

**SISTEMA DE SOPORTE A LA TOMA DE DECISIONES EN EL MARCO DE
COMPUTADORES PARA EDUCAR EN LA FASE DE PROFUNDIZACIÓN**



**JAIME MENDOZA
DEYRO ZÚÑIGA**

**UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERIA ELECTRONICA Y TELECOMUNICACIONES
INGENIERIA DE SISTEMAS
POPAYÁN
2010**

TABLA DE CONTENIDO

CAPÍTULO I. INTRODUCCION.....	1
JUSTIFICACION.....	3
Justificación social.....	3
Justificación Tecnológica.....	4
Justificación Académica.....	4
Innovación del Proyecto.....	4
OBJETIVOS.....	5
Objetivo general.....	5
Objetivos específicos.....	5
CAPITULO II. MARCO CONCEPTUAL.....	7
2.1. Programa de Computadores para Educar.....	7
Estrategia de Acompañamiento Educativo del Programa CPE.....	7
Fase de Profundización de la Estrategia de Acompañamiento Educativo.....	7
Competencias definidas en la Fase de Profundización 2008.....	8
2.2. Inteligencia de Negocios.....	8
2.3. Sistema de Soporte a la Toma de Decisiones.....	9
2.4. Bodega de Datos.....	10
Definición de Bodega de Datos.....	10
Terminología y conceptos de Bodega de Datos.....	11
4.1. Metodología para el desarrollo DW.....	12
2.5. Procesamiento Analítico en Línea.....	15
Definición de Procesamiento Analítico en Línea.....	15
Terminología y conceptos de OLAP.....	15
2.6. Minería de Datos.....	15
Definición de Minería de Datos.....	15
Terminología y conceptos de Minería de Datos.....	15
CAPITULO III. DESCRIPCION DE LA METODOLOGIA PARA LA RECOLECCION DE DATOS.....	17
3.1 Metodología para el desarrollo del instrumento de recolección de información ..	17
1. Definición De Objetivos E Hipótesis.....	17
Recolección De Información Adicional.....	18
Definición de las Variables.....	18
Definición de los objetivos.....	19
Definición de las hipótesis.....	19
2. Elaboración del cuestionario.....	20
Diseño del cuestionario.....	20
Determinación de la población.....	22
Determinación del tamaño de la población.....	22
Diseño de la muestra.....	24
Prueba Piloto.....	24
3. Aplicación del cuestionario.....	25
4. Tratamiento Estadístico de los Resultados.....	26
3.2. Herramienta de Recolección de Datos.....	28
CAPITULO IV. DESCRIPCIÓN DEL PROCESO DE DESARROLLO DEL DSS.....	30
4.1.1. PLANEACIÓN DEL PROYECTO.....	30
DEFINICIÓN DEL PROYECTO.....	30

