Key: persona	Clave subrogada de la dimensión	Tipo 0	1
Pk: persona	Clave operacional de la persona	Tipo 1	1
Apellido 1	Primer apellido de la persona	Tipo 1	Valdivieso
Apellido 2	Segundo apellido de la persona	Tipo 1	Jiménez
Nombre 1	Primer nombre de la persona	Tipo 1	Miriam
Nombre2	Segundo nombre de la persona	Tipo 1	Elizabeth
Id: genero	Identificador del atributo género	Tipo 1	F
Genero	Género de la persona	Tipo 1	Femenino
Id: estado civil	Identificador del atributo estado civil	Tipo 1	1
Estado civil	Estado civil de la persona	Tipo 1	Soltero
edad	Edad de la persona	Tipo 1	35
Id: edad etapa	Identificador del atributo edad etapa	Tipo 1	5
Edad etapa	Etapa de la vida a la que corresponde la edad de la persona	Tipo 1	Adulto
Estado registro	Estado del registro de la persona	Tipo 1	A

Tabla 17. Categorías de la Dimensión persona		
Categoría	Jerarquía	Ejemplo
Persona – Genero	genero	- F - M
Persona – Edad	edad etapa ↓ edad	- Infante (0-6) - 4 - Joven (18-25) - 19 - Adulto mayor (60) - 65
Persona – Estado civil	estado civil	- Casado - Divorciado
Persona – Estado	estado registro	- A

1.4.5.2. Dimensión persona – discapacidad

Tabla 18. Detalle de la tabla de Dimensión persona – discapacidad			
Atributo	Descripción	Cambio	Ejemplo
Key: persona discapacidad	Clave subrogada de la dimensión	Tipo 0	1
Pk: grado discapacidad	Clave operacional del grado de discapacidad	Tipo 1	1
Grado discapacidad	Grado de discapacidad general que presenta la persona	Tipo 1	Leve, Grave
Pk: deficiencia principal	Clave operacional de la deficiencia principal	Tipo 1	1
Deficiencia principal	Deficiencia principal que afecta a la persona	Tipo 1	Visual, Auditiva
Pk: causa deficiencia	Clave operacional de la causa de la deficiencia	Tipo 1	3
Causa deficiencia	Causa de la deficiencia	Tipo 1	Accidente laboral

Tabla 19. Categorías de la Dimensión persona – discapacidad		
Categoría	Jerarquía	Ejemplo
Deficiencia principal	deficiencia principal	- Auditiva - Física
Grado de discapacidad	grado discapacidad	- Leve - Grave
Causa de la deficiencia	causa deficiencia	- Enfermedad adquirida - Accidente de tránsito

1.4.5.3. Dimensión persona – hogar

Tabla 20. Detalle de la tabla de Dimensión persona – hogar			
Atributo	Descripción	Cambio	Ejemplo
Key: persona hogar	Clave subrogada de la dimensión	Tipo 0	1
Pk: con quien vive	Clave operacional de con quien vive	Tipo 1	2
Con quien vive	Con quien vive la persona	Tipo 1	Familia
Pk: tipo vivienda	Clave operacional del tipo de vivienda	Tipo 1	3
Tipo vivienda	Tipo de vivienda de la persona	Tipo 1	Arrendada

Tabla 21. Categorías de la Dimensión persona – hogar		
Categoría	Jerarquía	Ejemplo
Persona – Con quien vive	con quien vive	- Solo - Familia
Tipo de vivienda	tipo vivienda	- De la familia - Arrendada

1.4.5.4. Dimensión persona – trabajo

Tabla 22. Detalle de la tabla de Dimensiones persona – trabajo			
Atributo	Descripción	Cambio	Ejemplo
Key: persona trabajo	Clave subrogada de la dimensión	Tipo 0	1
Pk: nivel instrucción	Clave operacional del nivel de instrucción	Tipo 1	7
Nivel instrucción	Nivel de instrucción de la persona	Tipo 1	Superior
Trabaja	Estado laboral de la persona	Tipo 1	Si
Pk: trabaja descripción	Clave operacional de la actividad en la que labora la persona o de la razón por la que no trabaja	Tipo 1	1
Trabaja descripción	Actividad en la que labora la persona o la razón por la que no trabaja	Tipo 1	Por cuenta propia

Tabla 23. Categorías de la Dimensión persona – trabajo		
Categoría	Jerarquía	Ejemplo
Persona – Nivel de instrucción	nivel instrucción	- Alfabetizado - Secundaria
Trabajo	trabaja ↓ trabaja descripción	- Si - Sector privado - Por cuenta propia - No - Jubilado - Por su discapacidad

1.4.5.5. Dimensión localidad

Tabla 24. Detalle de la tabla de Dimensión localidad			
Atributo	Descripción	Cambio	Ejemplo
Key: localidad	Clave subrogada de la dimensión	Tipo 0	1
Pk: parroquia	Clave operacional de parroquia	Tipo 1	238
Pk: cantón	Clave operacional de cantón	Tipo 1	24
Pk: provincia	Clave operacional de provincia	Tipo 1	2
Región código	Código de la región	Tipo 1	01
Parroquia código	Código de la parroquia	Tipo 1	01-03-23
Cantón código	Código del cantón	Tipo 1	01-03
Provincia código	Código de la provincia	Tipo 1	01
Región	Región en la que vive la persona	Tipo 1	Sierra – Andina
Parroquia	Nombre de la parroquia	Tipo 1	Baños
Cantón	Nombre del cantón	Tipo 1	Cuenca
Provincia	Nombre de la provincia	Tipo 1	Azuay

Tabla 25. Categorías de la Dimensión localidad		
Categoría	Jerarquía	Ejemplo
Localidad	región ↓ provincia ↓ cantón ↓ parroquia	- Sierra – Andina - Pichincha - Quito - Calderón - Costa – Litoral - Guayas - Guayaquil - Rocafuerte

1.4.5.6. Dimensión tiempo

Tabla 26. Detalle de la tabla de Dimensión tiempo			
Atributo	Descripción	Cambio	Ejemplo
Key: tiempo	Clave subrogada de la dimensión	Tipo 0	1
Fecha completa	Fecha completa: año, mes, día	Tipo 0	1996-07-19
Día	Día en la fecha	Tipo 0	19
Mes	Mes en la fecha	Tipo 0	7
Mes nombre	Nombre del mes	Tipo 0	Julio
Año	Año en la fecha	Tipo 0	1996

Tabla 27. Categorías de la Dimensión tiempo		
Categoría	Jerarquía	Ejemplo
Tiempo	año ↓ mes nombre ↓ día	- 2010 - Abril - 17 - Junio - 11

1.4.5.7. Dimensión vehículo

Tabla 28. Detalle de la tabla de Dimensión vehículo			
Atributo	Descripción	Cambio	Ejemplo
Key: vehículo	Clave subrogada de la dimensión	Tipo 0	1
Pk: tipo vehículo	Clave operacional del tipo de vehículo	Tipo 1	1
Tipo vehículo	Tipo de vehículo	Tipo 1	Automático

Tabla 29. Categorías de la Dimensión vehículo		
Categoría	Jerarquía	Ejemplo
Tipo de vehículo	tipo vehículo	- Manual - Automático

1.4.5.8. Dimensión unidad autorizadora

Tabla 30. Detalle de la tabla de Dimensión unidad autorizadora			
Atributo	Descripción	Cambio	Ejemplo
Key: unidad autorizadora	Clave subrogada de la dimensión	Tipo 0	1
Pk: unidad autorizadora	Clave operacional de la unidad autorizada	Tipo 1	12
Nombre	Nombre de la unidad autorizada	Tipo 1	Hospital Eugenio Espejo
Id: entidad autorizada	Identificador del atributo entidad autorizada	Tipo 1	1
Nombre entidad	Nombre de la entidad autorizada	Tipo 1	Ministerio de salud publica
Director	Nombre del director de la unidad autorizadora	Tipo 1	Dr. Leonardo Pazmiño
Id: provincia	Identificador del atributo provincia	Tipo 1	16
Provincia	Provincia en la que se encuentra la unidad autorizadora	Tipo 1	Pichincha

Tabla 31. Categorías de la Dimensión unidad autorizada		
Categoría	Jerarquía	Ejemplo
Unidad autorizadora	provincia ↓ nombre	- Pichincha - Hospital militar de Quito - Azuay - José Carrasco Arteaga
Entidad autorizada	nombre entidad	- Ministerio de salud publica - Conadis

1.4.5.9. Dimensión ayudas

Tabla 32. Detalle de la tabla de dimensión ayudas			
Atributo	Descripción	Cambio	Ejemplo
Key: ayudas	Clave subrogada de la dimensión	Tipo 0	1
Pk: Ayuda clase	Clave operacional de la clase de ayuda	Tipo 1	1
Ayuda clase	Clase de ayuda	Tipo 1	Medicamento
Pk: ayuda	Clave operacional de la ayuda	Tipo 1	2
Ayuda descripción	Descripción de la ayuda	Tipo 1	Prótesis

Tabla 33. Categorías de la dimensión ayudas		
Categoría	Jerarquía	Ejemplo
Ayudas	ayuda clase ↓ ayuda descripción	- Medicamento - Acido valproico - Ayuda técnica - Ortesis - Protesis

1.4.5.10. Dimensión estado ayuda

Tabla 34. Detalle de la tabla de Dimensión estado ayuda			
Atributo	Descripción	Cambio	Ejemplo
Key: estado ayuda	Clave subrogada de la dimensión	Tipo 0	1
Pk: estado ayuda	Clave operacional del estado de la ayuda	Tipo 1	1
Estado ayuda	Estado de la ayuda	Tipo 1	pendiente

Tabla 35. Categorías de la Dimensión estado ayuda		
Categoría	Jerarquía	Ejemplo
Estado ayuda	estado ayuda	- Solicitado - Entregado

1.4.5.11. Dimensión entidad subvenciona

Tabla 36. Detalle de la tabla de Dimensión entidad subvenciona			
Atributo	Descripción	Cambio	Ejemplo
Key: entidad subvenciona	Clave subrogada de la dimensión	Tipo 0	1
Pk: entidad subvenciona	Clave operacional de la entidad subvenciona	Tipo 1	1
Nombre	Nombre de la entidad subvenciona	Tipo 1	Vicepresidencia
Siglas	Siglas de la entidad subvenciona	Tipo 1	VICE

Tabla 37. Categorías de la dimensión entidad subvenciona		
Categoría	Jerarquía	Ejemplo
Entidad subvenciona	nombre	- Vicepresidencia - Fundación hermano Miguel

1.4.5.12. Dimensión entidad provee

Tabla 38. Detalle de la tabla de dimensión entidad provee			
Atributo	Descripción	Cambio	Ejemplo
Key: entidad proveedora	Clave subrogada de la dimensión	Tipo 0	1
Pk: entidad proveedora	Clave operacional de la entidad proveedora	Tipo 1	1
Nombre	Nombre de la entidad proveedora	Tipo 1	Vicepresidencia
Siglas	Siglas de la entidad proveedora	Tipo 1	VICE

Tabla 39. Categorías de la Dimensión entidad provee		
Categoría	Jerarquía	Ejemplo
Entidad proveedora	nombre	- Gasespol - Mequimedix

1.4.5.13. Dimensión proyecto

Tabla 40. Detalle de la tabla de Dimensión proyecto			
Atributo	Descripción	Cambio	Ejemplo
Key: proyecto	Clave subrogada de la dimensión	Tipo 0	1
Pk: proyecto	Clave operacional del proyecto	Tipo 1	1
Título	Título del proyecto	Tipo 1	Ayudas Técnicas Fundación
Id: entidad	Identificador del atributo entidad	Tipo 1	7
Entidad	Nombre de la entidad asociada al proyecto	Tipo 1	Fundación hermano Miguel
Responsable	Nombre del responsable del proyecto	Tipo 1	Juan Pérez
Id: localidad	Identificador del atributo localidad	Tipo 1	197
Localidad	Lugar donde se ejecuta el proyecto	Tipo 1	Quito
Id: tipo proyecto	Identificador del atributo tipo proyecto	Tipo 1	2
Tipo proyecto	Tipo de proyecto	Tipo 1	Ayuda técnica

Tabla 41. Categorías de la Dimensión proyecto		
Categoría	Jerarquía	Ejemplo
Proyecto	tipo proyecto ↓ título	- Por concurso - Equipamiento básico para la escuela de educación especial Oswaldo Guayasamin - Ayuda técnica - Ayudas técnicas fundación hermano Miguel

1.4.6. IDENTIFICACIÓN DE LA FUENTE DE DATOS PARA TABLAS DE HECHOS Y DIMENSIONES

La fuente de datos es la base de datos del sistema web SIC, a continuación se lista las tablas de la base de datos para cada tablas de hechos y dimensión del Data mart.

Tabla 42. Identificación de fuentes de datos para las tablas de Data mart	
Data mart	BDD SIC
Tabla de hechos de personas discapacitadas	- persona
Tabla de hechos de autorización de vehículos	- solicitud_veh - autorizacion_veh
Tabla de hechos de ayudas	- medicamento_persona - ayudas_tecnicas
Dimensión persona	- persona - estado_civil
Dimensión persona – discapacidad	- grado_discapacidad - deficiencia_principal - causa_deficiencia
Dimensión persona – hogar	- con_quien_vive - tipo_vivienda
Dimensión persona – trabajo	- nivel_instruccion - tipo_trabajo - causa_no_trabaja
Dimensión localidad	- localidad
Dimensión tiempo	- persona - solicitud_veh - autorizacion_veh - medicamento_persona - ayudas_tecnicas
Dimensión vehículo	- tipo_vehiculo
Dimensión unidad autorizadora	- unidad_autorizadora - entidad_autorizadora - localidad
Dimensión ayudas	- tipo_ayuda_tecnica - medicamento_o_insumo
Dimensión estado ayuda	- estado_ayuda
Dimensión entidad subvenciona	- entidad_subvenciona
Dimensión entidad provee	- entidad_provee
Dimensión proyecto	- proyecto - entidad - localidad - tipo_proyecto

1.4.6.1. Mapeo detallado de la fuente de datos¹

El (*) al final del nombre de una columna, determina que su valor esta sujeto a transformaciones y/o será generado, estas actividades se detallan posteriormente en el proceso ETL.

Tabla 43. Mapeo detallado con la fuente de datos			
Destino			Fuente
Dimensión persona			
Columna	Tipo de dato	Tabla	Columna
Key: persona*	Integer	----	----
Pk: persona	Integer	persona	per_id
Apellido 1	Varchar(20)	persona	per_apellido1
Apellido 2	Varchar(20)	persona	per_apellido2
Nombre 1	Varchar(20)	persona	per_nombre1
Nombre2	Varchar(20)	persona	per_nombre2
Id: Genero	Varchar(1)	persona	per_sexo
Genero*	Varchar(10)	----	----
Id: estado civil	Integer	persona	est_id
Estado civil	Varchar(20)	estado_civil	est_descripcion
Id: edad*	Integer	persona	per_fecha_nac
Id: edad etapa*	Integer	----	----
Edad etapa*	Varchar(20)	----	----
Estado registro	Varchar(1)	persona	per_estado
Dimensión persona – discapacidad			
Columna	Tipo de dato	Tabla	Columna
Key: persona discapacidad*	Integer	----	----
Pk: grado discapacidad	Integer	grado_discapacidad	gra_id
Grado discapacidad	Varchar(20)	grado_discapacidad	gra_descripcion

¹ The Data Warehouse ETL Toolkit, chapter 3 – The logical data map

Pk: deficiencia principal	Integer	deficiencia_principal	dpr_id
Deficiencia principal	Varchar(30)	deficiencia_principal	dpr_descripcion
Pk: causa deficiencia	Integer	causa_deficiencia	cde_id
Causa deficiencia	Varchar(40)	causa_deficiencia	cde_descripcion
Dimensión persona - hogar			
Columna	Tipo de dato	Tabla	Columna
Key: persona hogar*	Integer	----	----
Pk: con quien vive	Integer	con_quien_vive	con_id
Con quien vive	Varchar(20)	con_quien_vive	con_descripcion
Pk: tipo vivienda	Integer	tipo_vivienda	viv_id
Tipo vivienda	Varchar(20)	tipo_vivienda	viv_descripcion
Dimensión persona – trabajo			
Columna	Tipo de dato	Tabla	Columna
Key: persona trabajo*	Integer	----	----
Pk: nivel instrucción	Integer	nivel_instruccion	niv_id
Nivel instrucción	Varchar(20)	nivel_instruccion	niv_descripcion
Trabaja*	Varchar(2)	----	----
Pk: trabaja descripción	Integer	- tipo_trabajo - causa_no_trabaja	- tra_id - ctr_id
Trabajo descripción	Varchar(30)	- tipo_trabajo - causa_no_trabaja	- tra_descripcion - ctr_descripcion
Dimensión localidad			
Columna	Tipo de dato	Tabla	Columna
Key: localidad*	Integer	----	----
Pk: parroquia	Integer	localidad	loc_id
Pk: cantón	Integer		loc_id
Pk: provincia	Integer		loc_id
Parroquia código	Varchar(20)		loc_codigo
Cantón código	Varchar(20)		loc_codigo
Provincia código	Varchar(20)		loc_codigo

Región código*	Varchar(20)				loc_descripcion
Parroquia	Varchar(100)				loc_descripcion
Cantón	Varchar(100)				loc_descripcion
Provincia	Varchar(100)				loc_descripcion
Región*	Varchar(100)		----		----
Dimensión tiempo					
Columna	Tipo de dato	Tabla	Columna		
Key: tiempo*	Integer	----	----		
Fecha completa*	Date				
Día*	Integer	- persona	- per_fecha_carnet		
Mes*	Integer	- solicitud_veh	- sol_fechaoriginal		
Año*	Integer	- autorización_veh	- sol_fecharecepcion		
		- medicamento_persona	- aut_fecha		
		- ayudas_tecnicas	- aut_fecha_autoriza		
			- mep_fecha_entrega		
			- ayu_fecha_entrega		
Mes nombre*	Varchar(30)	----	----		
Dimensión vehículo					
Columna	Tipo de dato	Tabla	Columna		
Key: vehículo*	Integer	----	----		
Pk: tipo vehículo	Integer	tipo_vehiculo	tve_id		
Tipo vehículo	Varchar(150)	tipo_vehiculo	tve_descripcion		
Dimensión unidad autorizadora					
Columna	Tipo de dato	Tabla	Columna		
Key: unidad autorizadora*	Integer	----	----		
Pk: unidad autorizadora	Integer	unidad_autorizadora	uni_id		
Nombre	Varchar(100)	unidad_autorizadora	uni_nombre		
Id: entidad autorizada	Integer	unidad_autorizadora	eau_id		
Nombre entidad	Varchar(120)	entidad_autorizadora	eau_nombre		
Director	unidad_autorizadora	unidad_autorizadora	uni_director		
Id: provincia	Integer	unidad_autorizadora	loc_id		

Provincia	Varchar(100)	localidad	loc_descripcion
Dimensión ayudas			
Columna	Tipo de dato	Tabla	Columna
Key: ayudas*	Integer	----	----
Ayuda clase*	Varchar(50)	----	----
Pk: ayuda	Integer	- tipo_ayuda_tecnica - medicamento_o_insumo	- tay_id - med_id
Ayuda descripción	Varchar(100)		- tay_descripcion - med_descripcion
Dimensión estado ayuda			
Columna	Tipo de dato	Tabla	Columna
Key: estado ayuda*	Integer	----	----
Pk: estado ayuda	Integer	estado_ayuda	eay_id
Estado ayuda	Varchar(50)	estado_ayuda	eay_descripcion
Dimensión entidad subvencionaria			
Columna	Tipo de dato	Tabla	Columna
Key: entidad subvencionaria*	Integer	----	----
Pk: entidad subvencionaria	Integer	entidad_subvencionaria	sub_id
Nombre	Varchar(50)		sub_nombre
Siglas	Varchar(50)		sub_siglas
Dimensión entidad provee			
Columna	Tipo de dato	Tabla	Columna
Key: entidad proveedora*	Integer	----	----
Pk: entidad proveedora	Integer	entidad_provee	pro_id
Nombre	Varchar(50)		pro_nombre
Siglas	Varchar(50)		pro_siglas
Dimensión proyecto			
Columna	Tipo de dato	Tabla	Columna

Key: proyecto*	Integer	----	----
Pk: proyecto	Integer	proyecto	pro_id
Titulo	Varchar(200)	proyecto	pro_titulo
Id: entidad	Integer	proyecto	ent_id
Entidad	Varchar(100)	entidad	ent_nombre
Responsable	Varchar(100)	proyecto	pro_responsable
Id: localidad	Integer	proyecto	loc_id
Localidad	Varchar(100)	localidad	loc_descripcion
Id: tipo proyecto	Integer	proyecto	tpr_id
Tipo proyecto	Varchar(50)	tipo_proyecto	tpr_descripcion
Tabla de hechos de personas discapacitadas			
Columna	Tipo de dato	Tabla	Columna
Fact: Personas*	Integer	----	----
Fact: Porcentaje discapacidad	Double	persona	per_porcentaje_disc
Fact: Edad discapacidad	Integer	persona	per_edad_disc
Tabla de hechos de autorización de vehículos			
Columna	Tipo de dato	Tabla	Columna
Fact: Vehículos*	Integer	----	----
Fact: Recepción(días)*	Integer	solicitud_veh	- sol_fechaoriginal - sol_fecharecepcion
Fact: Atención(días)*	Integer	- solicitud_veh - autorizacion_veh	- sol_fecharecepcion - aut_fecha
Fact: Resolución(días)*	Integer	autorizacion_veh	- aut_fecha - aut_fecha_autoriza
Fact: Total(días)*	Integer	----	----
Tabla de hechos de ayudas			
Columna	Tipo de dato	Tabla	Columna
Fact: Ayudas*	Integer	----	----
Fact: Monto	Double	- medicamento_persona - ayudas_tecnicas	- mep_monto - ayu_monto_total

Fact: Subvención	Double	- medicamento_persona - ayudas_tecnicas	- mep_valor - ayu_valor
------------------	--------	--	----------------------------

2. CAPÍTULO 2 : DISEÑO FÍSICO, CONSTRUCCIÓN Y DESPLIEGUE DEL DATA MART

2.1. DISEÑO FÍSICO¹

En esta etapa del ciclo de vida, se contemplan los pasos para convertir el diseño lógico en una base de datos física, las dimensiones y tablas de hechos se definen como tablas de base de datos respetando los tipo de datos asignados a los atributos conforme se planteo en el mapeo con la fuente de datos.

Las tareas realizadas comprenden:

- Elaboración de estándares, para nombrar a los objetos de la base de datos y el ETL.
- Construcción del modelo físico.
- Elaboración del plan de índices.
- Generación de código DDL² para la creación de la base de datos.
- Construcción de los cubos OLAP.

2.1.1. ELABORACIÓN DE ESTÁNDARES

Para elaborar los nombres de los elementos de la base de datos y los objetos del ETL se utilizó 3 componentes:

Nombre: <nombre> Describe al elemento, debe ser claro, no ambiguo, responde a la pregunta ¿Qué es este objeto?.

Clase: <clase> Describe la clasificación a la que pertenece el elemento, responde a la pregunta ¿Qué tipo de objeto es este?.

Calificador: <calificador> Componente opcional que permite ampliar la descripción del nombre o de la clase.

¹ The Data Warehouse Lifecycle Toolkit, chapter 15 – Completing the Physical Design

² Data definition language, ver Glosario

2.1.1.1. Estándar para los nombres de los objetos de la base de datos

Se determinó estándares para nombrar a las tablas, columnas e índices de la base de datos.

Estándar para los nombres de las tablas de la bdd

Tabla 44. Estándar para los nombres de las tablas de la bdd	
Formato	<clase> _<nombre>
Clases	dim: Representa tablas de tipo Dimensión. fact: Representa tablas de tipo Tabla de hechos.
Nombre	Corresponde al nombre determinado en el modelo dimensional.

Tabla 45. Nombres de las tablas de la bdd del Data mart		
Modelo lógico	Modelo físico	Abreviatura
Dimensión persona	dim_persona	dim_per
Dimensión persona - discapacidad	dim_persona_discapacidad	dim_ped
Dimensión persona - hogar	dim_persona_hogar	dim_peh
Dimensión persona - trabajo	dim_persona_trabajo	dim_pet
Dimensión localidad	dim_localidad	dim_loc
Dimensión tiempo	dim_tiempo	dim_tim
Dimensión vehículo	dim_vehiculo	dim_veh
Dimensión unidad autorizadora	dim_unidad_autoriza	dim_uni
Dimensión ayudas	dim_ayudas	dim_ayu
Dimensión estado ayuda	dim_estado_ayuda	dim_eay
Dimensión entidad subvenciona	dim_entidad_subvenciona	dim_ens
Dimensión entidad provee	dim_entidad_provee	dim_enp
Dimensión proyecto	dim_proyecto	dim_pry
Tabla de hechos de personas discapacitadas	fact_personas_discapacitadas	fac_per
Tabla de hechos de autorización de vehículos	fact_autorizacion_vehiculo	fac_veh
Tabla de hechos de ayudas	fact_ayudas	fac_ayu

Estándar para los nombres de las columnas de la bdd

Tabla 46. Estándar para los nombres de las columnas de la bdd	
Formato	<clase> _<nombre> _<calificador>
Clases	Abreviatura de la tabla a la cual pertenece la columna.
Nombre	Corresponde al nombre completo o parcial determinado en el modelo dimensional.
Calificador	Opcional key: Identifica a la columna como la clave subrogada de la tabla. id: Identifica a la columna como clave primaria heredada de la base de datos fuente.

Ejemplos:

Tabla 47. Ejemplos de nombres para las columnas de la bdd del Data mart	
Modelo lógico	Modelo físico
Key: persona	dim_per_key
Pk: estado civil	dim_per_estado_civil_id
Edad_etapa	dim_per_edad_etapa

Estándar para los nombres de los índices de la bdd

Tabla 48. Estándar para los nombres de los índices de la bdd	
Formato	<clase> _<nombre>
Clases	ik: Representa un índice de clave primaria de la tabla. is: Representa un índice compuesto por una columna, que no es la clave primaria de la tabla. im: Representa un índice compuesto por dos o mas columnas.
Nombre	El nombre depende de la clase de índice: ik: Abreviatura de la tabla. is: Nombre de la columna. im: Abreviatura de la tabla seguido de un nombre único.

2.1.1.2. Estándar para los nombres de los objetos del ETL

Se determinó estándares para nombrar los pasos que intervienen en los procesos de extracción, transformación y carga de datos; y para los nombres de los archivos del proceso ETL.

Tabla 49. Estándar para los nombres de los pasos del ETL	
Formato	<i>Pasos de carga y extracción:</i> <nombre> _<calificador2>(<calificador1>) <i>Pasos de transformación:</i> <nombre>(<calificador1>)
Nombre	<i>Pasos de carga y extracción:</i> Abreviatura de la tabla de Dimensión o Tabla de hechos a la cual representa el paso. <i>Pasos de transformación:</i> Descripción corta sobre la operación que realiza el paso.
Calificador1	Opcional para los pasos de transformación. DM: La operación se realiza sobre las tablas del Data mart. DS: La operación se realiza sobre la base de datos fuente. OR: Representa las registros antiguos de una Tabla de hechos. NR: Representa los registros nuevos de una Tabla de hechos. N: Operación de Inserción de datos. U: Operación de actualización de datos. D: Operación de eliminación de datos. Adicionalmente se puede utilizar la abreviatura de la tabla sobre la que ocurre la operación para evitar ambigüedades.
Calificador2	Opcional Información adicional que describe al paso de carga o extracción, para evitar ambigüedades.

Tabla 50. Estándar para los nombres de los archivos del ETL	
Formato	<clase>_<nombre>.<extensión>
Clase	job: Archivos que representan flujos tipo Job. elt: Archivos que representan flujos tipo Transformación. val: Archivos que representan flujos tipo Job para realizar validaciones. cal: Archivos que representan flujos tipo Transformación para realizar la definición de parámetros.
Nombre	Abreviatura de la tabla de dimensión o tabla de hechos sobre la cual se ejecuta el ETL.
Extensión	Extensión de los archivos: kjb: para archivos de tipo Job. ktr: para archivos de tipo Transformación.

2.1.2. CONSTRUCCIÓN DEL MODELO FÍSICO DE DATOS

Mediante el uso de una herramienta de modelado de datos, se paso el modelo lógico a un modelo físico respetando los estándares definidos para los objetos de la base de datos, tipos de datos, etc.

La herramienta permitió generar el código DDL en conformidad con lo anterior, necesario para crear la base de datos sobre PostgreSQL.

El script SQL se lo puede ubicar en el CD adjunto, ver Anexo D - /codigo_fuente/bdd/.

A continuación se presentan los tres modelos físicos construidos.

2.1.2.1. Modelo físico de la Tabla de hechos personas discapacitadas

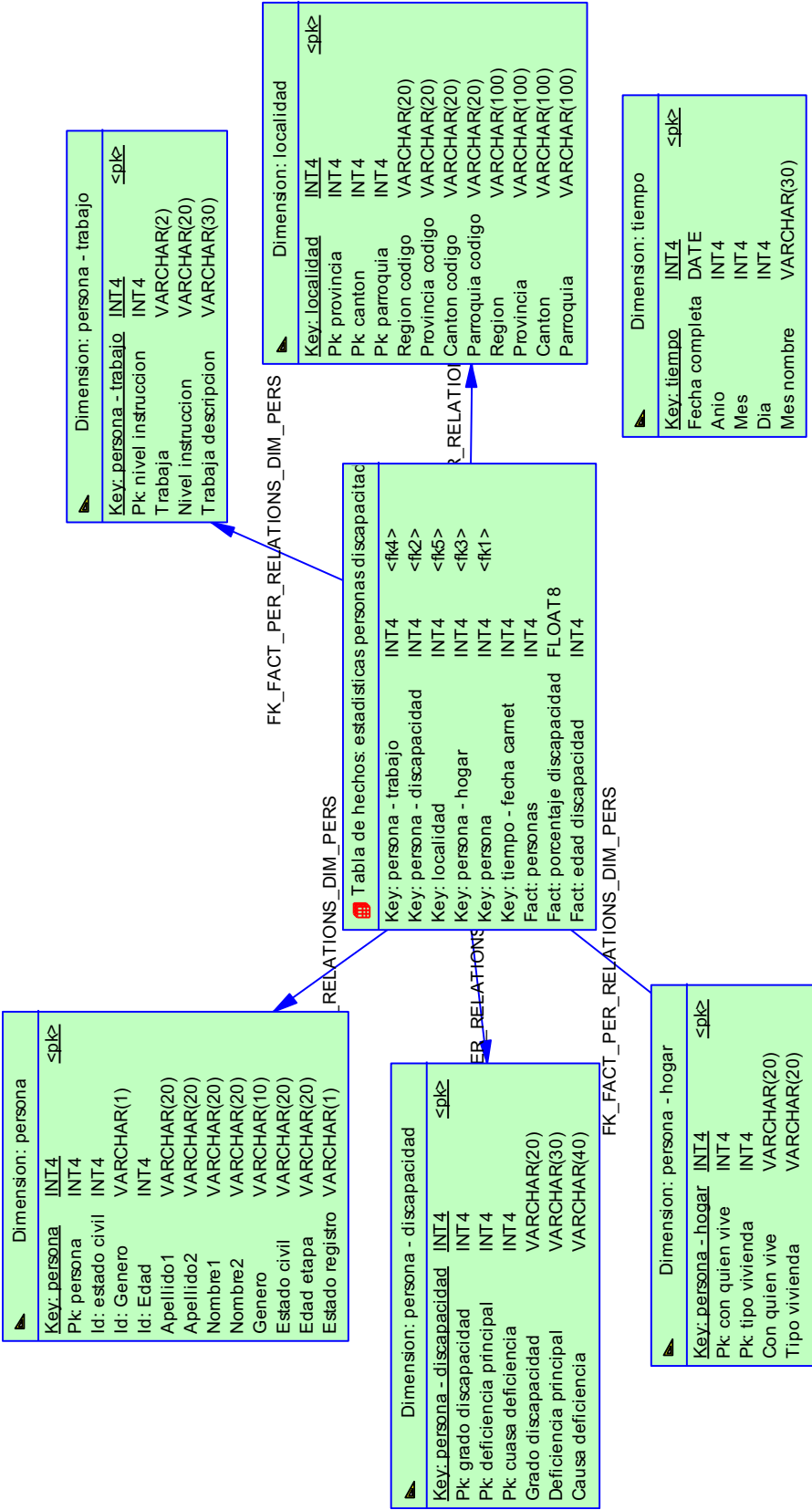


Figura 6. Modelo físico de la Tabla de hechos personas discapacitadas

2.1.2.2. Modelo físico de la Tabla de hechos autorización de vehículos

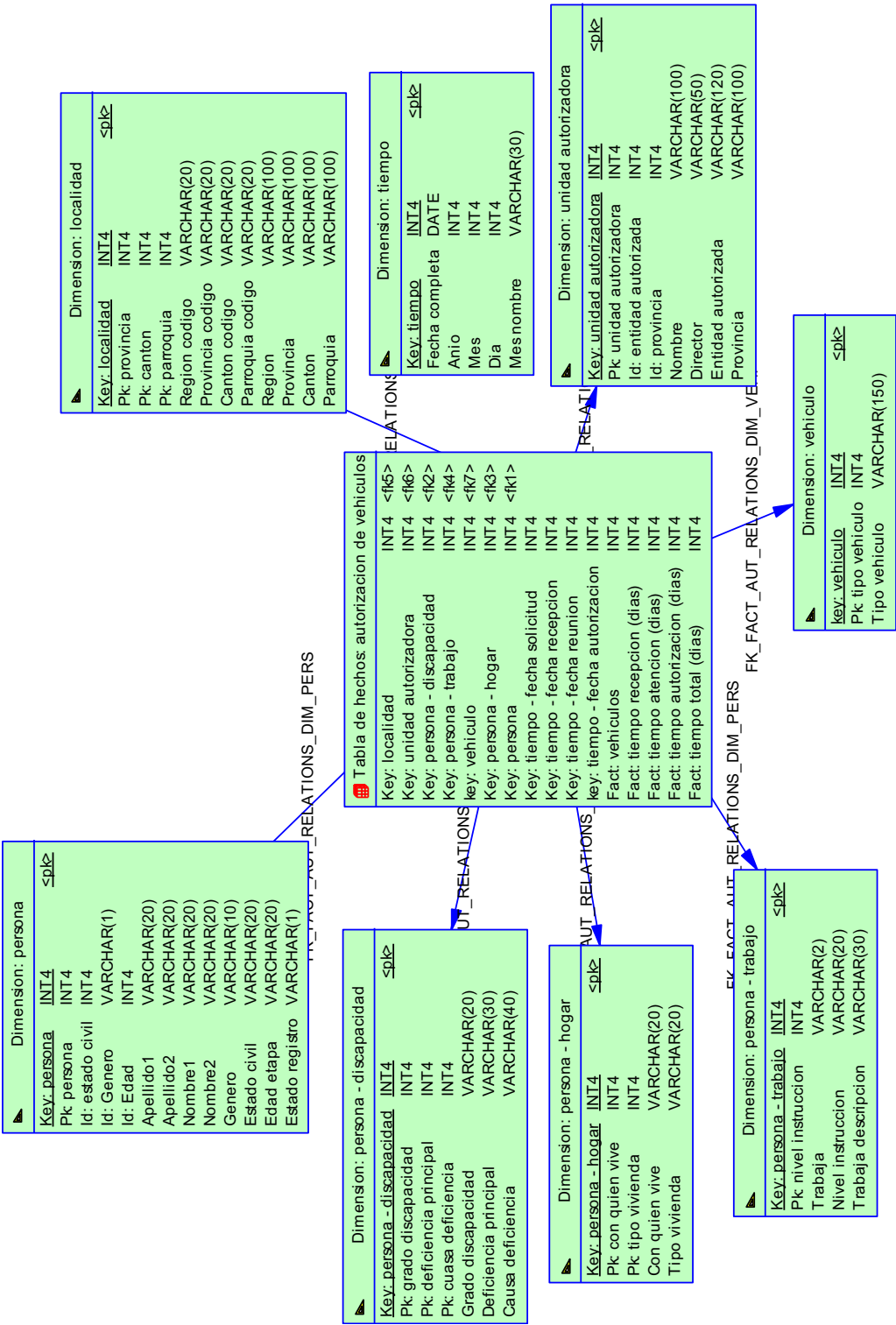


Figura 7. Modelo físico de la Tabla de hechos autorización de vehículos

2.1.2.3. Modelo físico de la Tabla de hechos ayudas

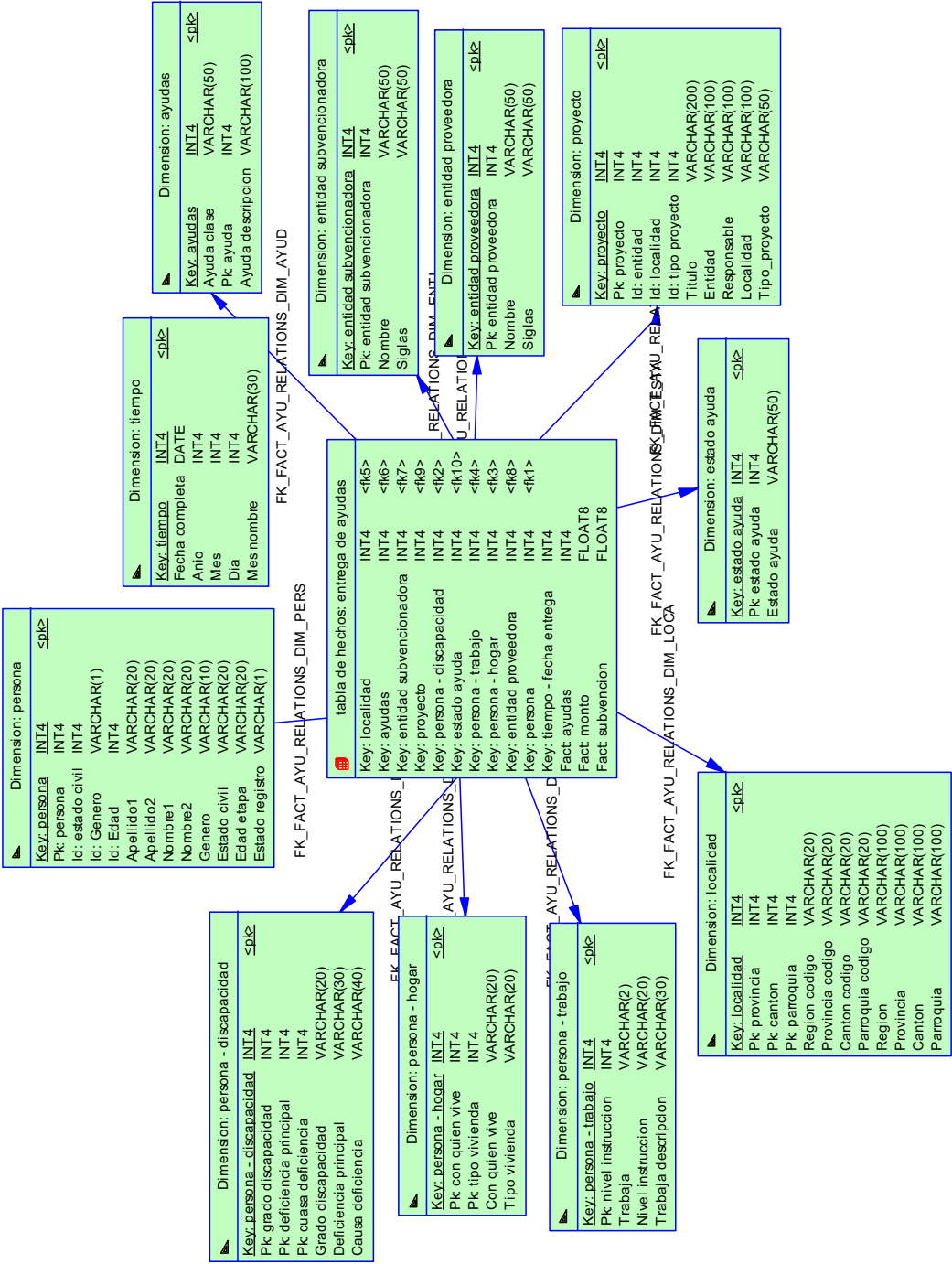


Figura 8. Modelo físico de la Tabla de hechos ayudas

2.1.3. ELABORACIÓN DEL PLAN DE ÍNDICES

Este plan es considerado como un plan inicial de índices, esta sujeto a cambios durante toda la vida del Data mart, en vista que mientras se utilice el Data mart se entenderá mejor los patrones de consulta y filtrado de tal manera que se puedan definir índices mas óptimos.

Todos los índices definidos son del tipo B-tree, debido a las especificaciones de la DBMS PostgreSQL.

En este plan se consideró los siguientes índices:

Índices para las Tablas de hechos.- Corresponde a un índice compuesto por todas las claves dimensionales de la Tabla de hechos, también se definieron índices para cada clave dimensional por separado. Estos índices tienen como propósito mejorar el rendimiento de las consultas, el planificador de consultas de la DBMS es el encargado de utilizar convenientemente estos índices.

Índices para las tablas de Dimensiones.- Se creo un índice para la clave subrogada de la tabla, índices para los atributos e índices compuestos para las jerarquías entre atributos utilizando las claves primarias asociadas tomadas de la bdd fuente. Estos índices tienen como propósito mejorar el rendimiento de las consultas, especialmente cuando se declaran filtros y ordenamientos.

Índices para el proceso ETL.- Para optimizar la carga y actualización de datos en las tablas de Dimensiones y especialmente en las Tablas de hechos, se definió índices simples y compuestos por las claves operacionales en las tablas de Dimensiones para agilizar el proceso de obtención de claves subrogadas.

Para asegurar el correcto funcionamiento del planificador de consultas, los índices y estadísticas deben estar actualizados, esta operación se definió posteriormente como parte del proceso ETL.

Tabla 51. Resumen de índices para cada Tabla de hechos y Dimensiones	
Tabla: dim_persona	
Índice	Columnas
ik_dim_per	dim_per_key
is_dim_per_id	dim_per_id
is_dim_per_estado_civil_id	dim_per_estado_civil_id
is_dim_per_genero_id	dim_per_genero_id
is_dim_per_edad	dim_per_edad_etapa_id dim_per_edad_id
Tabla: dim_persona_discapacidad	
Índice	Columnas
ik_dim_ped	dim_ped_key
im_dim_ped_ids	- dim_ped_grado_deficiencia_id - dim_ped_deficiencia_principal_id - dim_ped_causa_deficiencia_id
Tabla: dim_persona_hogar	
Índice	Columnas
ik_dim_peh	dim_peh_key
im_dim_peh_ids	- dim_peh_con_quien_vive_id - dim_peh_tipo_vivienda_id
Tabla: dim_persona_trabajo	
Índice	Columnas
ik_dim_pet	dim_pet_key
im_dim_pet_ids	- dim_pet_nivel_instruccion_id - dim_pet_trabaja_id - dim_pet_trabaja_descripcion_id
Tabla: dim_localidad	
Índice	Columnas
ik_dim_loc	dim_loc_key
im_dim_loc_codigos	- dim_loc_region_cod - dim_loc_provincia_cod - dim_loc_canton_cod - dim_loc_parroquia_cod
im_dim_loc_ids	- dim_loc_provincia_id - dim_loc_canton_id - dim_loc_parroquia_id
Tiempo: dim_tiempo	
Índice	Columnas
ik_dim_tim	dim_tim_key
is_dim_tim_fecha_completa	dim_tim_fecha_completa
im_dim_tim_amd	- dim_tim_anio - dim_tim_mes - dim_tim_dia
Tabla: dim_vehiculo	
Índice	Columnas
ik_dim_veh	dim_veh_key

is_dim_veh_tipo_vehiculo_id	dim_veh_tipo_vehiculo_id
Tabla: dim_unidad_autorizada	
Índice	Columnas
ik_dim_uni	dim_uni_key
is_dim_uni_id	dim_uni_id
im_dim_uni_ids	- dim_uni_provincia_id - dim_uni_id
Tabla: dim_ayudas	
Índice	Columnas
ik_dim_ayu	dim_ayu_key
im_dim_ayu_ayudas	- dim_ayu_clase_id - dim_ayu_id
Tabla: dim_estado_ayuda	
Índice	Columnas
ik_dim_eay	dim_eay_key
Tabla: dim_entidad_subvenciona	
Índice	Columnas
ik_dim_ens	dim_ens_key
is_dim_ens_id	dim_ens_id
Tabla: dim_entidad_provee	
Índice	Columnas
ik_dim_enp	dim_enp_key
is_dim_enp_id	dim_enp_id
Tabla: dim_proyecto	
Índice	Columnas
ik_dim_pry	dim_pry_key
is_dim_pry_id	dim_pry_id
im_dim_pry_ids	- dim_pry_tipo_proyecto_id - dim_pry_id
Tabla: fact_personas_discapacitadas	
Índice	Columnas
im_fac_per_keys	- dim_per_key - dim_pet_key - dim_peh_key - dim_pet_key - dim_loc_key - dim_tim_carnet_key
is_fac_per_dim_per_key	dim_per_key
is_fac_per_dim_ped_key	dim_pet_key
is_fac_per_dim_peh_key	dim_peh_key
is_fac_per_dim_pet_key	dim_pet_key
is_fac_per_dim_loc_key	dim_loc_key
is_fac_per_dim_tim_carnet_key	dim_tim_carnet_key
Tabla: fact_autorizacion_vehiculo	

Índice	Columnas
im_fac_veh_keys	<ul style="list-style-type: none"> - dim_per_key - dim_pet_key - dim_peh_key - dim_pet_key - dim_loc_key - dim_veh_key - dim_uni_key - dim_tim_solicitud_key - dim_tim_recepcion_key - dim_tim_reunion_key - dim_tim_autoriza_key
is_fac_veh_dim_per_key	dim_per_key
is_fac_veh_dim_ped_key	dim_pet_key
is_fac_veh_dim_peh_key	dim_peh_key
is_fac_veh_dim_pet_key	dim_pet_key
is_fac_veh_dim_loc_key	dim_loc_key
is_fac_veh_dim_veh_key	dim_veh_key
is_fac_veh_dim_uni_key	dim_uni_key
is_fac_veh_dim_tim_solicitud_key	dim_tim_solicitud_key
is_fac_veh_dim_tim_recepcion_key	dim_tim_recepcion_key
is_fac_veh_dim_tim_reunion_key	dim_tim_reunion_key
is_fac_veh_dim_tim_autoriza_key	dim_tim_autoriza_key
Tabla: fact_ayudas	
Índice	Columnas
im_fac_per_keys	<ul style="list-style-type: none"> - dim_per_key - dim_pet_key - dim_peh_key - dim_pet_key - dim_loc_key - dim_ayu_key - dim_ens_key - dim_enp_key - dim_pry_key - dim_eay_key - dim_tim_entrega_key
is_fac_ayu_dim_per_key	dim_per_key
is_fac_ayu_dim_ped_key	dim_pet_key
is_fac_ayu_dim_peh_key	dim_peh_key
is_fac_ayu_dim_pet_key	dim_pet_key
is_fac_ayu_dim_loc_key	dim_loc_key
is_fac_ayu_dim_ayu_key	dim_ayu_key
is_fac_ayu_dim_ens_key	dim_ens_key
is_fac_ayu_dim_enp_key	dim_enp_key
is_fac_ayu_dim_pry_key	dim_pry_key
is_fac_ayu_dim_eay_key	dim_eay_key
is_fac_per_dim_tim_entrega_key	dim_tim_entrega_key

2.1.4. CONSTRUCCIÓN DE LOS CUBOS OLAP

Como fue determinado en la selección de productos, el motor OLAP a ser utilizado es Mondrian, por lo tanto la construcción de cubos se la realizó mediante el esquema de etiquetas¹ XML propuesto por Pentaho - Mondrian.

La estructura básica del esquema es la siguiente:

```
<Schema>
  <Cube>
    <Dimension>
      <Hierarchy>
        <Level/>
      </Hierarchy>
    </Dimension>
    <DimensionUsage/>
    <Measure />
  </Cube>
</Schema>
```

Figura 9. Estructura básica del esquema Mondrian

<Schema>.- Colección de dimensiones compartidas y cubos, y otros objetos del Data mart.

<Dimension>.- Corresponde a colecciones jerarquizadas de los atributos de las tablas de dimensión, si es declarada al mismo nivel que los cubos se considera como una dimensión compartida, es decir puede ser utilizada por cualquier cubo dentro del esquema; caso contrario es una dimensión propia del cubo.

<Cube>.- Colección de dimensiones y medidas basadas en una misma tabla de hechos.

<Measure>.- Corresponde a cálculos y agregaciones a partir de las medidas de una tabla de hechos.

¹ <http://mondrian.pentaho.com/documentation/schema.php>

Algunos lineamientos utilizados en la declaración de dimensiones y cubos:

- Las dimensiones compartidas se definieron a partir de la información de la Tabla 12.
- Cada una de las categorías planteadas por dimensión en el modelado dimensional son declaradas utilizando la etiqueta <Dimension>.
- Cada una de las tablas de hechos planteadas en el modelado dimensional son declaradas utilizando la etiqueta <Cube>.
- Las dimensiones compartidas son declaradas dentro del cubo mediante la etiqueta <DimensionUsage> y las dimensiones propias son declaradas mediante la etiqueta <Dimension>.
- Las medidas definidos en el detalle de tablas de hechos son declarados mediante la etiqueta <Measure>.

A continuación se presenta un ejemplo para una dimensión y un cubo construidos bajo el esquema de etiquetas de Mondrian.