Evaluación de la disponibilidad organizacional para la ejecución del proyecto.....	30
PERSONAS RELACIONADAS CON EL PROYECTO	32
DESARROLLO DEL PLAN DEL PROYECTO.....	32
4.1.2. Definición de Requerimientos del Negocio.....	34
LISTAR LOS DATA MARTS DE LA BODEGA DE DATOS.....	36
LISTAR LAS DIMENSIONES DE LA BODEGA DE DATOS	36
MARCAR LAS INTERCEPCIONES.....	37
Alcance del proyecto	39
4.1.3. Modelado Dimensional	40
MODELADO INICIAL DE ALTO NIVEL.....	41
DESARROLLO DETALLADO DEL MODELO DIMENSIONAL.....	42
DIAGRAMAS DE LOS Data Mart	42
Revisión y validación	45
JERARQUÍAS.....	46
Jerarquías de Usuario	46
Jerarquía - Dimensión Área de Enseñanza.....	46
Jerarquía - Dimensión Beneficiario.....	46
Jerarquía – Dimensión Docente	47
Jerarquía – Dimensión Entidad Educativa.....	47
Jerarquía – Dimensión Fecha.....	47
Jerarquía – Dimensión Grado Enseñanza.....	47
Jerarquía – Dimensión Ruta CPE.....	47
Jerarquía – Dimensión Universidad.....	48
Jerarquía – Dimensión Respuesta Beneficiario.....	48
4.1.4. Diseño Físico del Data Warehouse	48
CONSTRUCCION DE LA BASE DE DATOS MULTIDIMENSIONAL.....	49
ESTRUCTURA DEL CUBO	50
USO DE DIMENSIONES	50
4.1.5. Diseño del Sistema ETL	53
4.1.6. Diseño de la Arquitectura Técnica.....	57
4.1.7. SELECCIÓN E INSTALACIÓN DEL PRODUCTO	57
4.1.8. Especificación de Aplicaciones de Usuario Final	60
Reportes Estándar.....	60
4.1.9. DESARROLLO DE LA APLICACIÓN DE USUARIO FINAL	61
Integración Herramienta OLAP.....	61
4.1.10. Despliegue del DW	65
4.1.11. Mantenimiento y crecimiento	66
4.1.12. Administración del proyecto.....	66
CAPITULO V. CASOS DE DISEÑO DE DW PARA EL MODELADO DIMENSIONAL DE	
HABILIDADES Y DESEMPEÑOS EN EL USO DE LAS TIC	67
5.1. DIMENSION DE HECHO	67
5.2. RELACIONES MUCHOS A MUCHOS.....	70
Puente Dimensión Área de Enseñanza - Dimensión Docente	71
Puente Dimensión Área Enseñanza - Dimensión Estudiante.....	71
Puente Dimensión Grado Enseñanza – Dimensión Docente	72
Puente Dimensión Respuesta – Dimensión Cuestionario	72
5.3. SUBDIMENSION.....	75
Subdimensiones del Hecho Encuesta Beneficiario	75
Subdimensiones de la Dimensión Localización.....	76

JERARQUÍAS DE ORGANIZACIÓN	77
Jerarquía – Dimensión Indicador	78
Jerarquía – Dimensión Respuesta.....	79
INCONVENIENTES PRESENTADOS EN EL DISEÑO.....	80
5.4. DIMENSION AUDITORIA	81
CAPITULO VI. DESCRIPCIÓN DEL PROCESO DE DESARROLLO DEL MODULO DE MINERIA DE DATOS	83
6.1. Entendimiento del negocio	83
Objetivo del Negocio.....	84
Objetivo de Minería de Datos	84
6.2. Entendimiento de datos.....	84
6.3. Preparación de datos	85
6.4. Modelado.....	87
Herramienta para la minería de Datos.....	87
Implementación de los modelos de minería de datos.....	90
Validación de los modelos de minería de datos	90
6.5. Evaluación.....	91
6.6. Despliegue	91
CAPÍTULO VII. CONCLUSIONES Y TRABAJOS FUTUROS	93
Conclusiones	93
Recomendaciones y Trabajos Futuros	94