Código para las categorías “Persona – Genero” y “Persona – Edad” de la dimensión persona .

```
<Dimension name="dim_per_genero" caption="Persona – Genero">
  <Hierarchy hasAll="true" allMemberName="Todos" primaryKey="dim_per_key">
    <Table name="dim_persona"/>
    <Level name="genero" column="dim_per_genero" uniqueMembers="true"/>
  </Hierarchy>
</Dimension>

<Dimension name="dim_per_edad" caption="Persona – Edad">
  <Hierarchy hasAll="true" allMemberName="Todos" primaryKey="dim_per_key">
    <Table name="dim_persona"/>
    <Level name="edad etapa" column="dim_per_edad_etapa_id"
nameColumn="dim_per_edad_etapa_id" uniqueMembers="true"/>
    <Level name="edad" column="dim_per_edad_id" uniqueMembers="true"/>
  </Hierarchy>
</Dimension>
```

Código para el cubo de la tabla de hechos de personas discapacitadas.

```
<Cube name="Personas discapacitadas">
  <Table name="fact_personas_discapacitadas"/>
  <DimensionUsage name="dim_per_genero" source="dim_per_genero"
foreignKey="dim_per_key"/>
  <DimensionUsage name="dim_per_edad" source="dim_per_edad"
foreignKey="dim_per_key"/>
  <DimensionUsage name="dim_per_estado_civil"
source="dim_per_estado_civil" foreignKey="dim_per_key"/>
  <DimensionUsage name="dim_per_estado" source="dim_per_estado"
foreignKey="dim_per_key"/>
  <DimensionUsage name="dim_ped_deficiencia_principal"
source="dim_ped_deficiencia_principal" foreignKey="dim_ped_key"/>
  <DimensionUsage name="dim_ped_grado_discapacidad"
source="dim_ped_grado_discapacidad" foreignKey="dim_ped_key"/>
  <DimensionUsage name="dim_ped_cuasa_deficiencia"
source="dim_ped_cuasa_deficiencia" foreignKey="dim_ped_key"/>
  <DimensionUsage name="dim_peh_con_quien_vive"
source="dim_peh_con_quien_vive" foreignKey="dim_peh_key"/>
  <DimensionUsage name="dim_peh_tipo_vivienda"
source="dim_peh_tipo_vivienda" foreignKey="dim_peh_key"/>
  <DimensionUsage name="dim_pet_nivel_instruccion"
source="dim_pet_nivel_instruccion" foreignKey="dim_pet_key"/>
  <DimensionUsage name="dim_pet_trabajo" source="dim_pet_trabajo"
foreignKey="dim_pet_key"/>
  <DimensionUsage name="dim_loc" source="dim_loc"
foreignKey="dim_loc_key"/>
  <Dimension name="dim_tim_carnet" caption="Tiempo carnetizacion"
foreignKey="dim_tim_carnet_key" type="TimeDimension">
    <Hierarchy hasAll="true" allMemberName="Todos"
primaryKey="dim_tim_key">
      <Table name="dim_tiempo"/>
      <Level name="anio" column="dim_tim_anio" type="Numeric"
uniqueMembers="true" levelType="TimeYears"/>
      <Level name="mes" nameColumn="dim_tim_mes_nombre"
column="dim_tim_mes" type="Numeric" uniqueMembers="false"
levelType="TimeMonths"/>
      <Level name="dia" column="dim_tim_dia" type="Numeric"
uniqueMembers="false" levelType="TimeDays"/>
    </Hierarchy>
  </Dimension>
  <Measure name="fac_per_num" caption="Personas" column="dim_per_key"
aggregator="count"/>
</Cube>
```

El archivo completo se ubica en el CD adjunto, ver Anexo D – /codigo_fuente/olap/.

2.2. DISEÑO Y DESARROLLO DE LA: EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS (ETL)¹

En esta etapa nos concentramos en el Back Room, especialmente en el área ETL donde se definen los procesos de extracción, limpieza, transformación y carga de datos, desde la base de datos fuente hacia la base de datos del Data mart, para llevar acabo esta etapa se determinaron las siguientes tareas:

- Elaboración del plan ETL
- Definición de Transformaciones y Cálculos
- ETL de las tablas de dimensión
- ETL de las tablas de hechos
- Construcción de los procesos ETL

2.2.1. ELABORACIÓN DEL PLAN ETL

El plan ETL detalla la forma como se realizó la carga de datos desde los sistemas fuente hacia el Data mart, se planteó tres procesos, a continuación se presenta el escenario para el área de ETL.

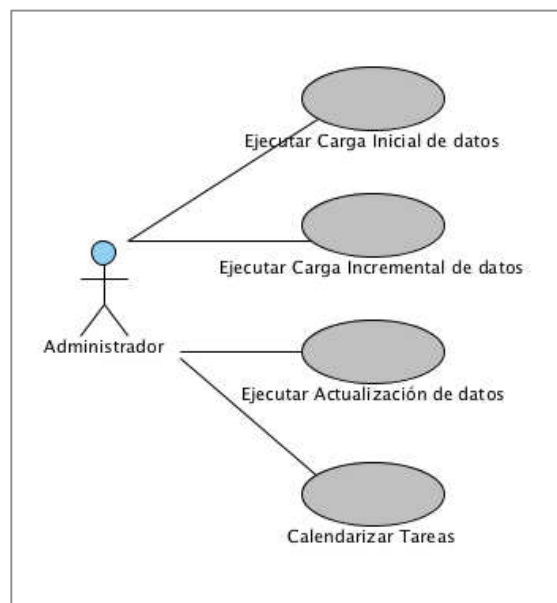


Figura 10. Escenario del área ETL

¹ The Data Warehouse Lifecycle Toolkit, chapter 16 – Data Staging

El administrador del sistema es el encargado de ejecutar los procesos ETL, estos procesos pueden ser ejecutados manualmente o pueden ser calendarizados.

Cada proceso ejecuta una serie de trabajos, los cuales cargan y mantienen actualizados los datos de las dimensiones y tablas de hechos

Cada trabajo consta de tres partes y son:

Validación.- Verifica la existencia de las tablas y archivos necesarios para realizar el proceso ETL, en caso de no encontrar alguna de las dependencias se abortara el proceso ETL.

ETL.- Es el flujo de datos, comprende la extracción de datos de la fuente, procesos de transformación de datos, generación y/o calculo de valores y finalmente la carga de los datos tratados a la base de datos del Data mart.

Actualización.- Actualización de índices y estadísticas de las tablas afectadas.

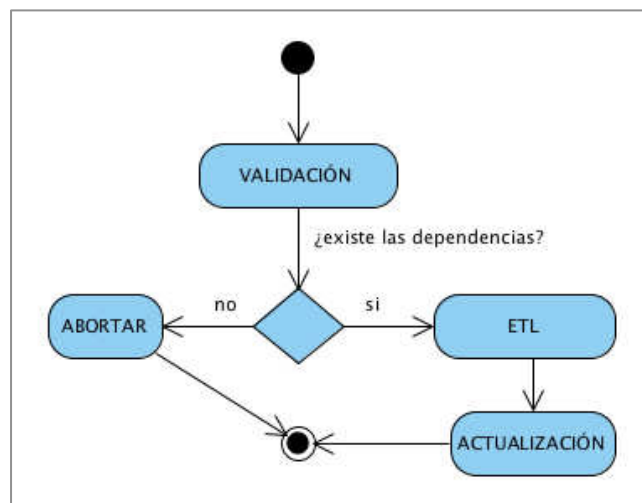


Figura 11. Diagrama del Trabajo para dimensiones

Adicionalmente, los trabajos para realizar la carga de datos sobre las tablas de hechos cuentan con pasos adicionales que permite capturar y guardar las fechas que sirven como parámetros para ejecutar el trabajo.

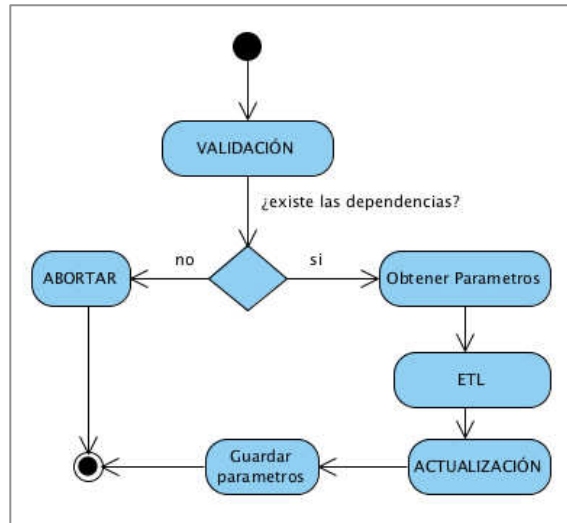


Figura 12. Diagrama del Trabajo para tablas de hechos

A continuación se detallan los tres procesos ETL planeados:

2.2.1.1. Carga inicial de datos

Este proceso carga datos sobre el Data mart por primera vez, el primer trabajo consiste en la creación de la base de datos, el segundo trabajo consiste en llenar la dimensión tiempo, como tercer paso se ejecutan los trabajos para llenar el resto de dimensiones y finalmente se ejecutan los trabajos para llenar las tablas de hechos.

Debido a que se trata de la carga inicial, los registros que se cargan a las tablas de hechos y dimensiones comprenden desde los mas antiguos hasta los ingresados en el sistema operacional la fecha en que se ejecuta la carga inicial.

2.2.1.2. Carga incremental de datos

Este proceso se ejecuta periódicamente y permite mantener actualizadas las dimensiones cargando nuevos registros, así como actualizando los datos ya registrados de acuerdo a la política de cambio que se planteo para cada atributo.

En el caso de las tablas de hechos, se carga los nuevos registros comprendidos entre la ultima ejecución de la carga incremental o inicial hasta la fecha anterior a la nueva ejecución.

2.2.1.3. Actualización de datos

Este proceso permite actualizar las dimensiones cargando nuevos registros así como actualizando los antiguos.

En el caso de las tablas de hechos, aunque es poco común realizar actualizaciones sobre los registros, este proceso permite actualizar los registros comprendidos entre dos fechas, una inicial y una final suministradas por el usuario.

2.2.2. DEFINICIÓN DE TRANSFORMACIONES Y CÁLCULOS

A continuación se detallan las transformaciones de datos y cálculos que se realizaron en el proceso ETL, la lógica de las transformaciones y cálculos esta expresada mediante pseudocódigo.

Tabla 52. Transformaciones y Cálculos en Dimensiones			
ID	Dimensión	Columna	Transformación / Calculo
1	dim_persona	dim_per_genero	<p>Se calcula en base al valor de la columna dim_per_genero_id Tal que, dim_per_genero_id es x, dim_per_genero es y:</p> <p>1) if x==F then y=Femenino 2) if x==M then y=Masculino 3) else y=Sin genero</p>
2	dim_persona	dim_per_edad_id	<p>Se calcula en base a la fecha de nacimiento y la actual, se resta y se obtienen los años: Tal que, dim_per_edad_id es y, fecha actual es x1, fecha de nacimiento es x2:</p> <p>y= cast_to_int(((convertir_a_dias(x1-x2))/365))</p>
3	dim_persona	dim_per_edad_etapa	<p>Se calcula en base al valor de la columna dim_per_edad_id Tal que, dim_per_edad_id es x, dim_per_edad_etapa es y:</p> <p>1) if x>=0 and x<=6 then y=Infantes 2) if x>=7 and x<=11 then y=Niños</p>

			<p>3) if $x \geq 12$ and $x \leq 17$ then $y = \text{Adolescentes}$</p> <p>4) if $x \geq 18$ and $x \leq 25$ then $y = \text{Jovenes}$</p> <p>5) if $x \geq 26$ and $x \leq 59$ then $y = \text{Adultos}$</p> <p>6) if $x \geq 60$ then $y = \text{Adultos mayores}$</p>
4	dim_persona_trabajo	dim_pet_trabaja_id	<p>Dato asignado en la extracción de datos</p> <p>Opciones:</p> <p>Si => para registros de la tabla tipo_trabajo</p> <p>No => para registros de la tabla causa_no_trabaja</p>
5	dim_tiempo	dim_tim_fecha_completa	<p>Se calcula a partir de una fecha constante (1900-01-01) y a cual se suma un numero de días secuenciales</p> <p>Tal que, dim_tim_fecha_completa es x, numero de días es i, fecha constante es y:</p> <pre> For($i=0$; $i < \text{numero_fechas}$; $i++$) { $x = y + i$; } </pre>
6	dim_tiempo	dim_tim_dia	<p>Se calcula a partir de la columna dim_tim_fecha_completa, extrayendo el componente "dia"</p> <p>Tal que, dim_tim_fecha_completa es x, dim_tim_dia es y:</p> <p>$y = \text{dia}(x)$</p>
7	dim_tiempo	dim_tim_mes	<p>Se calcula a partir de la columna dim_tim_fecha_completa, extrayendo el componente "mes"</p> <p>Tal que, dim_tim_fecha_completa es x, dim_tim_mes es y:</p> <p>$y = \text{mes}(x)$</p>
8	dim_tiempo	dim_tim_anio	<p>Se calcula a partir de la columna dim_tim_fecha_completa, extrayendo el componente "año"</p>

			<p>Tal que, dim_tim_fecha_completa es x, dim_tim_anio es y:</p> <p>$y = \text{año}(x)$</p>
9	dim_tiempo	dim_tim_mes_nombre	<p>Se calcula en base al valor de la columna dim_tim_mes.</p> <p>Tal que, dim_tim_mes_nombre es x, dim_tim_mes es y:</p> <p>1) if $x == 1$ then $y = \text{Enero}$ 2) if $x == 2$ then $y = \text{Febrero}$. . . 11) if $x == 11$ then $y = \text{Noviembre}$ 12) if $x == 12$ then $y = \text{Diciembre}$</p>
10	dim_ayudas	dim_ayu_clase_id	<p>Dato asignado en la extracción de datos</p> <p>Opciones:</p> <p>ayuda técnica => para registros de la tabla tipo_ayuda_tecnica medicamento => para registros de la tabla medicamento_o_insumo</p>
11	dim_localidad	dim_loc_region_cod	
12	dim_localidad	dim_loc_region	<p>Datos asignados en base al valor de la columna dim_loc_provincia_cod, la información sobre el código de región y la descripción para cada provincia se hallan en un archivo xml.</p> <p>Tal que, dim_loc_region_cod es $y1$, dim_loc_region es $y2$, dim_loc_provincia_cod es x, la información del archivo es provincias:</p> <pre> For($i=0$; $i < \text{count}(\text{provincias})$; $i++$) { if ($\text{provincias}[i][\text{codigo}] == x$) { $y1 = \text{provincia}[i][\text{region_codigo}]$ $y2 = \text{provincia}[i][\text{region_descrip}]$ } } </pre>

Tabla 53. Transformaciones y Cálculos en Tablas de hechos			
ID	Tabla	Columna	Transformación
1	fact_autorizacion_vehiculo	fac_veh_dias_recepcion	<p>Se calcula en base a las fechas de solicitud y recepción, se restan y se transforma el resultado a días: Tal que, fac_veh_dias_recepcion es y, sol_fecha_recepcion es $x1$, sol_fecha_original es $x2$:</p> $y = \text{cast_to_int}(\text{convertir_a_dias}(x1 - x2))$
2	fact_autorizacion_vehiculo	fac_veh_dias_atencion	<p>Se calcula en base a las fechas de recepción y reunión, se restan y se transforma el resultado a días: Tal que, fac_veh_dias_atencion es y, aut_fecha es $x1$, sol_fecha_recepcion es $x2$:</p> $y = \text{cast_to_int}(\text{convertir_a_dias}(x1 - x2))$
3	fact_autorizacion_vehiculo	fac_veh_dias_autoriza	<p>Se calcula en base a las fechas de reunión y la de autorización, se restan y se transforma el resultado a días: Tal que, fac_veh_dias_autoriza es y, aut_fecha_autoriza es $x1$, aut_fecha es $x2$:</p> $y = \text{cast_to_int}(\text{convertir_a_dias}(x1 - x2))$
4	fact_autorizacion_vehiculo	fac_veh_dias_total	<p>Se calcula en base a la suma de las medidas calculadas (1, 2, 3) Tal que, fac_veh_dias_total es f, fac_veh_dias_recepcion es $f1$, fac_veh_dias_atencion es $f2$, fac_veh_dias_autoriza es $f3$:</p> $f = f1 + f2 + f3$

2.2.3. ETL DE LAS TABLAS DE DIMENSIÓN¹

El flujo de datos base para la carga y actualización de las dimensiones consta de los siguientes pasos:

- 1) Extracción de datos de las tablas origen: Mediante código SQL se seleccionan las columnas que proveen los datos conforme a lo establecido en la Tabla 43, los nombres de las columnas se las redefine conforme al nombre de las columnas de la tabla de dimensión. En este paso se lleva a cabo una primera parte de las transformaciones.
- 2) Transformaciones y cálculos: Los registros entran a este paso y se lleva a cabo las transformaciones y cálculos definidos en la Tabla 52, de ser necesario este paso añade columnas a los registros.
- 3) Asignación de la clave subrogada: En base a las claves operacionales se asigna un número secuencial único al registro, esta es la clave subrogada.
- 4) Carga y actualización de Datos: Los registros del paso anterior son cargados a la tabla de dimensión, los registros nuevos son insertados mientras que los registros alterados son actualizados.

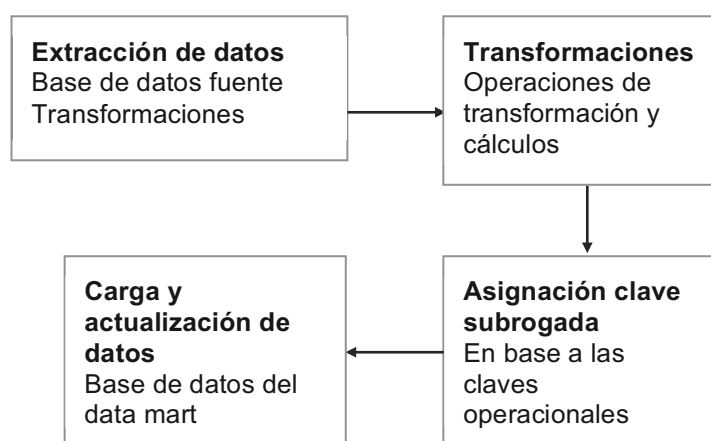


Figura 13. Diagrama del flujo de datos base para dimensiones

¹ The Data Warehouse ETL Toolkit, chapter 5 – Delivering Dimension Tables

A continuación se detalla el flujo de datos ETL para cada dimensión y su representación en la herramienta PDI¹ - kettle.

Para mayor información sobre los pasos utilizados en cada flujo ETL ver el documento de Metadatos del proceso ETL que se ubica en el CD adjunto, ver Anexo D – /documentos/documentos_tecnicos/metadatos_etl/.

2.2.3.1. ETL de la Dimensión persona

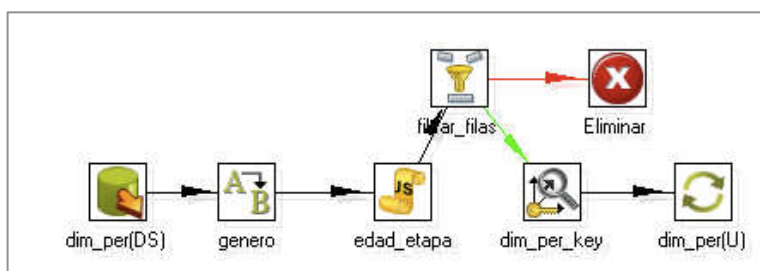


Figura 14. ETL de la Dimensión persona

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión persona, script SQL ver Anexo B1.
- 2) Se llevan a cabo las transformaciones 1, 2, 3 conforme a lo detallado en la Tabla 52, y se verifica que no existan registros con valores nulos.
- 3) Se asigna la clave subrogada para cada registro en base a la clave operacional: dim_per_id.
- 4) Se cargan y actualizan los registros en la dimensión.

¹ Pentaho Data Integration

2.2.3.2. ETL de la Dimensión persona – discapacidad

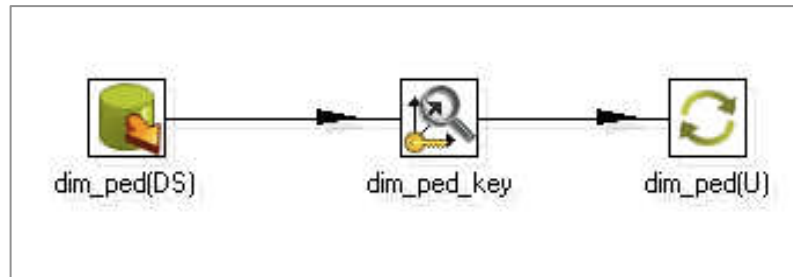


Figura 15. ETL de la Dimensión persona – discapacidad

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión persona – discapacidad, script SQL ver Anexo B2.
- 2) Se asigna la clave subrogada para cada registro en base a las claves operacionales:
 - dim_ped_grado_discapacidad_id
 - dim_ped_deficiencia_principal_id
 - dim_ped_causa_deficiencia_id
- 3) Se cargan y actualizan los registros en la dimensión.

2.2.3.3. ETL de la Dimensión persona – hogar

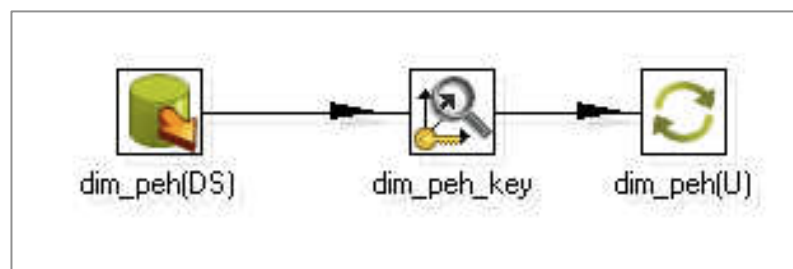


Figura 16. ETL de la Dimensión persona – hogar

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión persona – hogar, script SQL ver Anexo B3.
- 2) Se asigna la clave subrogada para cada registro en base a las claves operacionales:

- dim_peh_con_quien_vive_id
 - dim_peh_tipo_vivienda_id
- 3) Se cargan y actualizan los registros en la dimensión.

2.2.3.4. ETL de la Dimensión persona – trabajo

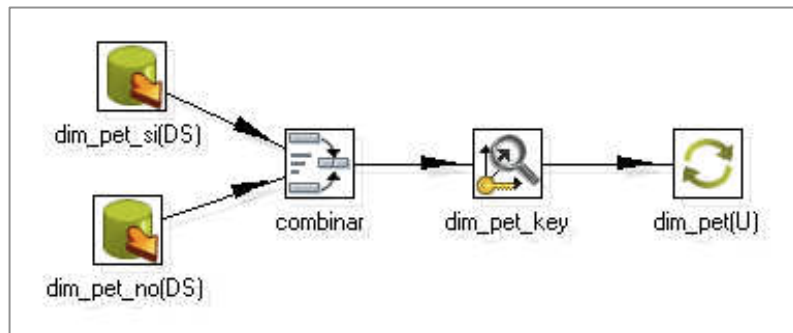


Figura 17. ETL de la Dimensión persona – trabajo

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión persona – trabajo, la extracción se realiza en dos etapas tomado en cuenta la transformación 4 conforme a lo detallado en la Tabla 52, la primera etapa extrae datos sobre tipos de trabajos y la segunda datos sobre causas de no trabajo, scripts SQL ver Anexo B4 y B5.
- 2) Se combinan los registros obtenidos de las extracciones.
- 3) Se asigna la clave subrogada para cada registro en base a las claves operacionales:
 - dim_pet_nivel_instruccion_id
 - dim_pet_trabaja_id
 - dim_pet_trabaja_descripcion_id
- 4) Se cargan y actualizan los registros en la dimensión.

2.2.3.5. ETL de la Dimensión localidad

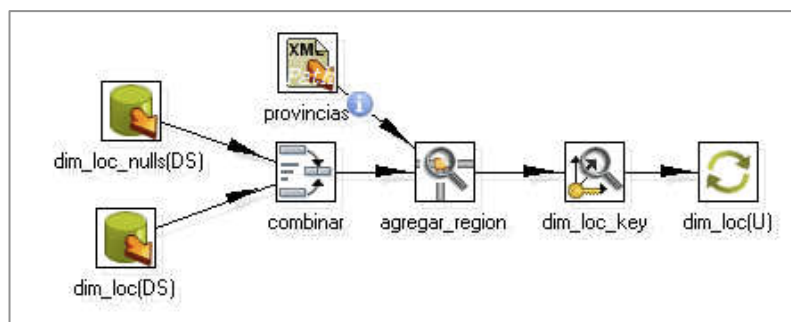


Figura 18. ETL de la Dimensión localidad

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión localidad, la extracción se realizó en dos etapas, la primera asigna datos por defecto para los códigos y descripciones de cantones y parroquias de cada provincia, la segunda obtiene esta información de la bdd fuente, script SQL ver Anexo B6 y B7.
- 2) Se combinan los registros de las extracciones.
- 3) Se lleva a cabo las transformaciones 11 y 12 conforme a lo detallado en la Tabla 52.
- 4) Se asigna la clave subrogada para cada registro en base a las claves operacionales:
 - dim_loc_provincia_id
 - dim_loc_canton_id
 - dim_loc_parroquia_id
- 5) Se cargan y actualizan los registros.

2.2.3.6. ETL de la Dimensión tiempo¹

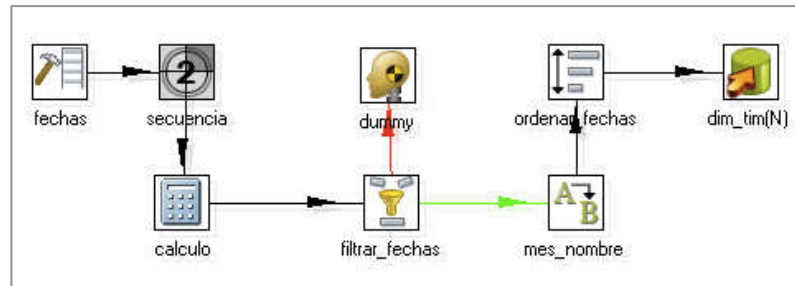


Figura 19. ETL de la Dimensión tiempo

- 1) Se generan 45000 registros con la fecha (1900-01-01) por defecto.
- 2) Para cada registro se asigna un número secuencial comenzando desde 1.
- 3) Se lleva a cabo las transformaciones 5, 6, 7, 8, 9 conforme a lo detallado en la Tabla 52.
- 4) Se filtran las fechas para ingresar solo valores hasta un año determinado, se ordena las fechas en base a la columna: dim_tim_fecha_completa.
- 5) Se cargan los registros en la dimensión.