LISTA DE FIGURAS

Figura 1. Pirámide de los niveles de decisión en una organización, tomado de [13]	9
Figura 2. Arquitectura de un Sistema de Soporte a la Toma de Decisiones (DSS), adaptada de [7].	9
Figura 3. Relación niveles - miembros de la dimensión Calendar (tomado de [20])	12
Figura 4. Ciclo de Vida Dimensional.....	13
Figura 5. Porcentaje de asistencia del docente a las jornadas de formación de la Fase de Profundización	26
Figura 6. Distribución de la población docente por su formación académica.....	27
Figura 7. Docente que ha identificado problemas en el aula o en la institución educativa y se pueden resolver utilizando las TIC	27
Figura 8. Impacto vs Viabilidad	38
Figura 9. Diagrama de Flujo del proceso de modelado dimensional (Adaptado de [8])	41
Figura 10. Diagrama Data Mart Acompañamiento Entidad Educativa	43
Figura 11. Diagrama Data Mart Respuesta Encuesta Beneficiario	44
Figura 12. Diagrama Data Mart Indicador Beneficiario	44
Figura 13. Diagrama Data Mart Jornada de Formación.....	45
Figura 14. Jerarquía – Dimensión Área de Enseñanza.....	46
Figura 15. Configuración ALL IN ONE para un Sistema de Inteligencia de Negocios.....	48
Figura 16. Interfaz de diseño para Cubos Multidimensionales	49
Figura 17. Estructura del Cubo, Medidas y Dimensiones	50
Figura 18. Uso de Dimensiones – Tipos de relaciones entre Dimensiones y Grupos de Medida.....	52
Figura 19. Diagrama de Alto Nivel ETL	54
Figura 20. Paquete ETL de la Dimensión Pregunta	55
Figura 21. Tablas temporales.....	56
Figura 22. Sistema de Auditoria para el DSS	56
Figura 23. Arquitectura un sistema Microsoft de DW/BI, adaptada de [7]	58
Figura 24. Plantilla Estándar.....	60
Figura 25. Reporte Estándar – Indicador por Entidad Educativa	61
Figura 26. Interfaz de selección del cubo	62
Figura 27. Área de trabajo Dundas OLAP Services	63
Figura 28. Consulta – Respuesta por Nivel Educativo (Grafica)	64

Figura 29. Consulta – Respuesta por Nivel Educativo (Grilla)	64
Figura 30. Vista Avanzada de Reportes OLAP	65
Figura 31. Dimensión Encuesta	68
Figura 32. Dimensión Respuesta	68
Figura 33. Dimensión Indicador	69
Figura 34. Dimensión Entidad Educativa	70
Figura 35. Puente Área de Enseñanza – Docente	71
Figura 36. Puente Área de Enseñanza – Estudiante	71
Figura 37. Puente Grado Enseñanza – Docente	72
Figura 38. Puente Respuesta – Cuestionario	73
Figura 39. Puente Respuesta – Pregunta - Encuesta	74
Figura 40. Puente Indicador - Encuesta	74
Figura 41. Subdimensiones Docente y Estudiante	76
Figura 42. Subdimensiones de la Dimensión Localización	77
Figura 43. Dimensión Organization, tomada de [19]	77
Figura 44. Jerarquía – Dimensión Indicador	79
Figura 45. Jerarquía – Dimensión Respuesta	79
Figura 46. Ejemplo niveles Jerarquía – Dimensión Respuesta	80
Figura 47. Sistema del Auditoría del DW	82
Figura 48. Fases del modelo de referencia de CRISP-DM, tomado de [9].	83
Figura 49. Configuración de Estructura de Minería de Datos	87
Figura 50. Configuración algoritmo de reglas de asociación	89
Figura 51. Entorno de Desarrollo de SQL Server Analysis Services para Modulo de Minería de Datos	90
Figura 52. Modelo de minería de datos en el Sistema de Soporte a la Toma de Decisiones de FP-CPE	92

LISTA DE TABLAS

Tabla 1. Rutas de la región Sur-Pacífico definidas por CPE-UNICAUCA	23
Tabla 2. Rutas de la región Sur-Pacífico definidas por CPE-UNICAUCA	23
Tabla 3. Personal administrativo, Técnico y de Desarrollo.....	34
Tabla 4. Temas Analíticos soportados por Indicador Beneficiario	35
Tabla 5. Matriz de Arquitectura Bus de la Bodega de Datos.	38
Tabla 6. Resultado del método de cuatro pasos de Ralph Kimball.....	41
Tabla 7. Comparación Herramientas OLAP.....	59
Tabla 8. Reportes Estándar	60
Tabla 9. Definición Estructura Genérica de Minería de Datos	85
Tabla 10. Estructuras de minería de datos para cada objetivo de minería de datos	86
Tabla 11. Estructuras de minería de datos, (E = Variable de Entrada, P = Variable de Predicción, X = No aplica).....	86
Tabla 12. Técnicas de minería de datos utilizadas en los modelos de minería de datos del proyecto	89

RESUMEN

Modelar y desarrollar un Sistema de Soporte a la Toma de Decisiones basado en tecnologías de Bodegas de datos, Procesamiento Analítico en Línea (OLAP) y Minería de datos , que permita identificar las estrategias de acompañamiento educativo que obtuvieron mejores resultados¹, en las instituciones educativas beneficiadas por el Programa Computadores para Educar durante la Fase de Profundización, ubicadas en la región Sur-pacífico 2 de Colombia durante el año 2007, teniendo en cuenta información de tipo socio-económica de los docentes y estudiantes.

El sistema está compuesto por: una Bodega de Datos en la cual se encuentran almacenados los datos; una herramienta OLAP que permite visualizar los reportes pre-establecidos y consultas Ad hoc (exploración y análisis multidimensional) y un módulo de Minería que permite la extracción automática de patrones, tendencias o correlaciones en los datos recolectados.

La fuente de datos del sistema se basa en una encuesta realizada a docentes y a estudiantes, sobre información personal, familiar, económica, relacionada con el uso del computador (tiempo, habilidades destrezas), institución educativa beneficiada, formador y convenio que acompaña el proceso, estrategia usada (impacto directo o indirecto), entre otras.

CAPÍTULO I. INTRODUCCION

En esta sección se presentara la definición del problema, la justificación del desarrollo del proyecto, el objetivo general y los objetivos específicos que pretende dar solución a la problemática planteada y finalmente se hace una descripción de la estructura de este documento.

Teniendo en cuenta la brecha digital interna del país, el Gobierno Nacional mediante el documento CONPES 3072 “Agenda de Conectividad” [1] diseñó una política de Estado para la masificación de las Tecnologías de la Información y las Comunicaciones (TIC) y se fortalecieron los Programas Compartel y Computadores para Educar (CPE), los cuales han permitido mejorar sustancialmente las condiciones de acceso a las TIC por parte de la población.

El Programa CPE [3] es un esfuerzo conjunto entre el gobierno nacional y la empresa privada, y *“tiene como misión primordial recibir, a título de donación, los equipos de cómputo dados de baja por las entidades públicas y privadas, así como de las personas naturales para reacondicionarlos y entregarlos en óptimas condiciones a instituciones educativas públicas en las distintas regiones del país”*[2], con el objetivo de brindar a estas instituciones acceso a las TIC y promover su uso por medio de la implementación de una estrategia de acompañamiento educativo.

¹ Mejores habilidades y desempeños en el uso de las TIC, expresados en competencias definidas por CPE para la Fase de Profundización [2].

Mediante esta estrategia “se realiza un proceso de formación y acompañamiento que permite compartir y construir con las comunidades educativas² un enfoque pedagógico y unas estrategias didácticas” para fortalecer el aprovechamiento significativo de las TIC en los procesos educativos [2]. Los actores encargados de ejecutar este proceso son las universidades con experiencia en la inclusión de las TIC en los contextos escolares [2]. La Universidad del Cauca con su Grupo de Investigación y Desarrollo en Tecnologías de la Información (GTI) viene trabajando con el Programa CPE desde el año 2006 por medio de los convenios del 2006, 2007, 2008 y 2009, apoyando y participando en la iniciativa de la inclusión de las TIC en la educación básica.

La inclusión de las TIC en la educación [4], contribuye fundamentalmente al desarrollo de un aprendizaje participativo y colaborativo, asimismo, afecta a la institución como organización social en todos sus elementos: el clima organizacional, los medios de instrucción, las relaciones docente-estudiante, estudiante-estudiante, los procesos de aprendizaje, los objetivos y filosofía institucional, los contenidos de instrucción, la metodología utilizada en los cursos y las habilidades que se desarrollan en los estudiantes.