2.2.3.7. ETL de la Dimensión vehículo

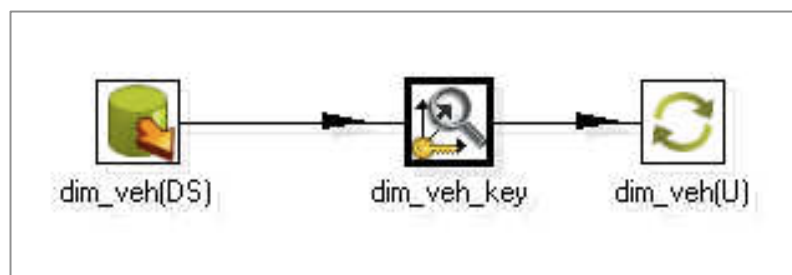


Figura 20. ETL de la Dimensión vehículo

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión vehículo, script SQL ver Anexo B8.
- 2) Se asigna la clave subrogada para cada registro en base a la clave operacional: dim_veh_tipo_vehiculo_id.

¹ Pentaho 3.2 Data Integration Beginner's Guide, chapter 6

- 3) Se cargan y actualizan los registros en la dimensión.

2.2.3.8. ETL de la Dimensión unidad autorizadora

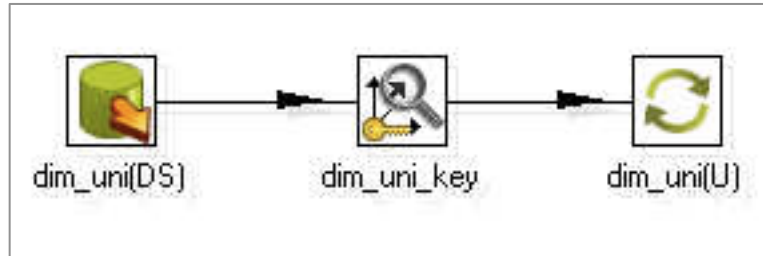


Figura 21. ETL de la Dimensión unidad autorizadora

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión unidad autorizadora, script SQL ver Anexo B9.
- 2) Se asigna la clave subrogada para cada registro en base a la clave operacional: dim_uni_id.
- 3) Se cargan y actualizan los registros en la dimensión.

2.2.3.9. ETL de la Dimensión ayudas

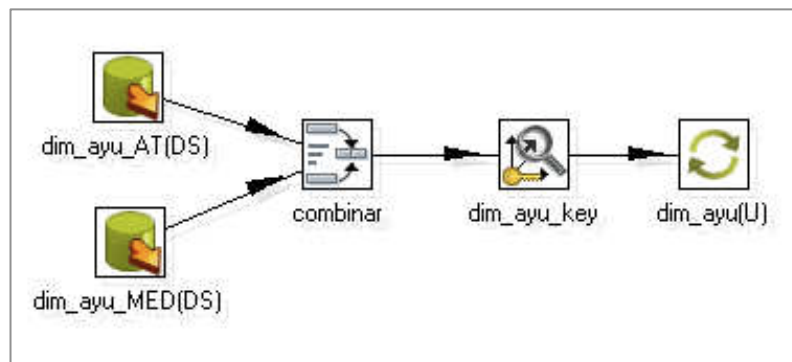


Figura 22. ETL de la Dimensión ayudas

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión ayudas, la extracción se realiza en dos etapas tomado en cuenta la transformación 10 conforme a lo detallado en la Tabla 52, la primera extrae datos sobre ayudas técnicas y la segunda extrae datos sobre medicamentos, scripts SQL ver Anexo B10 y B11.

- 2) Se combinan los registros obtenidos de las extracciones.
- 3) Se asigna la clave subrogada para cada registro en base a las claves operacionales:
 - dim_ayu_clase_id
 - dim_ayu_id
- 4) Se cargan y actualizan los registros en la dimensión.

2.2.3.10. ETL de la Dimensión estado ayuda

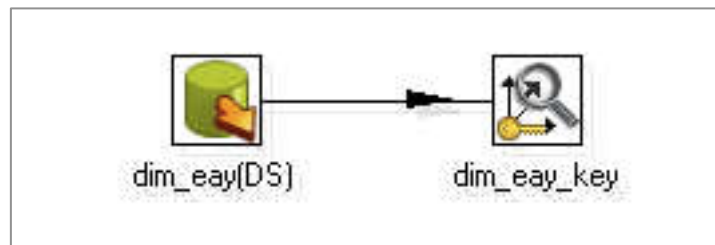


Figura 23. ETL de la Dimensión estado ayuda

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión estado ayuda, script SQL ver Anexo B12.
- 2) Se asigna la clave subrogada para cada registro en base a la clave operacional: dim_eay_id.
- 3) Se cargan y actualizan los registros en la dimensión.

2.2.3.11. ETL de la Dimensión entidad subvenciona

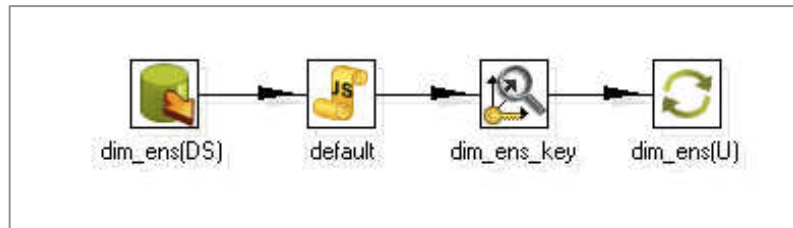


Figura 24. ETL de la Dimensión entidad subvenciona

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión entidad subvenciona, script SQL ver Anexo B13.
- 2) Se asigna la clave subrogada para cada registro en base a la clave operacional: dim_ens_id.
- 3) Se cargan y actualizan los registros en la dimensión.

2.2.3.12. ETL de la Dimensión entidad provee

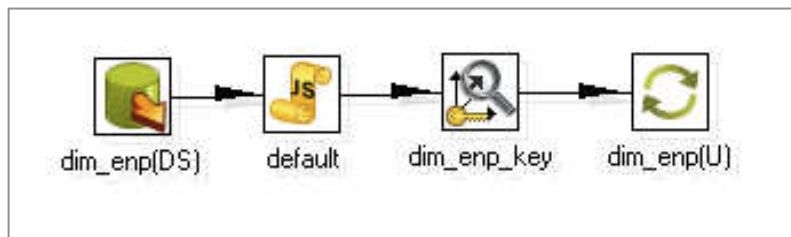


Figura 25. ETL de la Dimensión entidad provee

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión entidad provee, script SQL ver Anexo B14.
- 2) Se asigna la clave subrogada para cada registro en base a la clave operacional: dim_erp_id.
- 3) Se cargan y actualizan los registros en la dimensión.

2.2.3.13. ETL de la Dimensión proyecto

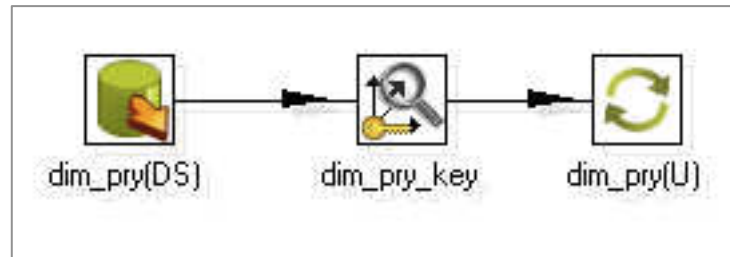


Figura 26. ETL de la Dimensión proyecto

- 1) Se extraen los datos conforme a lo establecido en la Tabla 43 para la dimensión proyecto, script SQL ver Anexo B15.
- 2) Se asigna la clave subrogada para cada registro en base a la clave operacional: dim_pry_id.
- 3) Se cargan y actualizan los registros en la dimensión.

2.2.4. ETL DE LAS TABLAS DE HECHOS¹

El flujo de datos base para la carga y actualización de las tablas de hechos consta de los siguientes pasos:

- 1) Extracción de datos de las tablas origen: Mediante código SQL se seleccionan las claves operacionales correspondientes a las dimensiones asociadas a la tabla de hechos, se seleccionan las columnas que proveen los datos para las medidas, los nombres de estas columnas se las redefine para que coincidan con las medidas de la tabla de hechos. En este paso se lleva a cabo una primera parte de las transformaciones esencialmente el cálculo de algunas medidas.

Los registros seleccionados están dentro de un rango de fechas, las cuales dependen del tipo de proceso ETL.

- 2) Transformaciones y cálculos: Se lleva a cabo las transformaciones y cálculos definidos en la Tabla 53, de ser necesario este paso añade columnas a los registros y realiza validaciones para comprobar la integridad de los cálculos, en caso de existir datos incoherentes estos son exportados a un archivo para su posterior análisis.
- 3) Obtención de claves dimensionales: Se busca reemplazar las claves operacionales con sus respectivas claves dimensionales, las claves dimensionales se añaden como nuevas columnas a los registros.

Las claves operacionales son removidas de los registros.

- 4) Validación de claves: Las claves obtenidas en el paso anterior son sometidos a una verificación para asegurar que no hay datos nulos, en caso de hallarse registros con claves nulas estos son exportados a un archivo para su posterior análisis. Los registros que pasan la verificación son ordenados en base a las claves dimensionales.

¹ The Data Warehouse ETL Toolkit, capítulo 6

- 5) Extracción de datos de la tabla de hechos: Mediante código SQL se seleccionan todas las columnas de la tabla y se las ordena por las claves dimensionales, los registros seleccionados están dentro de un rango de fechas, las cuales dependen del tipo de proceso ETL.
- 6) Comparación y Marcado de filas: Los registros de los pasos 4 y 5 son comparados en base a las claves dimensionales y dependiendo del resultado de la comparación se marca cada registro como: idéntico, alterado, nuevo, borrado; esta información se almacena en una nueva columna la cual se añade a los registros.
- 7) Actualización de Registros: Los registros del paso anterior son conmutados dependiendo de la marca, los nuevos son insertados, los alterados son actualizados y los borrados son eliminados de la tabla de hechos.

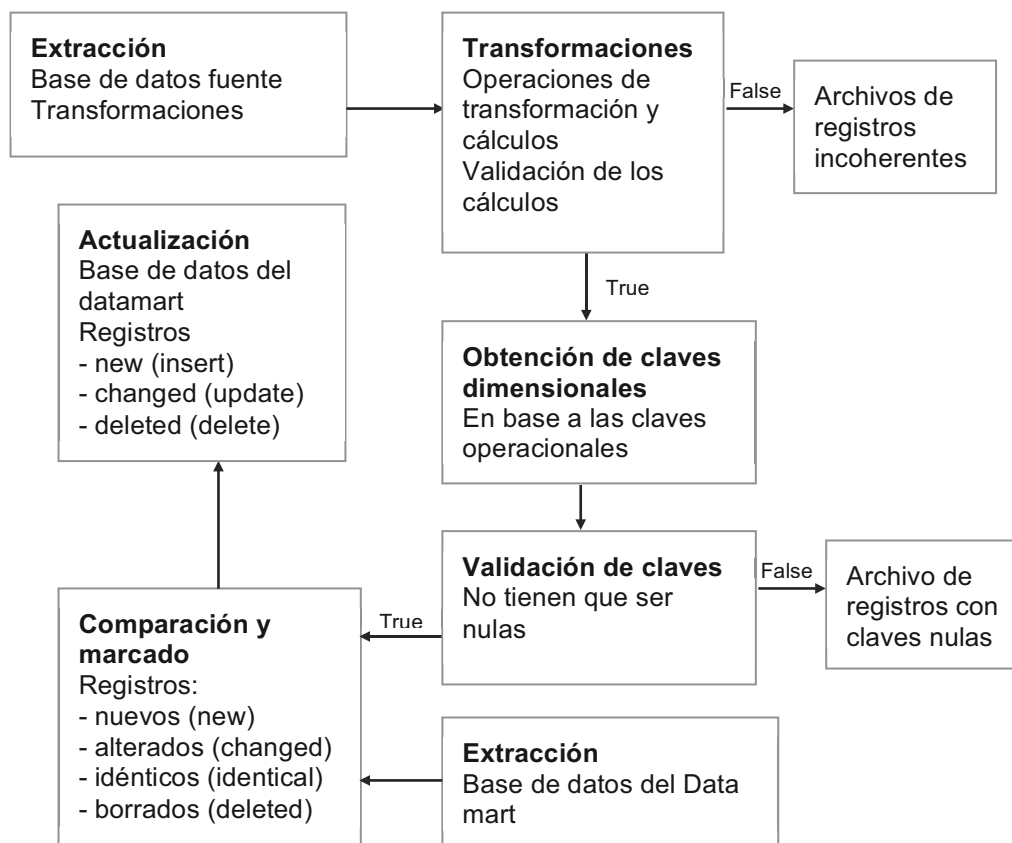


Figura 27. Diagrama del flujo de datos base para tablas de hechos

A continuación se detalla el flujo de datos ETL para cada tabla de hechos y su representación en la herramienta PDI - kettle.

Para mayor información sobre los pasos utilizados en cada flujo ETL ver el documento de Metadatos del proceso ETL que se ubica en el CD adjunto, ver Anexo D - /documentos/documentos_tecnicos/metadatos_etl/.

2.2.4.1. ETL de la Tabla de hechos de personas discapacitadas

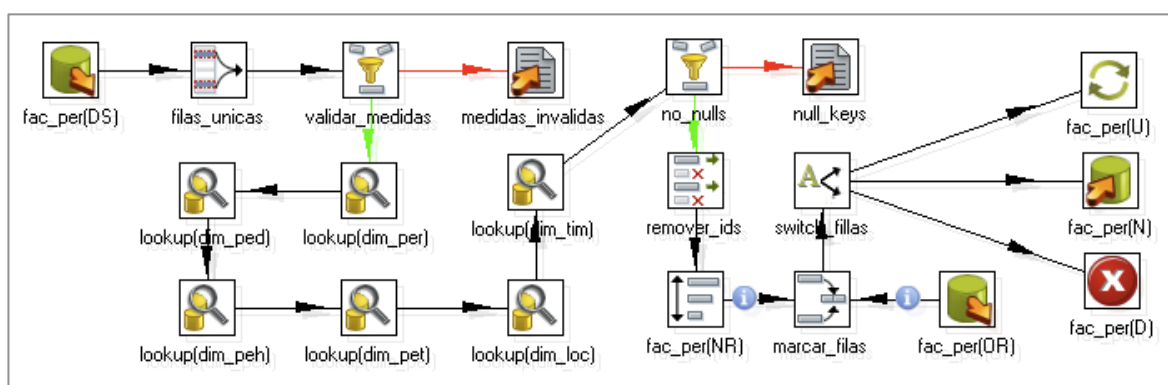


Figura 28. ETL de la Tabla de hechos de personas discapacitadas

- 1) Se extraen las claves operacionales de la fuente de datos, estas claves corresponden a las dimensiones asociadas a la tabla de hechos ver Tabla 12, se extraen los datos para las medidas de la tabla de hechos, script SQL ver Anexo B16.
- 2) Se valida que las medidas no sean nulas y que sean mayores que cero, los registros que no cumplan con la condición son enviados al archivo “fac_per_medidas_invalidos_yyyyMMdd.csv”.
- 3) Se busca las claves dimensionales en base a las claves operacionales.
- 4) Se comprueba que no existan registros con claves nulas, de ser el caso los registros son enviados al archivo “fac_per_null_key_yyyyMMdd.csv”. Se remueven las claves operacionales.
- 5) Se realiza la comparación de los registros con los existentes en la tabla de hechos (ver Anexo B17) y de acuerdo al resultado de la comparación son, insertados, actualizados o eliminados.

2.2.4.2. ETL de la Tabla de hechos de autorización de vehículos

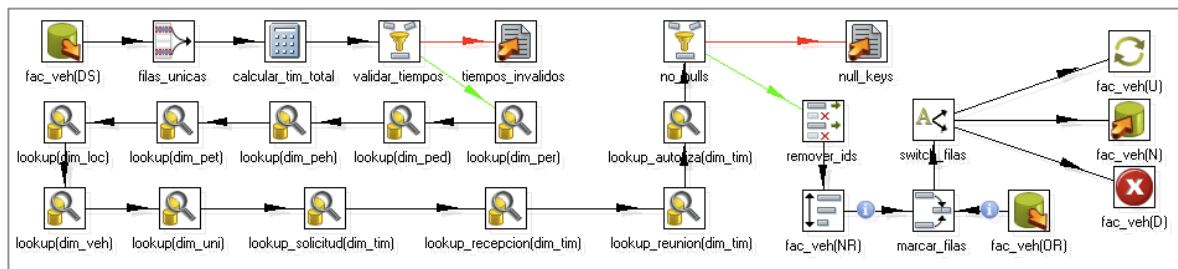


Figura 29. ETL de la Tabla de hechos de autorización de vehículos

- 1) Se extraen las claves operacionales de la fuente de datos, estas claves corresponden a las dimensiones asociadas a la tabla de hechos ver Tabla 12, se extraen los datos para las medidas de la tabla de hechos y se realizan las transformaciones 1, 2, 3 conforme a lo establecido en la Tabla 53, script SQL ver Anexo B18.
- 2) Se lleva a cabo la transformación 4 conforme a lo establecido en la Tabla 53 y se valida que las medidas calculadas sean mayores que cero, los registros que no cumplan con la condición son enviados al archivo “fac_veh_tiempos_invalidos_yyyyMMdd.csv”.
- 3) Se busca las claves dimensionales en base a las claves operacionales.
- 4) Se comprueba que no existan registros con claves nulas, de ser el caso los registros son enviados al archivo “fac_veh_null_key_yyyyMMdd.csv”. Se remueven las claves operacionales.
- 5) Se realiza la comparación de los registros con los existentes en la tabla de hechos (ver Anexo B19) y de acuerdo al resultado de la comparación son, insertados, actualizados o eliminados.

2.2.4.3. ETL de la Tabla de hechos de ayudas

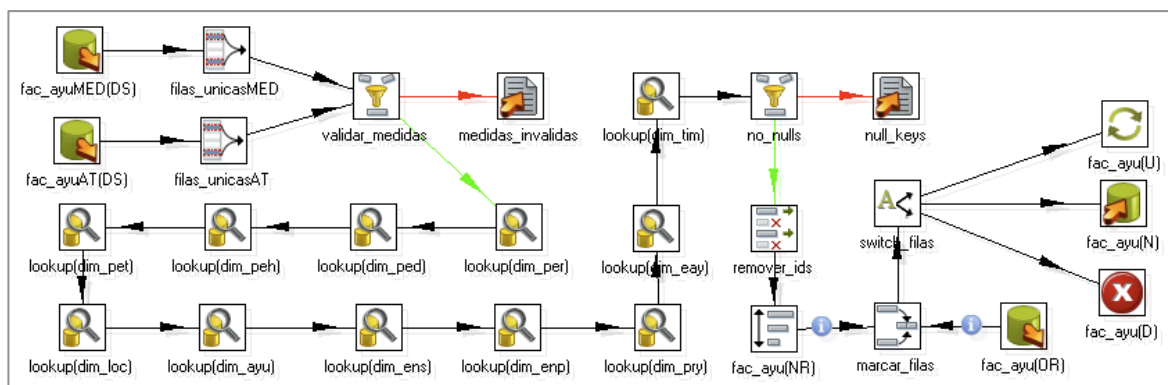


Figura 30. ETL de la Tabla de hechos de ayudas

- 1) Se extraen las claves operacionales de la fuente de datos, estas claves corresponden a las dimensiones asociadas a la tabla de hechos ver Tabla 12, se extraen los datos para las medidas de la tabla de hechos, la extracción se realizó en dos etapas, la primera para las ayudas técnicas y la segunda para los medicamentos o insumos, scripts SQL ver Anexo B20 y B21.
- 2) Se valida que las medidas no sean nulas y que sean mayores que cero, los registros que no cumplan con la condición son enviados al archivo “fac_ayu_medidas_invalidos_yyyyMMdd.csv”.
- 3) Se busca las claves dimensionales en base a las claves operacionales.
- 4) Se comprueba que no existan registros con claves nulas, de ser el caso los registros son enviados al archivo “fac_ayu_null_key_yyyyMMdd.csv”. Se remueven las claves operacionales.
- 5) Se realiza la comparación de los registros con los existentes en la tabla de hechos (ver Anexo B22) y de acuerdo al resultado de la comparación son, insertados, actualizados o eliminados.

2.2.5. CONSTRUCCIÓN DE LOS PROCESOS ETL

Como fue determinado en la selección de productos, la herramienta para llevar a cabo el proceso ETL es, Pentaho Data Integration – Kettle.

Se definieron dos conexiones de base de datos, una para acceder a la fuente de datos y otra para acceder a la base de datos del Data mart:

Tabla 54. Conexiones a base de datos		
Nombre	Descripción	Datos
datasource	Conexión a la base de datos fuente: sic Para extracción de datos (PostgreSQL)	Servidor: localhost Bdd: sic Puerto: 5432 Usuario: sic Clave: sic
datamart	Conexión a la base de datos: datamart_sic Para extracción y carga de datos (PostgreSQL)	Servidor: localhost Bdd: datamart_sic Puerto: 5432 Usuario: sic_analisis Clave: sicanalisis

Se construyo un Job para cada dimensión y tres Jobs para cada tabla de hechos, uno por cada proceso planeado como se definió en el plan ETL. Se construyó una Transformación para cada Flujo de datos tanto para dimensiones como para tablas de hechos, a continuación un ejemplo de Job para una dimensión y para una tabla de hechos.

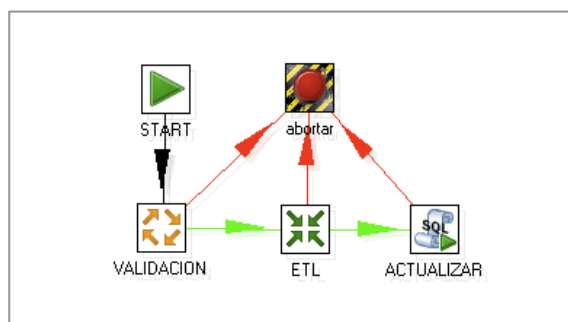


Figura 31. Trabajo para las Dimensiones

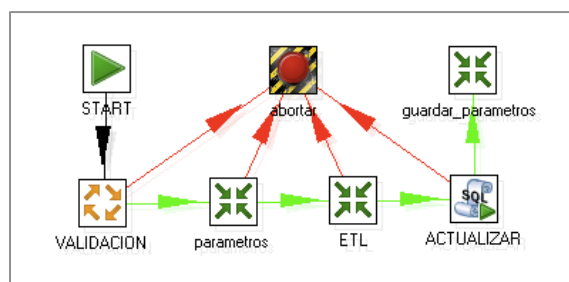


Figura 32. Trabajo para las Tablas de hechos

Los procesos definidos en el plan ETL fueron implementados mediante ficheros batch¹ los cuales realizan la ejecución de los Jobs.

El siguiente diagrama muestra la relación entre los componentes que conforman todo el proceso ETL.

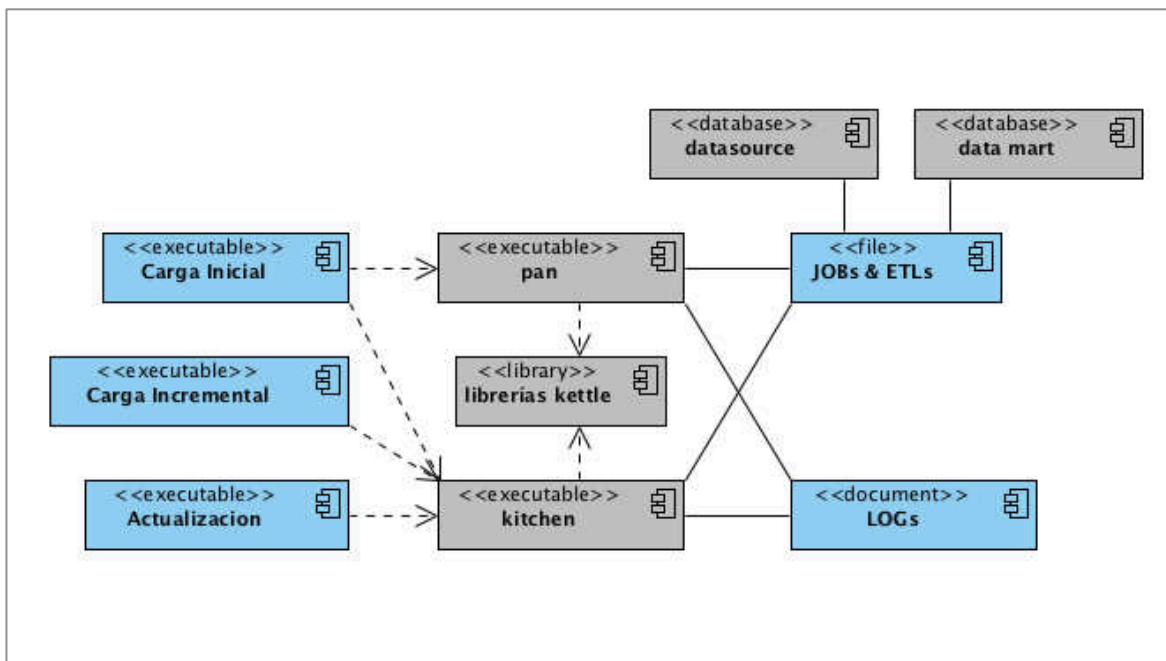


Figura 33. Diagrama de componentes para el proceso ETL

Carga Inicial: Archivo de procesamiento por lotes que implementa el proceso de carga inicial de datos.