Numerosos son los beneficios que aporta la inclusión de las TIC en el ámbito educativo, sin embargo como se menciona en el portal educativo Colombia Aprende³ del Ministerio de Educación Nacional (MEN) [5], “las TIC por sí solas no son útiles ni eficaces dentro del medio educativo; lo realmente importante es que los directivos, docentes y estudiantes, las apropien, aprendan a adaptarlas a sus prácticas pedagógicas y en los procesos de gestión”.

Dentro del contexto de las TIC [6], “la apropiación tiene que ver con un uso más avanzado de estas tecnologías y está orientado a la conformación e interconexión de espacios de creación y colaboración entre usuarios. Un adecuado nivel de apropiación permitirá la utilización de estas herramientas tecnológicas para estimular el aprendizaje y desarrollar habilidades que contribuyan a la creación de nuevo conocimiento”.

El éxito del proceso de apropiación de las TIC en el entorno educativo, depende en gran parte de una adecuada ejecución del proceso de formación [6]. Sin embargo, la apropiación tiene como característica principal que los usuarios, en este caso docentes y estudiantes, sean capaces de incluir nuevas formas de utilizar la información y de interactuar creativamente con las TIC. Si en este proceso no se alcanzan estos objetivos, estas herramientas serán “subutilizadas”, por ejemplo, utilizarlas solamente para procesar datos sin incluir la participación del usuario (producir conocimiento), y no se obtendrá el valor agregado que generan estas herramientas para avanzar en la formación de ciudadanos en una Sociedad del Aprendizaje.

En otras palabras, la inclusión de las TIC en el entorno educativo no sólo deben resolver el acceso a la información “saber qué” [6], sino que además, debe estimular la conformación de otro tipo de habilidades y desempeños “saber hacer”, que permitan enriquecer el proceso de enseñanza-aprendizaje, estimulando la creatividad, la colaboración, el conocimiento adaptable y la formación continua, entre otras cualidades.

² Alumnos, docentes, personal administrativo, padres de familia y comunidad en general [2].

³ Colombia Aprende “La red del conocimiento” es el portal educativo del MEN [5].

Actualmente, las universidades durante la ejecución del proceso de formación se viene trabajando en el fortalecimiento del “*saber hacer*” sobre las comunidades educativas beneficiadas por el Programa CPE [2], persiguiendo promover sobre éstas, el desarrollo de habilidades y desempeños, expresados en las competencias de la Fase de Profundización (pedagógicas, comunicativas y colaborativas, técnicas y tecnológicas).

En consecuencia de todo lo expresado anteriormente, el presente trabajo de grado plantea como alternativa de solución un Sistema de Soporte a la Toma de Decisiones (“*Decision-Making Support System*”, DSS de sus siglas en ingles) que permita identificar las estrategias de acompañamiento educativo que obtuvieron los mejores resultados utilizadas en el proceso de formación y acompañamiento de las instituciones educativas beneficiadas por el Programa CPE, en particular las instituciones de la región Sur Pacifico 2 (comprende los departamentos de Nariño, Putumayo y el sur del Cauca) que realizaron la Fase de Profundización en el año 2007. Al mismo tiempo, permite realizar exploraciones y análisis sobre la información socio-económica de los docentes y estudiantes, en función de los avances logrados en dichas competencias, como también encontrar patrones, tendencias o correlaciones ocultas en dicha información.

Todo lo anterior buscando dar a las universidades y a CPE un soporte tecnológico en la toma de decisiones, en el establecimiento de los convenios, y de las estrategias de acompañamiento educativo que se emplean en las jornadas de formación y acompañamiento en las diferentes instituciones beneficiadas por el Programa CPE.

JUSTIFICACION

Justificación social

Este proyecto de grado contribuye en el desarrollo de las comunidades educativas de bajo desarrollo ubicadas en la región Sur Pacifico 2, ya que el DSS es una herramienta estratégica para la Universidad del Cauca y el Programa CPE, por un lado dará a los gestores de las estrategias de acompañamiento un soporte u otro criterio en la toma de decisiones para el establecimiento de las estrategias de acompañamiento educativo que se emplean en las jornadas de formación, de tal manera que estas comunidades educativas, en particular docentes y estudiantes, adquieran mejores habilidades y desempeños en TIC al servicio del mejoramiento de la calidad y equidad de la educación, y la competitividad de los mismos y de la región, teniendo en cuenta que estas instituciones educativas se encuentran distantes geográficamente de las ciudades capitales, las cuales tienen mayor acceso al conocimiento.

Por otro lado, a los docentes indirectamente, les permitirá mejorar sus prácticas pedagógicas y metodológicas utilizadas en los cursos y por parte de los estudiantes, poder mejorar sus métodos de estudio y desarrollar sus habilidades en función del ideal de hombre que exige la Sociedad del Conocimiento y la Información⁴. Todo lo anterior en pro del proceso auto-formativo en la enseñanza y el aprendizaje.

⁴ Es una sociedad “*centrada en la persona, integradora y orientada al desarrollo, en que todos puedan crear, consultar, utilizar y compartir la información y el conocimiento, para que las*

Justificación Tecnológica

En este proyecto de grado se aplican metodologías y tecnologías de inteligencia de negocios para integrar, almacenar y analizar la información recopilada de docentes y estudiantes de las instituciones beneficiadas por el Programa CPE, logrando crear un marco de trabajo que ayude a estudiar e investigar, cómo algunas técnicas de Minería de Datos (“*Data Mining*”, DM de sus siglas en ingles), Bodegas de Datos (“*Data Warehouse*”, DW de sus siglas en ingles) y Procesamiento Analítico En Línea (“*On Line Analytical Processing*”, OLAP de sus siglas en ingles) pueden ser aplicadas para la identificación de estrategias de acompañamiento educativo en el marco del Programa CPE y en el descubrimiento de correlaciones, tendencias y patrones ocultos en la información socio-económica de la comunidad educativa, en particular docentes y estudiantes, de estas instituciones.

Justificación Académica

El proyecto de grado aporta al Grupo de Investigación y Desarrollo en Tecnologías de la Información (GTI), especialmente en las áreas de investigación DW, OLAP y DM, por el conocimiento que se desarrolló sobre como los DSS pueden apoyar a las instituciones educativas de regiones apartadas de Colombia beneficiadas por el Programa CPE, para que las estrategias de acompañamiento educativo que se utilicen en jornadas de formación, sean las más adecuadas de acuerdo a las condiciones particulares de cada institución y de la región.

Por otra parte, el DSS sirve como modelo para ser utilizado en las otras regiones en las cuales se encuentre en funcionamiento el Programa, de tal manera que se puedan obtener a futuro los mismos o mejores resultados de los logrados en la región Sur Pacifico 2, asimismo, dichos resultados se exterioricen a las otras universidades del convenio, para su uso y mejoramiento futuro.

Innovación del Proyecto

Hasta el momento CPE no cuenta con ningún sistema que permita realizar un análisis sobre las estrategias de acompañamiento educativo, de ahí que este trabajo de grado se constituye en una innovación a nivel tecnológico y educativo en este campo de investigación, ya que se plantea un DSS, que le permitirá a CPE identificar las estrategias de acompañamiento educativo con mejores resultados en las comunidades educativas beneficiadas.

Además en este proyecto se empleó tecnologías como DM (tecnologías que no se han usado hasta el momento en este contexto), que permiten realizar un proceso de descubrimiento de nuevas correlaciones, patrones y/o tendencias ocultas en la información almacenada en el DW, de tipo socio-económico y sobre las habilidades y desempeños expresados en competencias definidas por CPE para la FP (pedagógicas, comunicativas y colaborativas, técnicas y tecnológicas).

personas, las comunidades y los pueblos puedan emplear plenamente sus posibilidades en la promoción de su desarrollo sostenible y en la mejora de su calidad de vida, ...” [1].