Carga Incremental: Archivo de procesamiento por lotes que implementa el proceso de carga incremental de datos.

Actualizacion: Archivo de procesamiento por lotes que implementa el proceso de actualización de datos.

Pan²: Es un programa que permite ejecutar Transformaciones (flujos de datos) diseñadas en PDI – kettle.

Kitchen¹: Es un programa que permite ejecutar Jobs (Trabajos) diseñados en PDI – kettle.

¹ Archivo de procesamiento por lotes

² <http://wiki.pentaho.com/display/EAI/Pan+User+Documentation>

Librerías kettle: Conjunto de librerías jar necesarias para ejecutar los Jobs y las Transformaciones.

JOBs & ETLs: Conjunto de archivos XML generados mediante PDI – kettle, los cuales implementan los Trabajos y los flujos de datos respectivamente.

LOGs: Archivos de texto que registra todos los eventos concernientes a la ejecución de los Jobs.

Datasource: Base de datos fuente, sobre la cual se realiza la extracción de datos.

Data mart: Base de datos sobre la cual se realiza la carga de datos.

Los ejecutables y archivos resultantes de la implementación del proceso ETL se ubican en el CD adjunto, ver Anexo D – /codigo_fuente/etl/.

¹ <http://wiki.pentaho.com/display/EAI/Kitchen+User+Documentation>

2.3. CONSTRUCCIÓN DE LA APLICACIÓN DE USUARIO¹

Al iniciar esta etapa del ciclo de vida, dejamos prácticamente atrás todo lo referente al Back Room, y nos concentraremos en el Front Room.

La principal actividad de desarrollo que se lleva acabo en el Front Room es la creación de un primer grupo de reportes, los cuales adoptan el nombre de plantillas de la aplicación de usuario.

Plantillas de la aplicación de usuario.- Provee el diseño y la estructura de un reporte parametrizable, estos parámetros permiten al usuario generar docenas o cientos de reportes basados en la plantilla; es decir este conjunto de reportes sirven de ejemplo para la elaboración de nuevos reportes.

Un segundo esfuerzo esta enfocado en la creación de la aplicación Mapsic, por lo tanto las tareas que se realizaron en esta etapa fueron las siguientes:

- Definición de las plantillas de la aplicación de usuario
- Construcción de los reportes iniciales
- Desarrollo de la aplicación Mapsic

A continuación la figura 34, muestra la interacción entre los usuario y los sistemas del Front Room.

El usuario interactúa con el Data mart mediante los elementos del Front Room, es decir mediante herramientas de acceso a datos y aplicaciones desarrolladas sobre el Data mart. Como fue determinado en la selección de productos para el Front Room, la herramienta de acceso a datos es el sistema de reportes web Openi, adicionalmente se planteo el desarrollo de una pequeña aplicación GIS para realizar consultas interactivas.

¹ The Data Warehouse Lifecycle Toolkit, chapter 17 – Building End User Applications

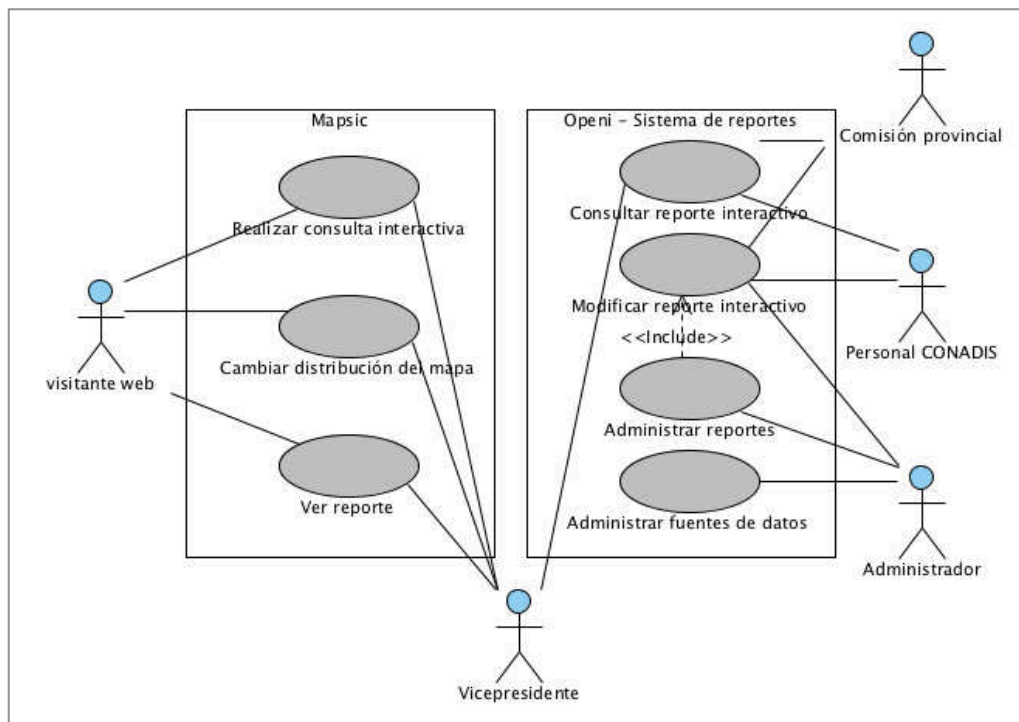


Figura 34. Escenario del Front Room

2.3.1. DEFINICIÓN DE LAS PLANTILLAS DE LA APLICACIÓN DE USUARIO

Las actividades realizadas en esta parte tuvieron como finalidad definir el grupo de plantillas de la aplicación de usuario a partir de un grupo de reportes candidatos, adicionalmente se determinó el árbol de navegación para los reportes.

2.3.1.1. Determinación del grupo de plantillas inicial

El primer paso fue identificar los reportes candidatos, como referencia se tomó los reportes de la sección de estadísticas del sitio web del CONADIS, y los requerimientos de análisis.

Lista de Reportes candidatos

- Tipo de deficiencia mas frecuente en niños de 5 años
- Causa de la deficiencia que originan discapacidades en menores de 5 años
- Discapacidad mas frecuente en mayores de 5 años
- Porcentaje de personas con discapacidad por sexo, región y sector

- Porcentaje de personas con discapacidad por provincias
- Personas con discapacidad y categoría ocupacional
- Distribución de ayudas técnicas por provincias
- Tiempo promedio para la autorización de vehículos
- Total de vehículos autorizados en el ultimo año
- Distribución de las personas con discapacidad que trabajan
- Distribución de Personas con discapacidad con deficiencia, limitaciones moderadas y graves.

El siguiente paso fue buscar similitudes entre los reportes con la finalidad de definir verdaderas plantillas y agrupar los reportes en función de los elementos que presentan, el resultado es la siguiente lista de reportes:

Análisis cruzado

- Causa de deficiencia por edades
- Distribución local de PCD¹ por situación laboral
- Distribución local de ayudas técnicas

Distribución por categoría

- Distribución de PCD por deficiencia principal
- Tiempo promedio para autorizar un vehículo

Promedio por categoría

- Porcentaje de PCD por genero

Tendencias por categoría

- PCD registradas por año

¹ personas con discapacidad

- Monto total en ayudas entregadas al año
- Vehículos autorizados al año

2.3.1.2. Especificación detallada de las plantillas

Tabla 55. Reporte plantilla - Distribución de PCD por deficiencia principal	
Descripción	Esta plantilla genera un reporte que totaliza el número de personas discapacitadas para cada deficiencia principal, permite detectar máximos y mínimos.
Parámetros por defecto	<i>Columna:</i> Medida Personas <i>Filas:</i> Deficiencia principal <i>Filtros:</i> Persona – estado (estado=A)
Medidas	Personas
Modificaciones	La fila puede ser cambiada por cualquier otra categoría, por ejemplo Grado de discapacidad, Persona – Edad, etc. Se puede aplicar cualquier tipo de filtro.

Tabla 56. Reporte plantilla - Porcentaje de PCD por genero	
Descripción	Esta plantilla genera un reporte que muestra el porcentaje de personas discapacitadas por genero, permite detectar predominancia.
Parámetros por defecto	<i>Columna:</i> Porcentaje <i>Filas:</i> Persona - Genero <i>Filtros:</i> Persona – estado (estado=A)
Medidas	Personas
Modificaciones	La fila puede ser cambiada por cualquier otra categoría, por ejemplo Grado de discapacidad, Persona – Edad, etc. Se puede aplicar cualquier tipo de filtro.
Nota	El Porcentaje se calcula mediante código MDX

Tabla 57. Reporte plantilla - Distribución local de PCD por situación laboral	
Descripción	Esta plantilla genera un reporte que muestra la cantidad de personas discapacitadas que trabajan y las que no trabajan, distribuidas por localidad: provincia, cantón, parroquia.
Parámetros por defecto	<i>Columna:</i> Medida Personas, Persona - Trabajo <i>Filas:</i> Localidad <i>Filtros:</i> Persona – estado (estado=A)
Medidas	Personas
Modificaciones	La categoría de la columna puede ser cambiada por cualquier otra categoría, por ejemplo Deficiencia principal, Persona – edad, Persona – Genero, etc. Se puede aplicar cualquier tipo de filtro. Se pueden seleccionar valores específicos de la categoría Localidad para detallar.

Tabla 58. Reporte plantilla - Causa de deficiencia mas frecuente por edades	
Descripción	Esta plantilla genera un reporte cruzado que muestra la cantidad de personas discapacitadas para cada intercesión entre las categorías Causa de la deficiencia y Persona - Edad, distribuidas, permite detectar máximos, mínimos y realizar comparaciones
Parámetros por defecto	<i>Columna:</i> Medida Personas, Persona - Edad <i>Filas:</i> Causa de la deficiencia <i>Filtros:</i> Persona – estado (estado=A)
Medidas	Personas
Modificaciones	La categoría de la columna o de la fila pueden ser cambiadas por cualquier otra categoría, por ejemplo Persona – Estado civil, Persona – Nivel de instrucción, Persona – Genero, etc. Se puede aplicar cualquier tipo de filtro.

Tabla 59. Reporte plantilla - PCD registradas por año	
Descripción	Esta plantilla genera un reporte que muestra el total de personas discapacitadas que se registraron por año en las provincias.
Parámetros por defecto	<i>Columna:</i> Medida Personas, Tiempo carnetizacion <i>Filas:</i> Localidad <i>Filtros:</i> Persona – estado (estado=A)
Medidas	Personas
Modificaciones	La fila puede ser cambiada por cualquier otra categoría, por ejemplo Persona – Genero, Persona – Edad, etc.

Tabla 60. Reporte plantilla - Distribución local de ayudas técnicas	
Descripción	Esta plantilla genera un reporte que muestra la cantidad de ayudas del tipo ayuda técnica distribuidas por localidad: provincia, cantón, parroquia.
Parámetros por defecto	<i>Columna:</i> Medida Ayudas, Ayuda (Ayudas técnicas) <i>Filas:</i> Localidad <i>Filtros:</i> Estado ayuda (estado=Entregado)
Medidas	Ayudas, Monto, Subvención
Modificaciones	La categoría de la columna puede ser cambiada por cualquier otra categoría, por ejemplo Medicamento, Entidad provee, Entidad subvenciona etc. Se puede aplicar cualquier tipo de filtro. Se pueden seleccionar valores específicos de la categoría Localidad para detallar.

Tabla 61. Reporte plantilla - Monto total en ayudas entregadas al año	
Descripción	Esta plantilla genera un reporte que muestra el monto total en ayudas que se entregaron por año en las provincias.
Parámetros por defecto	<i>Columna:</i> Medida Monto, Tiempo entrega <i>Filas:</i> Localidad <i>Filtros:</i> Estado ayuda (estado=Entregado)
Medidas	Ayudas, Monto, Subvención
Modificaciones	La fila puede ser cambiada por cualquier otra categoría, por ejemplo Persona – Genero, Deficiencia principal, Entidad provee, etc.

Tabla 62. Reporte plantilla - Vehículos autorizados al año	
Descripción	Esta plantilla genera un reporte que muestra el total de vehículos autorizados por año en las provincias.
Parámetros por defecto	<i>Columna:</i> Medida Vehículos, Tiempo autorización <i>Filas:</i> Localidad
Medidas	Vehículos, Recepción(días), Atención(días), Decisión(días), Total(días)
Modificaciones	La fila puede ser cambiada por cualquier otra categoría, por ejemplo Vehículo, Unidad autorizada, etc.

Tabla 63. Reporte plantilla - Tiempo promedio para autorizar un vehículo	
Descripción	Esta plantilla genera un reporte que muestra el tiempo promedio en días que toma el proceso de autorización de vehículos dependiendo del tipo de vehículo.
Parámetros por defecto	<i>Columna:</i> Medida Total(días) <i>Filas:</i> Vehículo
Medidas	Vehículos, Recepción(días), Atención(días), Decisión(días), Total(días)
Modificaciones	La fila puede ser cambiada por cualquier otra categoría, por ejemplo Localidad, Unidad autorizada La medida puede ser cambiada por Recepción(días), Atención(días), Decisión(días)

2.3.1.3. Diseño del árbol de navegación para los reportes

El objetivo principal es agrupar los reportes para facilitar y controlar el acceso, se considero la siguiente jerarquía para la organización de los reportes

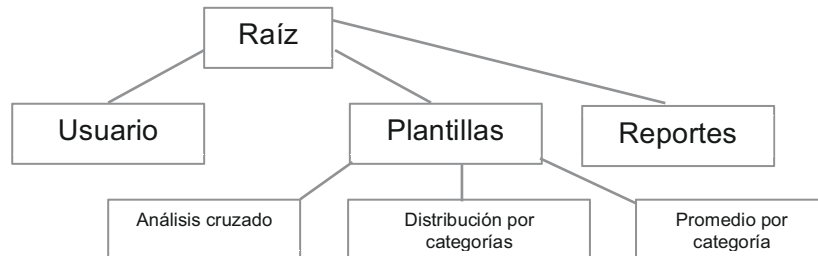


Figura 35. Árbol de navegación para los reportes

Primer nivel: Es el nodo raíz.

Segundo nivel: Agrupa los reportes en base a los permisos de acceso, el nodo “Plantilla” contiene los reportes iniciales, el nodo “Reportes” contiene reportes de acceso publico y el nodo “Usuario” contiene los reportes propios del usuario.

Tercer nivel: Los reportes plantilla se agrupan en base al tipo de análisis que representan.

Cuarto nivel en adelante: Estos niveles esta a criterio del administrador y los usuarios.

2.3.2. CONSTRUCCIÓN DE LOS REPORTE INICIALES

Finalmente los reportes plantilla son implementados bajo la herramienta de acceso a datos seleccionada, es decir bajo Openi.

El esquema de navegación de los reportes es implementado y viene a ser el menú para la selección de reportes en Openi.

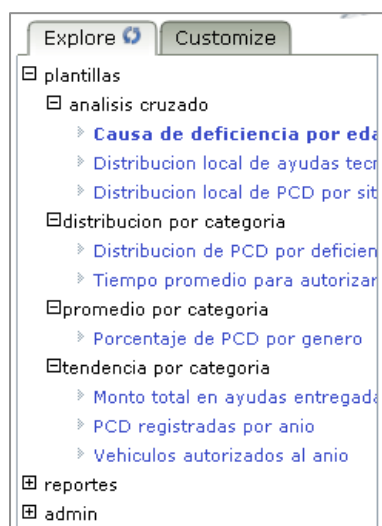


Figura 36. Árbol de navegación de reportes en Openi

Los reportes se construyeron en base a las especificaciones detalladas para cada plantilla, el código MDX generado para cada reporte se detalla en el Anexo C.

A continuación las Figuras 37, 38 y 39 muestran algunos ejemplos de los reportes corriendo sobre Openi.

Data

Chart

Drill :
☐ Hierarchy
☐ Replace
☐ Data Report
(Slicer: Persona - Estado->estado=A)

	Measures					
	Personas					
	Persona - Edad					
Causa de la deficiencia	Adolescente (12-17)	Adulto (26-59)	Adulto mayor (60)	Infante (0-6)	Joven (18-25)	Niño (7-11)
ENFERMEDAD ADQUIRIDA	1.952	24.826	19.218	353	3.260	917
ACCIDENTE DE TRANSITO	100	4.002	1.655	13	306	36
ACCIDENTE DE TRABAJO	14	2.858	1.538	7	104	4
ACCIDENTE DOMESTICO	256	3.418	1.764	30	518	103
ACCIDENTE DEPORTIVO	11	172	83	4	25	6
VIOLENCIA	24	1.107	334	1	87	6
DESASTRES NATURALES / GUERRA	6	39	22	2	7	9
CONGENITO / GENETICO	7.516	20.193	3.618	2.290	8.940	5.502
PROBLEMAS DE PARTO	1.909	3.298	190	459	2.385	1.226
TRASTORNOS NUTRICIONALES	10	21	19	3	18	8
NO SE SABE	610	2.937	1.542	75	950	268

Figura 37. Reporte - Causa de deficiencia por edades en Openi

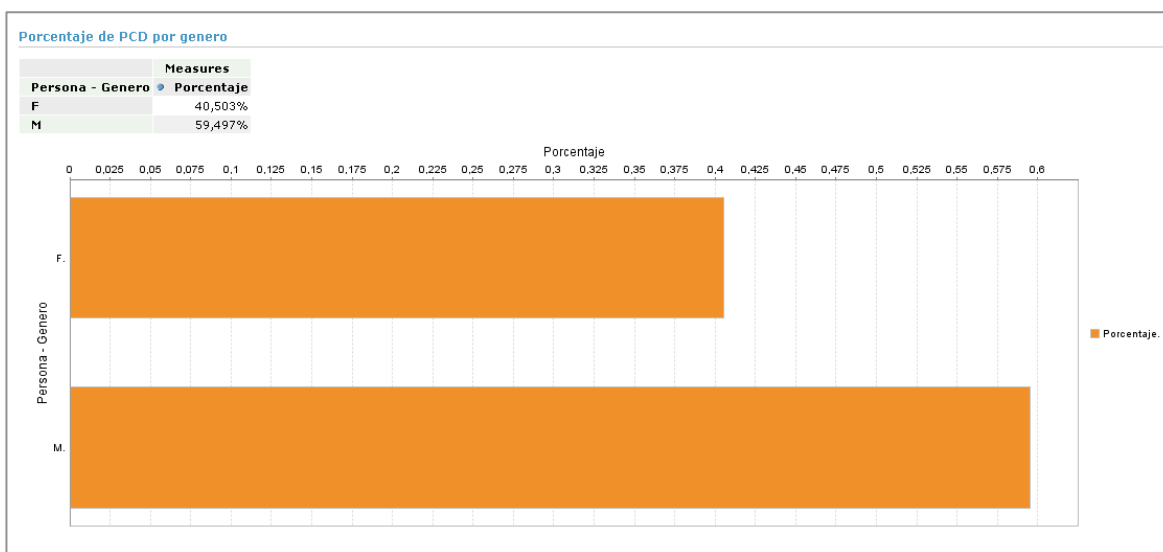


Figura 38. Reporte – Porcentaje de PCD por genero, vista de impresión

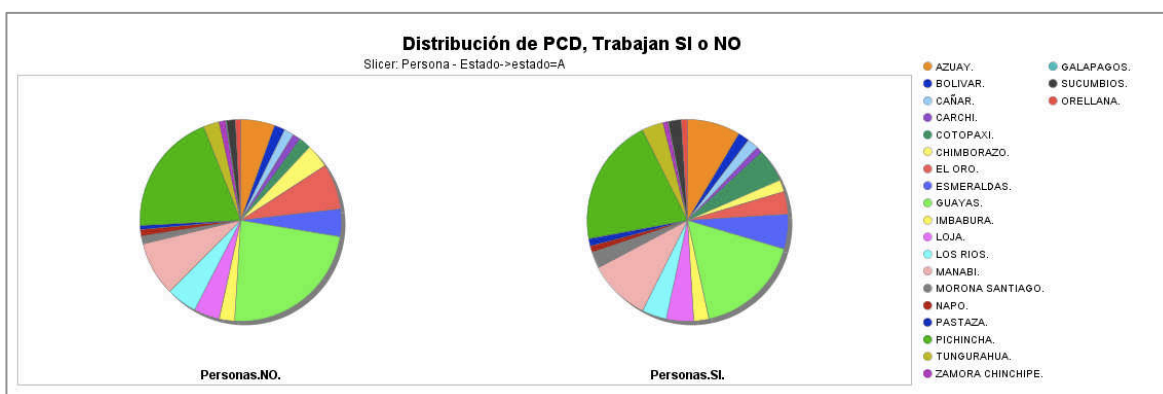


Figura 39. Reporte – Distribución de PCD por su situación laboral, gráfico

2.3.3. DESARROLLO DE LA APLICACIÓN - MAPSIC

Esta aplicación web tiene como propósito permitir a los usuarios realizar consultas interactivas, obtener información sobre el número y porcentaje de personas discapacitadas y su distribución provincial y cantonal en el Ecuador.

Los reportes están basados en las plantillas:

- Distribución de PCD por deficiencia principal, para reportes de distribución por categoría.
- Distribución local de PCD por situación laboral, para reportes cruzados entre Localidad y el resto de categorías.

Las categorías seleccionadas como parámetros para estos reportes son las siguientes:

- Persona – Genero
- Persona – Edad
- Deficiencia principal
- Causa de la deficiencia
- Localidad

La distribución por provincias y cantones se representa gráficamente en un mapa, la distribución puede ser:

Distribución por grupo de colores: La información se distribuye en 4 colores diferentes, cada color representa un rango de valores que va desde el número máximo de personas discapacitadas hasta el mínimo.

Distribución por escala de un solo color: La información se distribuye en tonalidades, desde blanco (255,255,255) para el valor mínimo y rojo (255,0,0) para el valor máximo del número de personas discapacitadas.

2.3.3.1. Especificación de los casos de uso

Tabla 64. Especificación del caso de uso: Realizar consultar interactiva		
Descripción general: Genera reportes en base a una serie de parámetros seleccionados por el usuario.		
Actores: Visitante web		
Pre-Condiciones: Existen datos en las tablas del Data mart, existen la información vectorial de los mapas.		
Pos-Condiciones: El usuario puede visualizar los reportes de distribución por categorías		
Flujo Base o Principal		
Paso	Actor(es)	Sistema
FB10	El visitante web asigna valores a los parámetros de consulta, seguidamente da clic en el botón “Consultar” o en una de las localidades del mapa.	
FB20		Se deshabilita el botón “Consultar” y el mapa mientras se generan los reportes y el mapa correspondiente, finalizada esta operación se refrescan los reportes de distribución por categorías y el mapa. Se habilita el botón “Consultar” y el mapa.

Tabla 65. Especificación del caso de uso: Cambiar la distribución del mapa		
Descripción general: Cambia la distribución grafica de datos en el mapa, dependiendo del tipo de distribución que seleccione el usuario; por defecto es ninguna.		
Actores: Visitante web		
Pre-Condiciones: Ejecutar la operación - Realizar consulta interactiva.		
Pos-Condiciones: El usuario puede visualizar la distribución gráfica seleccionada en el mapa.		
Flujo Base o Principal		
Paso	Actor(es)	Sistema
FB10	El visitante web selecciona el tipo de distribución.	
FB20		Se deshabilita el mapa mientras se genera el mapa acorde a la distribución seleccionada, finalizada esta operación se refresca el mapa y se habilita.

Tabla 66. Especificación del caso de uso: Ver reporte		
Descripción general: Permite cargar el reporte cruzado que seleccione el usuario.		
Actores: Visitante web		
Pre-Condiciones: Ejecutar la operación - Realizar consulta interactiva.		
Pos-Condiciones: El usuario puede visualizar el reporte cruzado que seleccionó.		
Flujo Base o Principal		
Paso	Actor(es)	Sistema
FB10	El visitante web a la sección de Reportes y da clic sobre el reporte que desea ver.	
FB20		Se despliega un indicador de carga mientras se arma el reporte cruzado, finalizada esta operación se esconde el indicador de carga y se muestra el reporte.

2.3.3.2. Arquitectura de la aplicación - Mapsic

Corresponde a un modelo en cuatro capas

Tabla 67. Capas de la aplicación - Mapsic	
Capa	Descripción
Presentación	Interfaz de usuario estilo web, utilizando xhtml, css y javascript. Presenta ayudas visuales tales como: resaltado de filas y columnas, ordenamiento por columnas, en los reportes. Paneles expandibles e indicadores de carga.
Comunicación	Las llamadas al servidor son asincrónicas, el refrescamiento del contenido de la pagina web es mediante Ajax ¹ .
Lógica	Procedimientos para la generación de reportes, procedimientos para la generación de mapas (GIS). Definición de constantes y variables de sesión.
Datos	Procedimientos para acceder y realizar consultas a bases de datos. Procedimientos para acceder a shape files (mapas vectoriales). Definición de metadatos

¹ Asynchronous JavaScript And XML, ver Glosario

2.3.3.3. Diseño de interfaces de usuario

La estructura base de la página web esta dividida en 2 bloques:

Cabecera: Contiene los logos del CONADIS y de la aplicación, los parámetros de consulta, y los parámetros de filtro agrupados en un panel expandible.

La valores para cada parámetro están agrupados en combos de selección.

Cuerpo: Contiene las secciones de Mapa y Reportes, las secciones se diseñaron por pestañas, cada sección es una pestaña, la sección por defecto es la de Mapa.

Los reportes están contenidos en paneles expandibles.

Las Figuras 40, 41 y 42 muestran la distribución del contenido y los objetos en la página web.

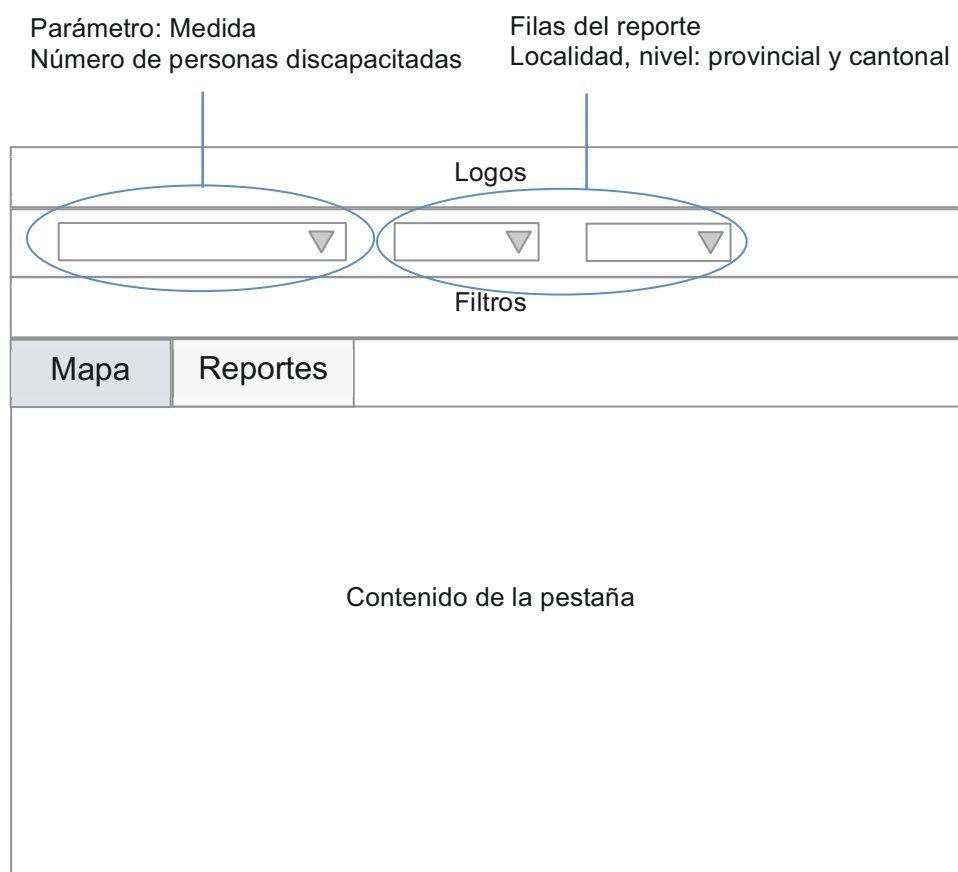


Figura 40. Diseño de la pagina web de la aplicación - Mapsic

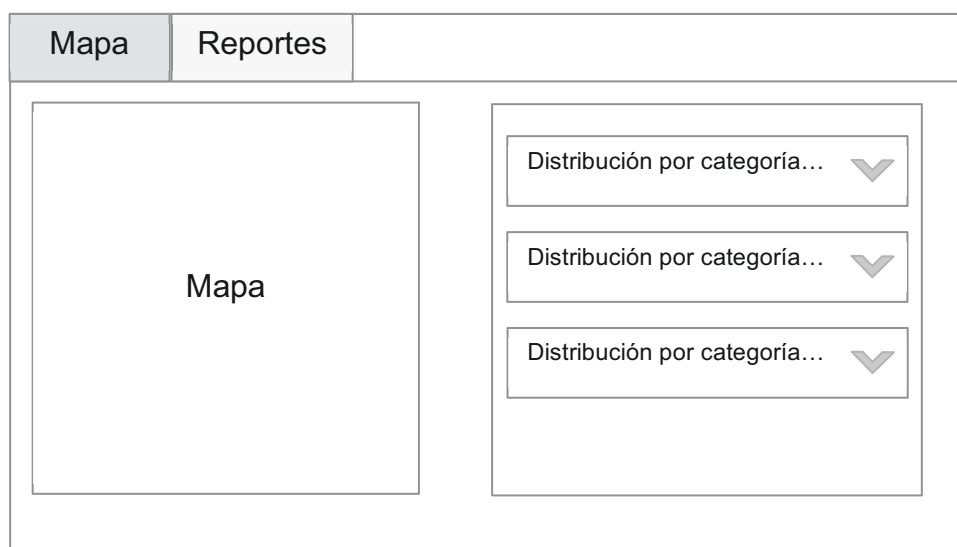


Figura 41. Contenido de la pestaña Mapa

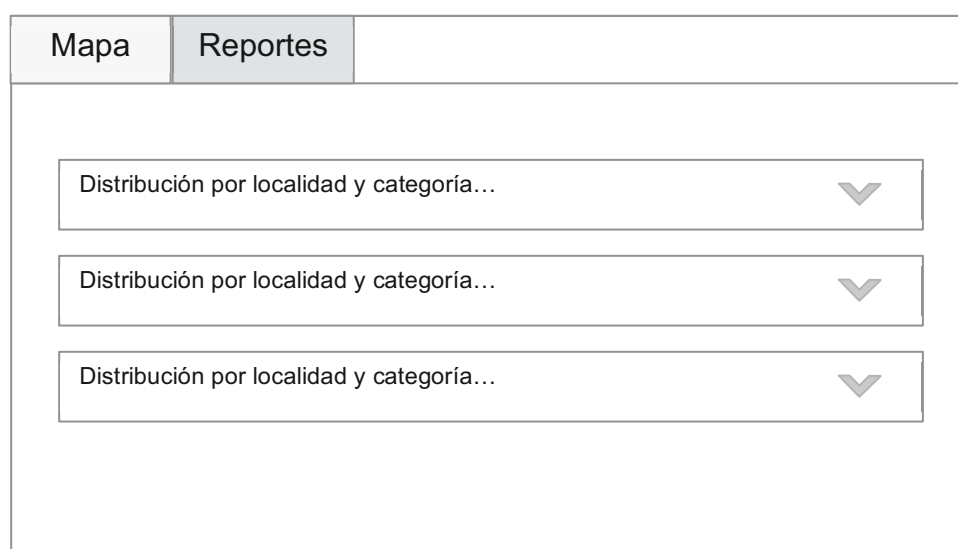


Figura 42. Contenido de la pestaña Reportes

2.3.3.4. Diagrama de secuencia de la aplicación - Mapsic

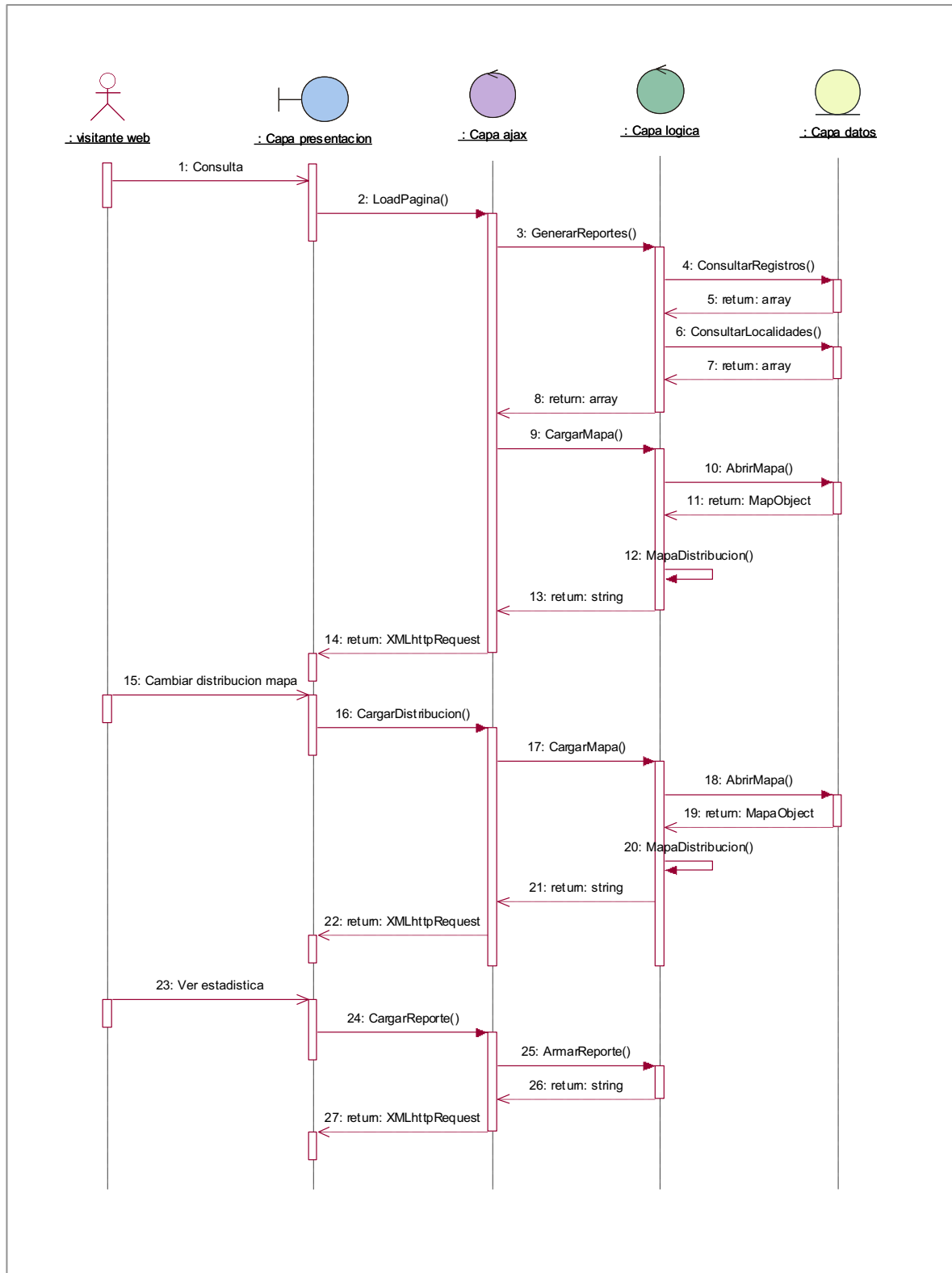


Figura 43. Diagrama de secuencia para la aplicación - Mapsic

2.3.3.5. Construcción de la aplicación - Mapsic

El siguiente diagrama muestra la relación que existe entre los componentes de la aplicación, el código fuente de cada componente se encuentra en el CD adjunto, ver Anexo D - /codigo_fuente/mapsic/.

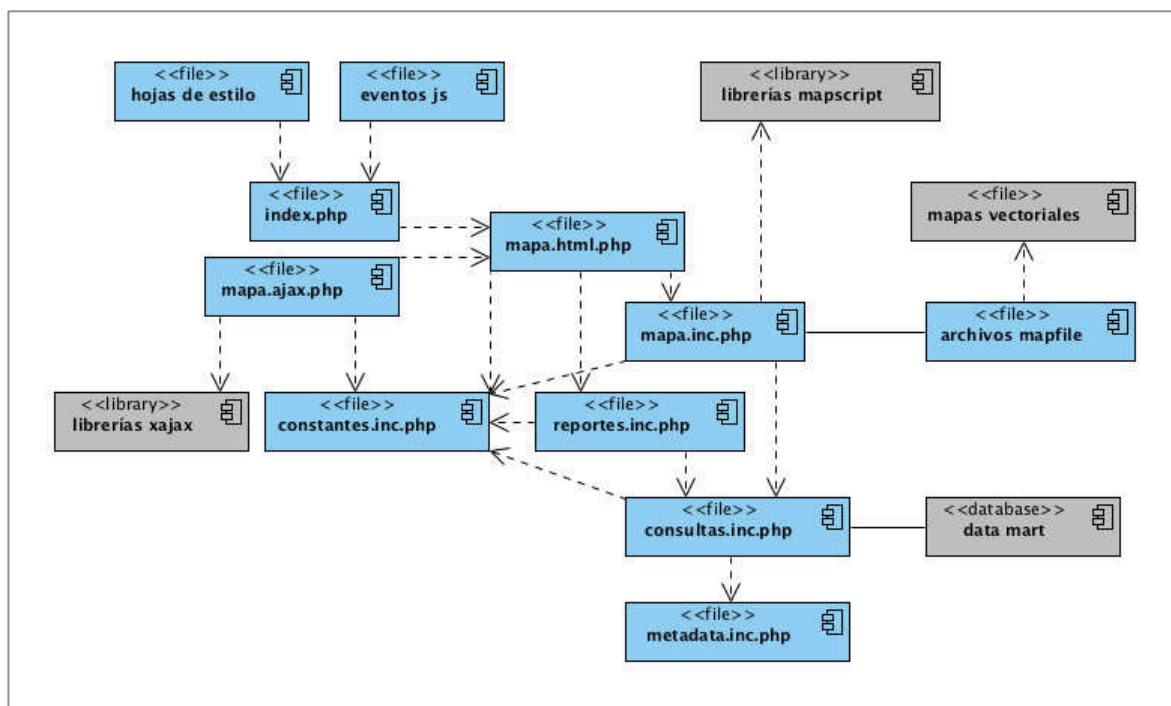


Figura 44. Diagrama de componentes de la aplicación - Mapsic

Index.php: Este archivo es la página por defecto de la aplicación, define la estructura de la pagina web mediante etiquetas xhtml¹, realiza llamadas a funciones en lenguaje php² para presentar los controles de la interfaz.

Mapa.ajax.php: Este archivo contiene todas las funciones necesarias para responder a las demandas de datos realizadas mediante Ajax.

Mapa.html.php: Este archivo contiene todas las funciones necesarias para crear los controles y otros elementos de la interfaz de usuario, entre estos se cuentan los combos de selección, imágenes, formularios.

Mapa.inc.php: Este archivo contiene todas las funciones necesarias para la creación y modificación de mapas mediante la inclusión de la librería mapscript.

¹ eXtensible Hypertext Markup Language, ver Glosario

² PHP Hypertext Pre-processor, ver Glosario

Reportes.inc.php: Este archivo contiene todas las funciones necesarias para la generación de reportes y mostrarlos en tablas.

Consultas.inc.php: Este archivo contiene todas las funciones necesarias para realizar conexiones y consultas a la base de datos.

Constantes.inc.php: Archivo de constantes de php que define datos numéricos y de texto, esta información esta disponible para todas las funciones a lo largo de la aplicación.

Metadata.inc.php: Archivo de constantes de php que define los datos de conexión, los nombres de tablas y columnas de la base de datos, esta información es utilizada para realizar las consultas.

Eventos.js: Grupo de archivos escritos en javascript¹ los cuales implementan funciones para manipular el DOM² de la pagina web, algunas de estas funciones incluyen expandir y colapsar paneles, resaltar celdas en las tablas, habilitar y deshabilitar controles, realizar llamadas Ajax.

Hojas de estilo: Archivos escritos en lenguaje css³ usados para definir la presentación de los elementos de la página web.

Archivos Mapfile⁴: Archivo que define la relación entre los objetos de MapServer, indica donde se encuentran los datos vectoriales e indica como se dibujaran los objetos.

¹ Definición, ver Glosario

² Document Object Model, ver Glosario

³ Cascading Style Sheets, ver Glosario

⁴ <http://mapserver.org/mapfile/index.html>

2.4. DESPLIEGUE¹

El despliegue es la convergencia de tecnología, datos y aplicaciones, es la integración de los componentes que se han ido desarrollando al transcurso del ciclo de vida del Data mart, con el fin de que los usuario puedan empezar a sacar provecho.

Claro que esta etapa también contempla capacitación para los usuarios y la preparación del ambiente para poder integrar cada uno de los componentes del Data mart.

Antes de colocar al Data mart en producción, es necesario llevarlo a través de un periodo de pruebas alfa y pruebas beta.

Periodo de pruebas alfa.- El objetivo de este periodo de pruebas es comprobar el correcto funcionamiento de cada uno de los componentes del Data mart, incluyendo la infraestructura a ser instalada, los procesos ETL, el rendimiento y las aplicaciones de usuario.

Estas pruebas se llevan a cabo utilizando una copia de las fuentes de datos que están en producción, de tal manera que los resultados reflejen el comportamiento del Data mart como si estuviera en producción, así se puede asegurar la calidad de los datos y las aplicaciones, este periodo es iterativo de hecho se puede extender tanto como sea necesario para asegurar calidad.

Periodo de pruebas beta.- Durante este periodo un grupo limitado de usuario tiene acceso al Data mart, el objetivo es llevar acabo pruebas con los usuarios, para comprobar la calidad y completitud de los datos, calidad de las aplicaciones de usuario y su relevancia para la organización. Este periodo adopta el nombre de liberación piloto.

El manual de usuario y el manual de instalación se ubican en el CD adjunto, ver Anexo D - /documentos/manuales/.

A continuación se detallan las pruebas realizadas sobre los procesos ETL, los reportes iniciales y la aplicación Mapsic.

¹ The Data Warehouse Lifecycle Toolkit, chapter 18 – Planning the Deployment

2.4.1. PRUEBAS SOBRE LOS PROCESOS ETL

El objetivo de estas pruebas es comprobar el correcto funcionamiento de los procesos de carga y actualización de datos.

Y comprobar que los procesos para validar las dependencias funcionan correctamente.

2.4.1.1. Casos de prueba para la Carga inicial

Tabla 68. Procedimiento de prueba para la Carga inicial

- 1) Dirigirse a la carpeta donde se instaló la herramienta “Data integration - kettle” (<kettle>)
- 2) Ir al siguiente directorio: <kettle>/datamart_sic_etl/ejecutables/
- 3) Ejecutar el archivo “Carga Inicial.bat”

Tabla 69. Caso de prueba 1 - Carga Inicial

Descripción general:	Comprobar que el proceso de carga inicial funciona correctamente.
Datos de Entrada:	----
Resultados Esperados:	<ul style="list-style-type: none"> - Se crea la base de datos - Se cargan datos sobre las tablas - Se crean los archivos de parámetros
Condiciones para esta prueba:	La base de datos datamart_sic no existe y no hay conexiones abiertas.
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	Se obtuvo los resultados esperados

Tabla 70. Caso de prueba 2 - Carga Inicial	
Descripción general:	Comprobar que el proceso no puede continuar si la base de datos existe y hay conexiones abiertas.
Datos de Entrada:	----
Resultados Esperados:	Se produce un error al ejecutar el trabajo de creación de la base de datos.
Condiciones para esta prueba:	La base de datos datamart_sic ya existe y hay conexiones abiertas.
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	Se produjo el error, la base de datos no puede ser eliminada dado que existen conexiones abiertas, por lo tanto el proceso de creación no puede continuar.

2.4.1.2. Casos de prueba para la Carga Incremental

Tabla 71. Procedimiento de prueba para la Carga inicial
<ol style="list-style-type: none"> 1) Dirigirse a la carpeta donde se instalo la herramienta "Data integration - kettle" (<kettle>) 2) Ir al siguiente directorio: <kettle>/datamart_sic_etl/ejecutables/ 3) Ejecutar el archivo "Carga Incremental.bat"

Tabla 72. Caso de prueba 1 - Carga Incremental	
Descripción general:	Comprobar que el proceso de carga incremental funciona correctamente.
Datos de Entrada:	----
Resultados Esperados:	<ul style="list-style-type: none"> - Se cargan y/o actualizan datos en las tablas de dimensiones - Se cargan nuevos registros a las tablas de hechos
Condiciones para esta prueba:	<ul style="list-style-type: none"> - Se debe haber ejecutado la Carga Inicial al menos una vez - Deben existir los archivos de parámetros
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera

Resultados Obtenidos:	Se obtuvo los resultados esperados
-----------------------	------------------------------------

2.4.1.3. Casos de prueba para la Actualización

Tabla 73. Procedimiento de prueba para la Actualización

- 1) Dirigirse a la carpeta donde se instalo la herramienta “Data integration - kettle” (<kettle>)
- 2) Ir al siguiente directorio: <kettle>/datamart_sic_etl/ejecutables/
- 3) Ejecutar el archivo “Actualizacion.bat”

Tabla 74. Caso de prueba 2 – Actualización

Descripción general:	Comprobar que el proceso de actualización de dimensiones funciona correctamente.
Datos de Entrada:	Opción = 1
Resultados Esperados:	Se cargan y/o actualizan datos en las tablas de dimensiones
Condiciones para esta prueba:	Se debe haber ejecutado la Carga Inicial al menos una vez
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	Se obtuvo los resultados esperados

Tabla 75. Caso de prueba 2 – Actualización

Descripción general:	Comprobar que el proceso de actualización de tablas de hechos funciona correctamente.
Datos de Entrada:	Opción = {2,3,4,5} Fecha inicial = yyyyMMdd Fecha final = yyyyMMdd Tal que, yyyyMMdd es una fecha valida donde yyyy es el año expresando en cuatro cifras, MM el mes comprendido entre {1-12}, dd el día comprendido entre {1-31}
Resultados Esperados:	Se actualizan los datos en las tablas de hechos

Condiciones para esta prueba:	Se debe haber ejecutado la Carga Inicial al menos una vez
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	Se obtuvo los resultados esperados

Tabla 76. Caso de prueba 3 – Actualización	
Descripción general:	Comprobar que el proceso de actualización de tablas de hechos produce un error si se ingresa mal cualquiera de las fechas provistas.
Datos de Entrada:	<p>Opción = {2,3,4,5}</p> <p>Fecha inicial = dato</p> <p>Fecha final = dato</p> <p>Tal que, “dato” representa una cadena alfa numérica que no corresponde a una fecha con formato yyyyMMdd</p>
Resultados Esperados:	Se produce un error con ID:1, y se debe revisar el log
Condiciones para esta prueba:	Se debe haber ejecutado la Carga Inicial al menos una vez
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	Se obtuvo los resultados esperados

2.4.1.4. Casos de prueba para la validación de dependencias

Tabla 77. Procedimiento de prueba para la validación de dependencias

- 1) Dirigirse a la carpeta donde se instalo la herramienta “Data integration - kettle” (<kettle>)
- 2) Ir al siguiente directorio: <kettle>/datamart_sic_etl/
- 3) Realizar un respaldo de la carpeta “etl” y eliminarla
- 4) Ejecutar los archivos “Carga inicial.bat”, “Carga Incremental.bat” y “Actualizacion.bat”

Tabla 78. Caso de prueba – Validación de dependencias

Descripción general:	Comprobar que los procesos ETL no pueden continuar cuando faltan uno o mas archivos necesarios.
Datos de Entrada:	----
Resultados Esperados:	Se produce un error con ID:7, y se debe revisar el log
Condiciones para esta prueba:	----
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	Se obtuvo los resultados esperados

2.4.2. PRUEBAS SOBRE LOS REPORTES INICIALES

El objetivo de estas pruebas es verificar que la información de los reportes es la correcta, se pretende identificar incoherencias en la agregación de las medidas, errores de sintaxis en las consultas MDX, etc.

Actualmente las medidas están sujetas a las siguientes agregaciones:

- Medidas: Personas, Vehículos, Ayudas, Monto, Subvención, a la Agregación sumatoria.
- Medidas: Porcentaje discapacidad, Recepción(días), Atención(días), Resolución(días), Total(días) a la Agregación promedio.

También se realizó pruebas para verificar que la medida Porcentaje este bien calculada.

Se elaboró un caso de prueba para cada tipo de agregación utilizando al menos una medida de cada cubo.

Tabla 79. Procedimiento de prueba para verificación de medidas

- 1) Seleccionar un reporte en el cual se presente la medida calculada según la agregación a probar
- 2) Ejecutar la consulta MDX correspondiente
- 3) Anotar los valores obtenidos
- 4) Construir en SQL la misma consulta
- 5) Ejecutar la consulta SQL
- 6) Anotar los valores obtenidos
- 7) Comparar los valores obtenidos en ambos casos

2.4.2.1. Caso de prueba para la medida: Persona

Tabla 80. Caso de prueba para la medida: Persona, agregación: sumatoria	
Descripción general:	Para esta prueba se seleccionó el reporte: Distribución de PCD por deficiencia principal
Datos de Entrada:	<p><u>Consulta MDX:</u> SELECT {[Measures].[fac_per_num]} ON columns, {[dim_ped_deficiencia_principal].[Todos].Children} ON rows FROM [Personas discapacitadas] WHERE ([dim_per_estado].[Todos].[A])</p> <p><u>Consulta SQL:</u> SELECT dd.dim_ped_deficiencia_principal, sum(f.fac_per_num) as Personas FROM fact_personas_discapacitadas f, dim_persona_discapacidad dd, dim_persona dp WHERE f.dim_ped_key=dd.dim_ped_key and f.dim_per_key=dp.dim_per_key and dp.dim_per_estado='A' GROUP BY dd.dim_ped_deficiencia_principal</p>
Resultados Esperados:	No hay diferencia al comparar los valores obtenidos con la consulta MDX vs. los obtenidos con la consulta SQL
Condiciones para esta prueba:	<ul style="list-style-type: none"> - Existen datos en la bdd datamart_sic - Están cargados los cubos en Mondrian
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	No hay diferencia

2.4.2.2. Caso de prueba para la medida: Vehículos

Tabla 81. Caso de prueba para la medida: Vehículos, agregación: sumatoria	
Descripción general:	Para esta prueba se seleccionó el reporte: Vehículos autorizados al año
Datos de Entrada:	<p><u>Consulta MDX:</u> SELECT NON EMPTY CrossJoin({[Measures].[fac_veh_num]}, {[dim_tim_autoriza].[Todos].[2007], [dim_tim_autoriza].[Todos].[2008]}) ON columns, NON EMPTY {[dim_loc].[Todos].Children} ON rows FROM [Autorizacion de vehiculos]</p> <p><u>Consulta SQL:</u> SELECT dl.dim_loc_region, dt.dim_tim_anio, sum(f.fac_veh_num) as Vehiculos FROM fact_autorizacion_vehiculo f, dim_localidad dl, dim_tiempo dt WHERE f.dim_loc_key=dl.dim_loc_key and</p>

	f.dim_tim_autoriza_key=dt.dim_tim_key and (dt.dim_tim_anio=2007 or dt.dim_tim_anio=2008) GROUP BY dl.dim_loc_region, dt.dim_tim_anio ORDER BY dim_loc_region
Resultados Esperados:	No hay diferencia al comparar los valores obtenidos con la consulta MDX vs. los obtenidos con la consulta SQL
Condiciones para esta prueba:	- Existen datos en la bdd datamart_sic - Están cargados los cubos en Mondrian
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	No hay diferencia

2.4.2.3. Caso de prueba para la medida: Ayudas

Tabla 82. Caso de prueba para la medida: Ayudas, agregación: sumatoria	
Descripción general:	Para esta prueba se seleccionó el reporte: Distribución local de ayudas técnicas
Datos de Entrada:	<p><u>Consulta MDX:</u> SELECT {[Measures].[fac_ayu_num], [dim_ayu].[Todos].[ayuda tecnica]} ON columns, Hierarchize(Union({[dim_loc].[Todos]}, [dim_loc].[Todos].Children)) ON rows FROM [Ayudas] WHERE ([dim_eay].[Todos].[ENTREGADO])</p> <p><u>Consulta SQL:</u> SELECT dl.dim_loc_region, da.dim_ayu_clase, sum(f.fac_ayu_num) as Ayudas FROM fact_ayudas f, dim_localidad dl, dim_ayudas da, dim_estado_ayuda de WHERE f.dim_loc_key=dl.dim_loc_key and f.dim_ayu_key=da.dim_ayu_key and f.dim_eay_key=de.dim_eay_key and da.dim_ayu_clase='ayuda tecnica' and de.dim_eay_estado='ENTREGADO' GROUP BY dl.dim_loc_region, da.dim_ayu_clase ORDER BY dim_loc_region</p>
Resultados Esperados:	No hay diferencia al comparar los valores obtenidos con la consulta MDX vs. los obtenidos con la consulta SQL
Condiciones para esta prueba:	- Existen datos en la bdd datamart_sic - Están cargados los cubos en Mondrian
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables	Luis Mosquera

de Ejecución de la Prueba:	
Resultados Obtenidos:	No hay diferencia

2.4.2.4. Caso de prueba para las medidas: Monto, Subvención

Tabla 83. Caso de prueba para las medidas: Monto, Subvención, agregación: sumatoria	
Descripción general:	Para esta prueba se seleccionó el reporte: Monto total en ayudas entregadas al año Se añadió la medida Subvención a las columnas
Datos de Entrada:	<u>Consulta MDX:</u> SELECT CrossJoin({[Measures].[fac_ayu_monto], [Measures].[fac_ayu_subvencion]}, {[dim_tim_solicita].[Todos].[2008]}) ON columns, {[dim_ayu].[Todos].children} ON rows FROM [Ayudas] WHERE ([dim_eay].[Todos].[ENTREGADO]) <u>Consulta SQL:</u> SELECT da.dim_ayu_clase, dt.dim_tim_anio, sum(f.fac_ayu_monto) as Monto, sum(f.fac_ayu_subvencion) FROM fact_ayudas f, dim_ayudas da, dim_tiempo dt, dim_estado_ayuda de WHERE f.dim_ayu_key=da.dim_ayu_key and f.dim_tim_entrega_key=dt.dim_tim_key and f.dim_eay_key=de.dim_eay_key and dt.dim_tim_anio=2008 and de.dim_eay_estado='ENTREGADO' GROUP BY da.dim_ayu_clase, dt.dim_tim_anio
Resultados Esperados:	No hay diferencia al comparar los valores obtenidos con la consulta MDX vs. los obtenidos con la consulta SQL
Condiciones para esta prueba:	<ul style="list-style-type: none"> - Existen datos en la bdd datamart_sic - Están cargados los cubos en Mondrian
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	No hay diferencia

2.4.2.5. Caso de prueba para las medidas de tiempo

Tabla 84. Caso de prueba para las medidas de tiempo, agregación: promedio	
Descripción general:	<p>Las medidas son: Recepción(días), Atención(días), Resolución(días), Total(días)</p> <p>Para esta prueba se seleccionó el reporte: Tiempo promedio para autorizar un vehículo</p> <p>Se añadió todas las medidas antes mencionadas a la columna</p>
Datos de Entrada:	<p>Consulta MDX:</p> <pre>SELECT {[Measures].[fac_veh_dias_total], [Measures].[fac_veh_dias_recepcion], [Measures].[fac_veh_dias_atencion], [Measures].[fac_veh_dias_autoriza]} ON columns, {[dim_veh].[Todos].children} ON rows FROM [Autorizacion de vehiculos]</pre> <p>Consulta SQL:</p> <pre>SELECT dv.dim_veh_tipo_vehiculo, avg(f.fac_veh_dias_total) as "Total(dias)", avg(f.fac_veh_dias_recepcion) as "Recepcion(dias)", avg(f.fac_veh_dias_atencion) as "Atencion(dias)", avg(f.fac_veh_dias_autoriza) as "Resolucion(dias)" FROM fact_autorizacion_vehiculo f, dim_vehiculo dv WHERE f.dim_veh_key=dv.dim_veh_key GROUP BY dv.dim_veh_tipo_vehiculo</pre>
Resultados Esperados:	No hay diferencia al comparar los valores obtenidos con la consulta MDX vs. los obtenidos con la consulta SQL
Condiciones para esta prueba:	<ul style="list-style-type: none"> - Existen datos en la bdd datamart_sic - Están cargados los cubos en Mondrian
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	No hay diferencia

2.4.2.6. Caso de prueba para la medida Porcentaje

Tabla 85. Caso de prueba para la medida Porcentaje	
Descripción general:	Para esta prueba se seleccionó el reporte: Porcentaje de PCD por genero Se añadió todas las medidas antes mencionadas a la columna
Datos de Entrada:	<p>Consulta MDX:</p> <pre>WITH MEMBER [Measures].[Porcentaje] AS '([Measures].[fac_per_num] / ([Measures].[fac_per_num], [dim_per_genero].CurrentMember.Parent))',format_string = 'Percent' SELECT {[Measures].[Porcentaje]} ON columns, {[dim_per_genero].[Todos].children} ON rows FROM [Personas discapacitadas]</pre> <p>Consulta SQL:</p> <pre>SELECT dp.dim_per_genero, cast(cast(count(f.dim_per_key)as float8)/(select count(dim_per_key) FROM fact_personas_discapacitadas)*100 as decimal(4,2)) as "Porcentaje" from fact_personas_discapacitadas f, dim_persona dp WHERE f.dim_per_key=dp.dim_per_key GROUP BY dp.dim_per_genero</pre>
Resultados Esperados:	No hay diferencia al comparar los valores obtenidos con la consulta MDX vs. los obtenidos con la consulta SQL
Condiciones para esta prueba:	<ul style="list-style-type: none"> - Existen datos en la bdd datamart_sic - Están cargados los cubos en Mondrian
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	No hay diferencia

2.4.3. PRUEBAS SOBRE LA APLICACIÓN MAPSIC

El objetivo de estas pruebas es encontrar errores en la generación de los reportes, y/o errores en la distribución de los datos en el mapa.

2.4.3.1. Casos de prueba para los reportes

Tabla 86. Procedimiento de prueba para los Reportes de la aplicación Mapsic	
<ol style="list-style-type: none"> 1) Ingresar a la aplicación mediante un navegador web 2) Seleccionar la pestaña de Reportes 3) Seleccionar valores para los parámetros de consulta 4) Dar clic en el botón “Consultar” 5) Desplegar cada uno de los reportes 6) Verificar que la estructura del reporte sea la adecuada, y que los datos sean iguales a los de su similar en Openi 	

Tabla 87. Caso de prueba 1 - Reportes de la aplicación Mapsic	
Descripción general:	Corresponde a un error encontrado en la generación de reportes
Datos de Entrada:	Consultar = {Personas discapacitadas} En = {Ecuador} Por = {Cantón}
Resultados Esperados:	La estructura de los reportes es la adecuada, sus valores corresponden con los de su similar en Openi
Condiciones para esta prueba:	- Existen datos en la bdd datamart_sic - Están cargados los cubos en Mondrian
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	La estructura de los reportes presenta errores, no cargo el nombre de los cantones en las cabeceras de las filas, cargo el de las provincias descuadrando completamente los reportes Los datos obtenidos sin embargo si corresponden a los de sus similares en Openi.

Tabla 88. Caso de prueba 2 - Reportes de la aplicación Mapsic	
Descripción general:	Escenario bajo el cual el caso de prueba tuvo éxito
Datos de Entrada:	Consultar = {Personas discapacitadas} En = {Ecuador, (Azuay – Zamora Chinchipe)} Por = {Provincia, Cantón} * “Cantón” solo para el conjunto (Azuay – Zamora Chinchipe)
Resultados Esperados:	La estructura de los reportes es la adecuada, sus valores corresponden con los de su similar en Openi.
Condiciones para esta prueba:	- Existen datos en la bdd datamart_sic - Están cargados los cubos en Mondrian
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	La estructura de los reportes es la adecuada y los datos obtenidos si corresponden a los de sus similares en Openi.

2.4.3.2. Casos de prueba para la distribución de datos en el mapa

Tabla 89. Procedimiento de prueba para la distribución de datos en el mapa
<ol style="list-style-type: none"> 1) Ingresar a la aplicación mediante un navegador web 2) Seleccionar la pestaña de Mapa 3) Seleccionar valores para los parámetros de consulta 4) Dar clic en el botón “Consultar” 5) Seleccionar el tipo de distribución para el mapa 6) Verificar que los datos de la distribución por localidad estén bien ubicados en el mapa.

Tabla 90. Caso de prueba para la distribución de datos en el mapa	
Descripción general:	Se utilizó la distribución por grupo de colores, para este caso, cada valor de la distribución por localidad se ubica en uno de los rangos del grupo de colores del mapa, la localidad del mapa debe pintarse conforme a ese color.
Datos de Entrada:	Consultar = {Personas discapacitadas} En = {Ecuador, (Azuay – Zamora Chinchipe)} Por = {Provincia, Cantón} Distribución = {Distribuir por grupo de colores}
Resultados Esperados:	Que cada localidad del mapa este pintada de acuerdo al color que le corresponde en función de los datos de la distribución por localidad
Condiciones para esta prueba:	- Existen datos en la bdd datamart_sic - Están cargados los cubos en Mondrian
Fechas de Ejecución de la Prueba:	Martes, 23 de noviembre de 2010
Responsables de Ejecución de la Prueba:	Luis Mosquera
Resultados Obtenidos:	Se obtuvo los resultados esperados

3. CAPITULO 3 : CONCLUSIONES Y RECOMENDACIONES

3.1. CONCLUSIONES

- El ciclo de vida propuesto por Kimball facilita las tareas de desarrollo del Data warehouse, en este caso las del Data mart, define los productos que se deben obtener tras finalizar cada etapa, además de presentar plantillas, formatos y diagramas para la documentación.
- El esfuerzo aplicado en estructurar consultas SQL para el análisis de datos es alto, especialmente en bases de datos transaccionales esta tarea es compleja; ahora mediante el Data mart el cual presenta una base de datos apropiada para realizar análisis y consultar grandes cantidades de datos, el estructurar consultas en SQL involucra menos esfuerzo, sin embargo es más eficaz utilizar MDX mediante los cubos OLAP.
- Una de las actividades mas importantes dentro del Data mart es el proceso ETL, especialmente el aseguramiento de la calidad de los datos, por lo tanto es importante contar con sistemas fuentes confiables que validen y controlen el registro de datos, estos mecanismos facilitan en gran medida el tratamiento de los datos.
- El Data mart mejora la experiencia de los usuarios poniendo a su disposición la información que necesitan cuando la necesitan, por lo tanto es importante seleccionar herramientas apropiadas que se ajusten a las necesidades de la organización.

3.2. RECOMENDACIONES

- Recomiendo realizar una retro alimentación en base a la información y análisis que actualmente esta en capacidad de brindar el Data mart, con la finalidad de facilitar el planteamiento de nuevos requerimientos los cuales pueden ser tomados en cuenta para un nuevo ciclo de vida del Data mart.
- En vista de la importancia de los metadatos a lo largo del ciclo de vida del Data mart, muchas de las veces tener documentada esta información en archivos digitales o físicos no es suficiente, recomiendo estructurar catálogos en línea para los metadatos, de tal manera que puedan estar sincronizados con las aplicaciones y sistemas de la organización.
- Recomiendo detallar información geo-referenciada para las personas discapacitadas registradas en el CONADIS, así como de entidades asociadas, con la finalidad de aprovechar al máximo los beneficios de las herramientas GIS.
- Para llevar acabo el proceso de carga incremental de datos, recomiendo calendarizar la ejecución del programa respectivo en función de la frecuencia con que registra nueva información el sistema web SIC.

BIBLIOGRAFÍA

Libros

- 1) KIMBALL Ralph, REEVES Laura, ROSS Margy, THORNTHWAITE Warren, The Data Warehouse Lifecycle Toolkit, 1era Edición, Wiley, 2001
- 2) KIMBALL Ralph, ROSS Margy, The Data Warehouse Toolkit, 2da Edición, Wiley, 2002
- 3) KIMBALL Ralph, CASERTA Joe, The Data Warehouse ETL Toolkit, Wiley, 2004
- 4) ROLDÁN María Carina, Pentaho 3.2 Data Integration Beginner's Guide, Packt Publishing Ltd., 2010
- 5) RAINARDI Vincent, Building a Data Warehouse, Apress, 2008
- 6) WANG John, Encyclopedia of Data Warehousing and Mining, Idea Group, 2006

Páginas web

- 1) Wikipedia, <http://es.wikipedia.org/>
- 2) Kimball group, Design Tip #81 Fact Table Surrogate Key, <http://www.ralphkimball.com/html/designtipsPDF/DesignTips2006/KU81FactTableSurrogateKeys.pdf>
- 3) Pentaho, How to Design a Mondrian Schema, <http://mondrian.pentaho.com/documentation/schema.php>
- 4) Arquitectura técnica del Data warehouse, http://biblioteca.usac.edu.gt/tesis/08/08_7695.pdf
- 5) El Rincon del BI, Construcción procesos ETL, <http://churriwifi.wordpress.com/2010/05/10/16-3-construccion-procesos-etl-utilizando-kettle-pentaho-data-integration/>
- 6) Pentaho Community, Pentaho Data Integration Steps, <http://wiki.pentaho.com/display/EAI/Pentaho+Data+Integration+Steps>
- 7) Openi.org, OpenI Wiki, <http://wiki.openi.org/installation/quick-start>
- 8) MapServer, Mapfile, <http://mapserver.org/mapfile/index.html>
- 9) CONADIS, Estadísticas, <http://conadis.gov.ec/estadisticas.htm>

GLOSARIO

Ajax.- Acrónimo de Asynchronous JavaScript And XML, es una técnica de desarrollo web para crear aplicaciones interactivas, mantiene comunicación asíncrona con el servidor, de esta forma es posible realizar cambios sobre las páginas sin necesidad de recargarlas.

Bottom-up.- Las partes individuales se diseñan con detalle y luego se enlazan para formar componentes más grandes, que a su vez se enlazan hasta que se forma el sistema completo.

B-tree.- En informática, un árbol-B es un tipo de estructura de datos de árboles. Representa una colección de datos ordenados de manera que se permite una inserción y borrado eficientes de elementos. Es un índice, multinivel, dinámico, con un límite máximo y mínimo en el número de claves por nodo.

Caso de uso.- Es una técnica para la captura de requisitos potenciales de un nuevo sistema o una actualización de software. Cada caso de uso proporciona uno o más escenarios que indican cómo debería interactuar el sistema con el usuario o con otro sistema para conseguir un objetivo específico.

Categoría.- Cada una de las Jerarquías de atributos presentes en una dimensión.

Clave subrogada.- Una Clave Subrogada es un campo numérico de una tabla cuyo único requisito es almacenar un valor numérico único para cada fila de la tabla.

CONADIS.- Consejo Nacional de Discapacidades

CSS.- Cascading Style Sheets (hojas de estilo en cascada). Es un lenguaje usado para definir la presentación de un documento estructurado escrito en HTML o XML.

Data mart.- Un Data mart es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones.

Data warehouse.- Almacén de datos, es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.

DBMS.- Database management system (sistemas de gestión de bases de datos) son un tipo de software muy específico, dedicado a servir de interfaz entre la base de datos, el usuario y las aplicaciones que la utilizan.

DDL.- Data Definition Language (Lenguaje de definición de datos). Es un lenguaje proporcionado por el sistema de gestión de base de datos, que permite a los usuarios de la misma llevar a cabo las tareas de definición de las estructuras que almacenarán los datos así como de los procedimientos o funciones que permitan consultarlos.

Denormalización.- Es el proceso de procurar optimizar el desempeño de una base de datos por medio de agregar datos redundantes.

DOM.- Document Object Model (modelo de objetos del documento). Es esencialmente un API que proporciona un conjunto estándar de objetos para representar documentos HTML y XML, los programas pueden acceder y modificar el contenido, estructura y estilo de los documentos HTML y XML.

Drilldown.- Es un método de exploración para datos multidimensionales, consiste en moverse desde un nivel a otro nivel mas detallado dentro de una jerarquía.

ETL.- Extract, Transform and Load (Extraer, transformar y cargar). Es el proceso que permite mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse.

GIS.- Geographic Information System (Sistema de Información Geográfica). Es una integración organizada de hardware, software y datos geográficos diseñada para capturar, almacenar, manipular, analizar y desplegar en todas sus formas la información geográficamente referenciada.

Javascript.- Se utiliza principalmente en su forma del lado del cliente (client-side), implementado como parte de un navegador web permitiendo mejoras en la interfaz de usuario y páginas web dinámicas.

KPI.- En la terminología empresarial, un indicador clave de rendimiento (KPI) es una medida cuantificable para valorar los éxitos empresariales.

En Analysis Services, un KPI es un conjunto de cálculos asociados a un grupo de medida de un cubo, que se usa para evaluar el éxito empresarial. Normalmente, estos cálculos son una combinación de expresiones MDX o miembros calculados.

MDX.- MultiDimensional eXpressions (expresiones multidimensionales). Es un lenguaje de consulta para bases de datos multidimensionales sobre cubos OLAP.

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

PHP.- Es un lenguaje de programación interpretado, diseñado originalmente para la creación de páginas web dinámicas. Es usado principalmente en interpretación del lado del servidor.

SQL.- structured query language (lenguaje de consulta estructurado). Es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones en éstas.

Tablas Pivot.- Las tablas dinámicas, también llamadas pivot tables, son una herramienta para análisis de bases de datos. Se encargan de resumir y ordenar la información contenida en estas.

XHTML.- eXtensible Hypertext Markup Language (lenguaje extensible de marcado de hipertexto). Es el lenguaje de marcado pensado para sustituir a HTML como estándar para las páginas web.

XMLA.- XML for Analysis (XML para Análisis). Es un estándar para el acceso a datos en sistemas de análisis, como son OLAP y minería de datos.

ANEXOS

ANEXO A – Información

A1: Preguntas para las entrevistas

Entrevista Director

¿Cuáles son los objetivos de la organización?

¿Qué podría evitar conseguir los objetivos anteriores?

¿Cuáles considera que son las métricas que permiten medir el éxito de la organización?

¿Cuáles son las áreas o departamentos de la organización?

¿Cuáles son los procesos dentro del departamento de servicios sociales?

¿Cuáles son las entidades que participan en los procesos anteriores?

¿Cuáles son los reportes que con mas frecuencia solicita?

¿Cómo obtiene actualmente esos reportes?

¿Qué hace con esa información?

¿Ha solicitado alguna vez información no prevista en los reportes anteriores?

¿Si lo anterior es verdad, cuanto tiempo tomo obtener la información?

¿Qué otros reportes le gustaría obtener y de que tipo?

¿Cuánta información histórica se utiliza para los reportes?

¿Considera que existe algún cuello de botella que impide obtener eficazmente la información?

¿Cree que si se mejora el acceso a la información, esto repercutirá positivamente a la organización?

Entrevista Jefe de sistemas

- ¿Con que fuentes de información cuentan (Sistemas Operacionales)?
- ¿Cuál es el proceso utilizado para difundir la información a los usuarios?
- ¿Qué herramientas utilizan actualmente para acceder y analizar la información?
- ¿Qué tipo de análisis se generan rutinariamente (reportes)?
- ¿Cómo manejan las solicitudes de información no prevista?
- ¿Existe algún factor que puede causar pérdida de calidad en la información?
- ¿Cuál es el principal cuello de botella en el proceso actual de acceso a la información?
- ¿Que le parece la posibilidad de implantar un Data Mart?

A2: Información organizacional del CONADIS¹

Visión

El Consejo Nacional de Discapacidades, será un organismo dinamizador y articulador de todos los sectores de la comunidad para desarrollar acciones de prevención, atención e integración, con el propósito de prevenir las discapacidades y elevar la calidad de vida de las personas con discapacidad.

Misión

Es un organismo público y autónomo que genera políticas e información, planifica y coordina acciones con los sectores público y privado, en el ámbito de las discapacidades y canaliza recursos nacionales e internacionales en este campo.

Funciones

- Formular políticas nacionales relacionadas con las discapacidades

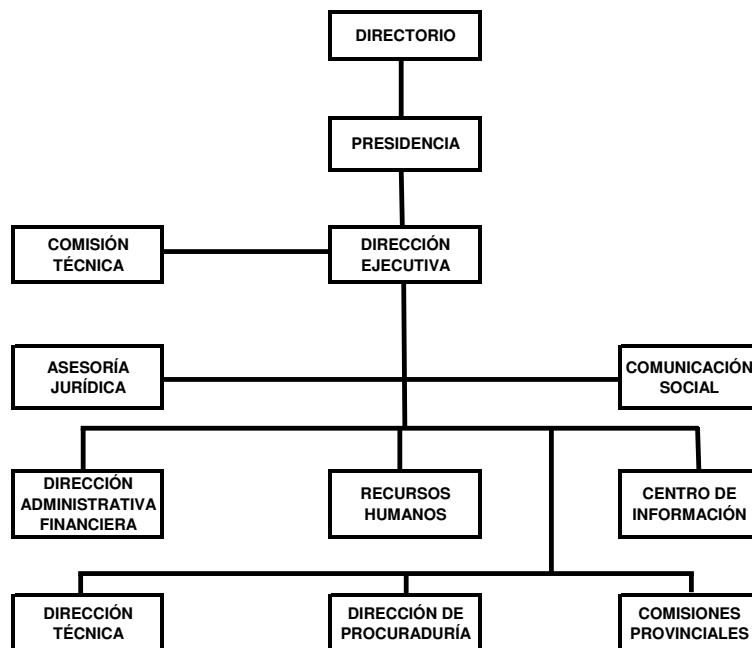
¹ <http://conadis.gov.ec/>

- Establecer el Plan Nacional de Discapacidades
- Defender jurídicamente los derechos de las personas con discapacidad
- Coordinar, efectuar el seguimiento y evaluar las acciones que realizan en discapacidades las entidades del sector público y privado .
- Canalizar recursos nacionales e internacionales.
- Realizar o impulsar investigaciones en el área de las discapacidades
- Vigilar el cumplimiento de la Ley y la aplicación de sanciones o quienes la incumplan.

Órganos del CONADIS

1. DIRECTORIO
2. DIRECCIÓN EJECUTIVA
3. COMISIÓN TÉCNICA ; Y
4. COMISIONES PROVINCIALES

Estructura Orgánica



A3: Sistema web del CONADIS (SIC)¹

Este sistema web se basa en las especificaciones y requerimientos del sistema anterior que poseía el CONADIS, con algunos cambios que se han ido implementando por las necesidades que se presentan en la organización.

Procesos que efectúa el CONADIS a través del uso del Sistema Informático Web del CONADIS:

- Registro Nacional de Discapacidades
- Ayudas Técnicas
- Medicamentos e Insumos
- Solicitud y Autorización de Vehículos
- Solicitud y Autorización de Bienes
- Defensa de Derechos
- Instituciones

Mantenimiento y Administración

¹ Manual de usuario del Sistema web SIC

ANEXO B – Scripts del Proceso ETL

B1: Script sql – dim_per(DS)

Paso del ETL de la dimensión persona.

Extrae información sobre la persona, para el ETL de la dimensión persona.

```
SELECT
p.per_id as dim_per_id,
ec.est_id as dim_per_estado_civil_id,
p.per_apellido1 as dim_per_apellido1,
p.per_apellido2 as dim_per_apellido2,
p.per_nombre1 as dim_per_nombre1,
p.per_nombre2 as dim_per_nombre2,
p.per_sexo as dim_per_genero,
ec.est_descripcion as dim_per_estado_civil,
cast((date_part('year',current_date) - date_part('year',per_fecha_nac)) as int)
as dim_per_edad_id,
p.per_estado as dim_per_estado
FROM
persona p,
estado_civil ec
WHERE
p.tpe_id=1 and
p.est_id=ec.est_id
ORDER BY
dim_per_id
```

B2: Script sql – dim_ped(DS)

Paso del ETL de la dimensión persona – discapacidad.

Extrae información sobre discapacidades, para el ETL de la dimensión persona – discapacidad.

```
SELECT
gd.gra_id as dim_ped_grado_discapacidad_id,
dp.dpr_id as dim_ped_deficiencia_principal_id,
cd.cde_id as dim_ped_causa_deficiencia_id,
gd.gra_descripcion as dim_ped_grado_discapacidad,
dp.dpr_descripcion as dim_ped_deficiencia_principal,
cd.cde_descripcion as dim_ped_causa_deficiencia
FROM
grado_discapacidad gd,
deficiencia_principal dp,
causa_deficiencia cd
ORDER BY
dim_ped_grado_discapacidad_id,
dim_ped_deficiencia_principal_id,
dim_ped_causa_deficiencia_id
```

B3: Script sql – dim_peh(DS)

Paso del ETL de la dimensión persona – hogar.

Extrae información sobre la situación familiar de la persona de la bdd fuente, para el ETL de la dimensión persona – hogar.

```
SELECT
cq.con_id as dim_peh_con_quien_vive_id,
tv.viv_id as dim_peh_tipo_vivienda_id,
cq.con_descripcion as dim_peh_con_quien_vive,
tv.viv_descripcion as dim_peh_tipo_vivienda
FROM
con_quien_vive cq,
tipo_vivienda tv
ORDER BY
dim_peh_con_quien_vive_id,
dim_peh_tipo_vivienda_id
```

B4: Script sql – dim_pet_si(DS)

Paso del ETL de la dimensión persona – trabajo.

Extrae información sobre las personas que si tienen trabajo, asigna el valor de “SI” para la columna dim_pet_trabaja_id, para el ETL de la dimensión persona – trabajo.

```
SELECT
ni.niv_id as dim_pet_nivel_instruccion_id,
'SI' as dim_pet_trabaja_id,
tt.tra_id as dim_pet_trabaja_descripcion_id,
ni.niv_descripcion as dim_pet_nivel_instruccion,
tt.tra_descripcion as dim_pet_trabaja_descripcion
FROM
nivel_instruccion ni,
tipo_trabajo tt
order by
dim_pet_nivel_instruccion_id,
dim_pet_trabaja_id,
dim_pet_trabaja_descripcion_id
```

B5: Script sql – dim_pet_no(DS)

Paso del ETL de la dimensión persona – trabajo.

Extrae información sobre las personas que no tienen trabajo, asigna el valor de “NO” para la columna dim_pet_trabaja.

```
SELECT
ni.niv_id as dim_pet_nivel_instruccion_id,
'NO' as dim_pet_trabaja,
ni.niv_descripcion as dim_pet_nivel_instruccion,
ct.ctr_descripcion as dim_pet_trabaja_descripcion
FROM
nivel_instruccion ni,
causa_no_trabaja ct
order by
dim_pet_nivel_instruccion_id,
dim_pet_trabaja
```

B6: Script sql – dim_loc_nulls(DS)

Paso del ETL de la dimensión localidad.

Extrae información sobre provincias de la bdd fuente, asigna “0” para el código de cantones y parroquias y “desconocida” a la descripción, extrae información sobre provincias y cantones de la bdd fuente, asigna “0” para el código de parroquias y “desconocida” a la descripción.

```
SELECT
loc_id as dim_loc_provincia_id,
1 as dim_loc_canton_id,
1 as dim_loc_parroquia_id,
loc_codigo as dim_loc_provincia_cod,
'0' as dim_loc_canton_cod,
'0' as dim_loc_parroquia_cod,
loc_descripcion as dim_loc_provincia,
'Desconocida' as dim_loc_canton,
'Desconocida' as dim_loc_parroquia
FROM
localidad
WHERE
tlo_id=1
UNION
SELECT
lp.loc_id as dim_loc_provincia_id,
lc.loc_id as dim_loc_canton_id,
1 as dim_loc_parroquia_id,
lp.loc_codigo as dim_loc_provincia_cod,
lc.loc_codigo as dim_loc_canton_cod,
'0' as dim_loc_parroquia_cod,
lp.loc_descripcion as dim_loc_provincia,
lc.loc_descripcion as dim_loc_canton,
'Desconocida' as dim_loc_parroquia
FROM
```

```

localidad lp,
localidad lc
WHERE
lp.tlo_id=1 and
lc.tlo_id=2 and
lp.loc_codigo=lc.loc_codigo_sup
ORDER BY
dim_loc_provincia_id,
dim_loc_canton_id,
dim_loc_parroquia_id

```

B7: Script sql – dim_loc(DS)

Paso del ETL de la dimensión localidad.

Extrae información sobre provincias y cantones y parroquia.

```

SELECT
lp.loc_id as dim_loc_provincia_id,
lc.loc_id as dim_loc_canton_id,
lr.loc_id as dim_loc_parroquia_id,
lp.loc_codigo as dim_loc_provincia_cod,
lc.loc_codigo as dim_loc_canton_cod,
lr.loc_codigo as dim_loc_parroquia_cod,
lp.loc_descripcion as dim_loc_provincia,
lc.loc_descripcion as dim_loc_canton,
lr.loc_descripcion as dim_loc_parroquia
FROM
localidad lp,
localidad lc,
localidad lr
WHERE
lp.tlo_id=1 and
lc.tlo_id=2 and
lr.tlo_id=3 and
lp.loc_codigo=lc.loc_codigo_sup and
lc.loc_codigo=lr.loc_codigo_sup
ORDER BY
dim_loc_provincia_id,
dim_loc_canton_id,
dim_loc_parroquia_id

```

B8: Script sql – dim_veh(DS)

Paso del ETL de la dimensión vehículo.

Extrae información sobre vehículos.

```

SELECT
tvh_id as dim_veh_tipo_vehiculo_id,
tvh_descripcion as dim_veh_tipo_vehiculo
FROM
tipo_vehiculo
ORDER BY
tvh_id

```

B9: Script sql – dim_uni(DS)

Paso del ETL de la dimensión unidad autorizada.

Extrae información sobre unidades autorizadoras.

```
SELECT
au.uni_id as dim_uni_id,
au.eau_id as dim_uni_entidad_autorizada_id,
au.loc_id as dim_uni_provincia_id,
au.uni_nombre as dim_uni_nombre,
au.uni_director as dim_uni_director,
ea.eau_nombre as dim_uni_entidad_autorizada,
l.loc_descripcion as dim_uni_provincia
FROM
unidad_autorizadora au,
entidad_autorizadora ea,
localidad l
WHERE
au.eau_id=ea.eau_id and
au.loc_id=l.loc_id
ORDER BY
dim_uni_id
```

B10: Script sql – dim_ayuAT(DS)

Paso del ETL de la dimensión ayudas.

Extrae información sobre ayudas técnicas, asigna el valor “ayuda técnica” a la columna dim_ayu_clase.

```
SELECT
'ayuda tecnica' as dim_ayu_clase,
tay_id as dim_ayu_id,
tay_descripcion as dim_ayu_descripcion
FROM
tipo_ayuda_tecnica
ORDER BY
dim_ayu_clase,
dim_ayu_id
```

B11: Script sql – dim_ayuMED(DS)

Paso del ETL de la dimensión ayudas.

Extrae información sobre medicamentos e insumos, asigna el valor “medicamento” a la columna dim_ayu_clase.

```
SELECT
'medicamento' as dim_ayu_clase,
med_id as dim_ayu_id,
med_descripcion as dim_ayu_descripcion
FROM
medicamento_o_insumo
ORDER BY
dim_ayu_clase,
dim_ayu_id
```

B12: Script sql – dim_eay(DS)

Paso del ETL de la dimensión estado ayuda.

Extrae los estados para ayudas técnicas y medicamentos.

```
SELECT
eay_id as dim_eay_id,
eay_descripcion as dim_eay_estado
FROM
estado_ayuda
ORDER BY
eay_id
```

B13: Script sql – dim_ens(DS)

Paso del ETL de la dimensión entidad subvenciona

Extrae información sobre las entidades que subvencionan ayudas.

```
SELECT
sub_id as dim_ens_id,
sub_nombre as dim_ens_nombre,
sub_siglas as dim_ens_siglas
FROM
entidad_subvenciona
ORDER BY
dim_ens_id
```

B14: Script sql – dim_enp(DS)

Paso del ETL de la dimensión entidad provee.

Extrae información sobre las entidades que proveen ayudas.

```
SELECT
pro_id as dim_enp_id,
pro_nombre as dim_enp_nombre,
pro_siglas as dim_enp_siglas
FROM
entidad_provee
ORDER BY
dim_enp_id
```

B15: Script sql – dim_pry(DS)

Paso del ETL de la dimensión dim_proyecto.

Extrae información sobre proyectos de inversión.

```
SELECT
p.pro_id as dim_pry_id,
p.ent_id as dim_pry_entidad_id,
p.loc_id as dim_pry_localidad_id,
p.tpr_id as dim_pry_tipo_proyecto_id,
p.pro_titulo as dim_pry_titulo,
e.ent_nombre as dim_pry_entidad,
p.pro_responsable as dim_pry_responsable,
l.loc_descripcion as dim_pry_localidad,
tp.tpr_descripcion as dim_pry_tipo_proyecto
FROM
proyecto p,
entidad e,
localidad l,
tipo_proyecto tp
WHERE
p.ent_id=e.ent_id and
p.loc_id=l.loc_id and
p.tpr_id=tp.tpr_id
ORDER BY
dim_pry_id
```


B16: Script sql – fact_per(DS)

Paso del ETL de la tabla de hechos personas discapacitadas.

Extrae las claves operacionales asociadas a las dimensiones de la tabla de hechos fact_personas_discapacitadas y las columnas para las medidas.

```
SELECT
p.per_id as dim_per_id,
p.gra_id as dim_ped_grado_discapacidad_id,
pd.dpr_id as dim_ped_deficiencia_principal_id,
cd.cde_id as dim_ped_causa_deficiencia_id,
p.con_id as dim_peh_con_quien_vive_id,
p.viv_id as dim_peh_tipo_vivienda_id,
p.niv_id as dim_pet_nivel_instruccion_id,
p.per_trabaja as dim_pet_trabaja_id,
CASE per_trabaja
WHEN 'SI' THEN tt.tra_id
ELSE ct.ctr_id
END dim_pet_trabaja_descripcion_id,
p.loc_id_provincia as dim_loc_provincia_id,
p.loc_id_canton as dim_loc_canton_id,
p.loc_id_parroquia as dim_loc_parroquia_id,
cast(to_char(p.per_fecha_carnet,'YYYY-MM-DD') as date) as
dim_tim_fecha_carnet_id,
p.per_porcentaje_disc as fac_per_porcentaje_discapacidad,
p.per_edad_disc as fac_per_edad_discapacidad
FROM
persona p,
per_dpr pd,
per_causa_def cd,
tipo_trabajo tt,
causa_no_trabaja ct
WHERE
p.tpe_id=1 and
pd.pdp_peso=1 and
p.per_id=pd.per_id and
p.per_id=cd.per_id and
p.tra_id=tt.tra_id and
p.ctr_id=ct.ctr_id and
p.per_fecha_carnet between cast('${START_DATE}' as date) and cast('${END_DATE}'
as date)
ORDER BY
dim_per_id
```

B17: Script sql – fac_per(OR)

Extrae los datos de la tabla de hechos personas discapacitadas.

```
SELECT
fac_per_porcentaje_discapacidad,
fac_per_edad_discapacidad,
dim_per_key,
dim_ped_key,
dim_peh_key,
dim_pet_key,
dim_loc_key,
dim_tim_carnet_key
```

```

FROM
fact_personas_discapacitadas f,
dim_tiempo d
WHERE
dim_tim_carnet_key=dim_tim_key AND
dim_tim_fecha_completa between cast('${START_DATE}' as date) and
cast('${END_DATE}' as date)
ORDER BY
dim_per_key,
dim_ped_key,
dim_peh_key,
dim_pet_key,
dim_loc_key,
dim_tim_carnet_key

```

B18: Script sql – fac_veh(DS)

Paso del ETL de la tabla de hechos de autorización de vehículos.

Extrae las claves primarias asociadas a las dimensiones de la tabla de hechos de autorización de vehículos, extrae las fechas relacionadas al proceso de autorización de vehículos.

```

SELECT
sv.sol_id as fac_veh_solicitud_id,
av.aut_id as fac_veh_autoriza_id,
sv.per_id as dim_per_id,
p.gra_id as dim_ped_grado_discapacidad_id,
pd.dpr_id as dim_ped_deficiencia_principal_id,
cd.cde_id as dim_ped_causa_deficiencia_id,
p.con_id as dim_peh_con_quien_vive_id,
p.viv_id as dim_peh_tipo_vivienda_id,
p.niv_id as dim_pet_nivel_instruccion_id,
p.per_trabaja as dim_pet_trabaja_id,
CASE per_trabaja
WHEN 'SI' THEN tt.tra_id
ELSE ct.ctr_id
END dim_pet_trabaja_descripcion_id,
p.loc_id_provincia as dim_loc_provincia_id,
p.loc_id_canton as dim_loc_canton_id,
p.loc_id_parroquia as dim_loc_parroquia_id,
av.tvh_id as dim_veh_tipo_vehiculo_id,
av.uni_id as dim_uni_id,
cast(to_char(sv.sol_fechaoriginal,'YYYY-MM-DD') as date) as
dim_tim_fecha_solicitud_id,
cast(to_char(sv.sol_fecharecepcion,'YYYY-MM-DD') as date) as
dim_tim_fecha_recepcion_id,
cast(to_char(av.aut_fecha,'YYYY-MM-DD') as date) as dim_tim_fecha_reunion_id,
cast(to_char(av.aut_fecha_autoriza,'YYYY-MM-DD') as date) as
dim_tim_fecha_autoriza_id,
cast(extract(day from (sv.sol_fecharecepcion - sv.sol_fechaoriginal)) as int) as
fac_veh_dias_recepcion,
cast(extract(day from (av.aut_fecha - sv.sol_fecharecepcion)) as int) as
fac_veh_dias_atencion,
cast(extract(day from (av.aut_fecha_autoriza - av.aut_fecha)) as int) as
fac_veh_dias_autoriza
FROM
solicitud_veh sv,
autorizacion_veh av,
persona p,

```

```

per_dpr pd,
per_causa_def cd,
tipo_trabajo tt,
causa_no_trabaja ct
WHERE
p.tpe_id=1 and
pd.pdp_peso=1 and
sv.sol_id=av.sol_id and
sv.per_id=p.per_id and
p.per_id=pd.per_id and
p.per_id=cd.per_id and
p.tra_id=tt.tra_id and
p.ctr_id=ct.ctr_id and
av.aut_fecha_autoriza between cast('${START_DATE}' as date) and
cast('${END_DATE}' as date)
ORDER BY
fac_veh_solicitud_id,
fac_veh_autoriza_id

```

B19: Script sql – fac_veh(OR)

Extrae los datos de la tabla de hechos de autorización de vehículos

```

SELECT
fac_veh_dias_recepcion,
fac_veh_dias_atencion,
fac_veh_dias_autoriza,
fac_veh_dias_total,
dim_per_key,
dim_ped_key,
dim_peh_key,
dim_pet_key,
dim_loc_key,
dim_veh_key,
dim_uni_key,
dim_tim_solicitud_key,
dim_tim_recepcion_key,
dim_tim_reunion_key,
dim_tim_autoriza_key
FROM
fact_autorizacion_vehiculo,
dim_tiempo
WHERE
dim_tim_autoriza_key=dim_tim_key and
dim_tim_fecha_completa between cast('${START_DATE}' as date) and
cast('${END_DATE}' as date)
ORDER BY
dim_per_key,
dim_ped_key,
dim_peh_key,
dim_pet_key,
dim_loc_key,
dim_veh_key,
dim_uni_key,
dim_tim_solicitud_key,
dim_tim_recepcion_key,
dim_tim_reunion_key,
dim_tim_autoriza_key

```

B20: Script sql – fac_ayuMED(DS)

Paso del ETL de la tabla de hechos ayudas.

Extrae las claves operacionales asociadas a las dimensiones de la tabla de hechos ayudas, extrae los montos y subvenciones asociadas a medicamentos.

```
SELECT
mp.mep_id as fac_ayu_solicitud_id,
mp.per_id as dim_per_id,
p.gra_id as dim_ped_grado_discapacidad_id,
pd.dpr_id as dim_ped_deficiencia_principal_id,
cd.cde_id as dim_ped_causa_deficiencia_id,
p.con_id as dim_peh_con_quien_vive_id,
p.viv_id as dim_peh_tipo_vivienda_id,
p.niv_id as dim_pet_nivel_instruccion_id,
p.per_trabaja as dim_pet_trabaja_id,
CASE per_trabaja
WHEN 'SI' THEN tt.tra_id
ELSE ct.ctr_id
END dim_pet_trabaja_descripcion_id,
p.loc_id_provincia as dim_loc_provincia_id,
p.loc_id_canton as dim_loc_canton_id,
p.loc_id_parroquia as dim_loc_parroquia_id,
2 as dim_ayu_clase_id,
mp.med_id as dim_ayu_id,
mp.sub_id as dim_ens_id,
mp.pro_ent_id as dim_enp_id,
mp.pro_id as dim_pry_id,
mp.eay_id as dim_eay_id,
cast(to_char(mp.mep_fecha_solicita,'YYYY-MM-DD') as date) as
dim_tim_solicita_id,
mp.mep_monto as fac_ayu_monto,
mp.mep_valor as fac_ayu_subvencion
FROM
medicamento_persona mp,
persona p,
per_dpr pd,
per_causa_def cd,
tipo_trabajo tt,
causa_no_trabaja ct
WHERE
p.tpe_id=1 and
pd.pdp_peso=1 and
mp.per_id=p.per_id and
p.per_id=pd.per_id and
p.per_id=cd.per_id and
p.tra_id=tt.tra_id and
p.ctr_id=ct.ctr_id and
mp.mep_fecha_solicita between cast('${START_DATE}' as date) and
cast('${END_DATE}' as date)
ORDER BY
fac_ayu_solicitud_id
```

B21: Script sql – fac_vehAT(DS)

Paso del ETL de la tabla de hechos ayudas.

Extrae las claves operacionales asociadas a las dimensiones de la tabla de hechos ayudas, extrae los montos y subvenciones asociadas a las ayudas técnicas.

```
SELECT
ay.ayu_id as fac_ayu_solicitud_id,
ay.per_id as dim_per_id,
p.gra_id as dim_ped_grado_discapacidad_id,
pd.dpr_id as dim_ped_deficiencia_principal_id,
cd.cde_id as dim_ped_causa_deficiencia_id,
p.con_id as dim_peh_con_quien_vive_id,
p.viv_id as dim_peh_tipo_vivienda_id,
p.niv_id as dim_pet_nivel_instruccion_id,
p.per_trabaja as dim_pet_trabaja_id,
CASE per_trabaja
WHEN 'SI' THEN tt.tra_id
ELSE ct.ctr_id
END dim_pet_trabaja_descripcion_id,
p.loc_id_provincia as dim_loc_provincia_id,
p.loc_id_canton as dim_loc_canton_id,
p.loc_id_parroquia as dim_loc_parroquia_id,
1 as dim_ayu_clase_id,
ay.tay_id as dim_ayu_id,
cast(ay.ayu_ent_sub as int) as dim_ens_id,
cast(ay.ayu_ent_prov_ser as int) as dim_enp_id,
ay.pro_id as dim_pry_id,
ay.eay_id as dim_eay_id,
cast(to_char(ay.ayu_fecha_solicita,'YYYY-MM-DD') as date) as
dim_tim_solicita_id,
ay.ayu_monto_total as fac_ayu_monto,
ay.ayu_valor as fac_ayu_subvencion
FROM
ayudas_tecnicas ay,
persona p,
per_dpr pd,
per_causa_def cd,
tipo_trabajo tt,
causa_no_trabaja ct
WHERE
p.tpe_id=1 and
pd.pdp_peso=1 and
ay.per_id=p.per_id and
p.per_id=pd.per_id and
p.per_id=cd.per_id and
p.tra_id=tt.tra_id and
p.ctr_id=ct.ctr_id and
ay.ayu_fecha_solicita between cast('${START_DATE}' as date) and
cast('${END_DATE}' as date)
ORDER BY
fac_ayu_solicitud_id
```

B22: Script sql – fac_veh(OR)

Extrae los datos de la tabla de hechos ayudas

```
SELECT
fac_ayu_monto,
fac_ayu_subvencion,
dim_per_key,
dim_ped_key,
dim_peh_key,
dim_pet_key,
dim_loc_key,
dim_ayu_key,
dim_ens_key,
dim_enp_key,
dim_pry_key,
dim_eay_key,
dim_tim_solicita_key
FROM
fact_ayudas,
dim_tiempo
WHERE
dim_tim_solicita_key=dim_tim_key and
dim_tim_fecha_completa between cast('${START_DATE}' as date) and
cast('${END_DATE}' as date)
ORDER BY
dim_per_key,
dim_ped_key,
dim_peh_key,
dim_pet_key,
dim_loc_key,
dim_ayu_key,
dim_ens_key,
dim_enp_key,
dim_pry_key,
dim_eay_key,
dim_tim_solicita_key
```

ANEXO C – Consultas MDX para las plantillas de aplicación de usuario

C1: Reporte: Distribución de PCD¹ por deficiencia principal

```
SELECT      {[Measures].[fac_per_num]}      ON      columns,
{[dim_ped_deficiencia_principal].[Todos].Children} ON rows FROM [Personas
discapacitadas] WHERE ([dim_per_estado].[Todos].[A])
```

C2: Reporte: Porcentaje de PCD por genero

```
WITH      MEMBER [Measures].[Porcentaje] AS '([Measures].[fac_per_num] /
([Measures].[fac_per_num],
[dim_per_genero].CurrentMember.Parent))',format_string = 'Percent' SELECT
{[Measures].[Porcentaje]} ON columns, {[dim_per_genero].[Todos].children} ON
rows FROM [Personas discapacitadas]
```

C3: Reporte: Distribución local de PCD por situación laboral

```
SELECT      Union(CrossJoin({[Measures].[fac_per_num]},
{[dim_pet_trabajo].[Trabaja      (SI/NO)].[NO]}),
CrossJoin({[Measures].[fac_per_num]},      {[dim_pet_trabajo].[Trabaja
(SI/NO)].[SI]})) ON columns, {[dim_loc].[Todos].Children} ON rows FROM
[Personas discapacitadas] WHERE ([dim_per_estado].[Todos].[A])
```

C4: Reporte: Causa de deficiencia por edades

```
SELECT      Hierarchize(CrossJoin({[Measures].[fac_per_num]},
[dim_per_edad].[Todos].Children))      ON      columns,
{[dim_ped_cuasa_deficiencia].[Todos].Children} ON rows FROM [Personas
discapacitadas] WHERE ([dim_per_estado].[Todos].[A])
```

¹ personas con discapacidad

C5: Reporte: PCD registradas por año

```
SELECT                                CrossJoin({[Measures].[fac_per_num]},  
{[dim_tim_carnet].[Todos].[2007], [dim_tim_carnet].[Todos].[2008]}) ON columns,  
{[dim_loc].[Todos].children} ON rows FROM [Personas discapacitadas]
```

C6: Reporte: Distribución local de ayudas técnicas

```
SELECT {[Measures].[fac_ayu_num], [dim_ayu].[Todos].[ayuda tecnica]} ON  
columns, Hierarchize(Union({[dim_loc].[Todos]}, [dim_loc].[Todos].Children)) ON  
rows FROM [Ayudas] WHERE ([dim_eay].[Todos].[ENTREGADO])
```

C7: Reporte: Monto total en ayudas entregadas al año

```
SELECT                                CrossJoin({[Measures].[fac_ayu_monto]},  
{[dim_tim_solicita].[Todos].[2008], [dim_tim_solicita].[Todos].[2009]}) ON  
columns, {[dim_loc].[Todos].children} ON rows FROM [Ayudas] WHERE  
([dim_eay].[Todos].[ENTREGADO])
```

C8: Reporte: Vehículos autorizados al año

```
SELECT      NON      EMPTY      CrossJoin({[Measures].[fac_veh_num]},  
{[dim_tim_autoriza].[Todos].[2007], [dim_tim_autoriza].[Todos].[2008]}) ON  
columns, NON EMPTY {[dim_loc].[Todos].Children} ON rows FROM [Autorizacion  
de vehiculos]
```

C9: Reporte: Tiempo promedio para autorizar un vehículo

```
SELECT      {[Measures].[fac_veh_dias_total]} ON      columns,  
{[dim_veh].[Todos].children} ON rows FROM [Autorizacion de vehiculos]
```


ANEXO D – Estructura del directorio del CD adjunto

- ❖ codigo_fuente
 - bdd
 - etl
 - ejecutables
 - jobs
 - calendarizaciones
 - transformaciones
 - validaciones
 - olap
 - reportes_plantilla
 - mapsic
- ❖ documentos
 - documentos_tecnicos
 - diccionario_datos
 - metadatos_etl
 - manuales
 - manual_instalacion
 - manual_usuario
- ❖ instaladores
 - ambiente
 - aplicaciones
 - data_mart
 - etl
 - olap
- ❖ otros