SISTEMA DE SOPORTE A LA TOMA DE DECISIONES EN EL MARCO DE COMPUTADORES PARA EDUCAR
(CPE) EN LA FASE DE PROFUNDIZACIÓN (FP)

OBJETIVOS

Objetivo general

- Modelar y desarrollar un Sistema de Soporte a la Toma de Decisiones (DSS) basado en tecnologías de DW, OLAP y DM, que permita identificar las estrategias de acompañamiento educativo que obtuvieron mejores resultados⁵, en las instituciones educativas beneficiadas por el Programa Computadores para Educar (CPE) durante la Fase de Profundización, ubicadas en la región Sur-pacífico 2 de Colombia durante el año 2007, teniendo en cuenta información de tipo socio-económica de los docentes y estudiantes.

Objetivos específicos

- Determinar⁶ el conjunto de casos de diseño de DW que se debe tener en cuenta para el modelado dimensional de información de habilidades y desempeños en el uso de las TIC adquiridos por docentes y estudiantes de instituciones educativas del programa CPE.
- Construir un DW para las instituciones educativas de la Fase de Profundización región Sur-pacífico 2, utilizando los casos de diseño determinados para este contexto, que almacene e integre información de tipo socio-económica, y sobre habilidades y desempeños en el uso de las TIC adquiridos por docentes y estudiantes de 35 instituciones educativas durante la Fase de Profundización.
- Modelar y construir un módulo de DM, que permita realizar un proceso de descubrimiento de nuevas correlaciones, patrones y/o tendencias ocultas en la información almacenada en el DW.
- Integrar una herramienta OLAP al DSS, que permita realizar consultas para identificar las estrategias de acompañamiento educativo con los mejores resultados en las instituciones educativas seleccionadas para el desarrollo de este proyecto.

ESTRUCTURA DEL DOCUMENTO

El documento se compone de los siguientes capítulos:

- Capítulo I. Introducción: Permite tener una visión general del proyecto, el origen, la justificación y los objetivos del proyecto.
- Capítulo II. Marco Conceptual: Contiene los conceptos teóricos que se necesitaron para el desarrollo del proyecto, esta información está relacionada con los conceptos fundamentales de DW, DSS, OLAP y DM.
- Capítulo III. Descripción de la Metodología para la Recolección de Datos: Se hace una descripción del proceso de desarrollo que se siguió para la recolección de la información de los docentes, estudiantes, formadores e instituciones educativas. Esta información abarca aspectos socio-económicos y de competencias de los docentes y estudiantes. El instrumento que se utilizó fue la encuesta.

⁵ Mejores habilidades y desempeños en el uso de las TIC, expresados en competencias definidas por CPE para la Fase de Profundización [2].

⁶ Distinguir, discernir.

- Capítulo IV. Descripción del Proceso de Desarrollo del DSS: Se hace una descripción de todo el ciclo de vida usado para el desarrollo del sistema de DSS, desde su concepción inicial hasta su implementación y despliegue, junto con la integración de la herramienta OLAP. Incluye la presentación de algunos reportes estándar definidos por el usuario.
- Capítulo V. Casos de Diseño de DW para el Modelado Dimensional de Habilidades y Desempeños en el Uso de las TIC: Contiene la descripción tanto a nivel conceptual, lógico y aspectos de implementación de los diferentes casos que se presentaron durante el modelamiento e implementación del DW para este campo de la educación. Además se describen algunos problemas presentados durante el modelado dimensional.
- Capítulo VI. Descripción del Proceso de Desarrollo del Modulo de Minería: Se hace una descripción de todo el ciclo de vida usado para el desarrollo del módulo de Minería de datos, con algunas relaciones encontradas entre las variables socio-económicas y las competencias acompañados por los docentes.
- Capítulo VII. Conclusiones y Trabajo Futuro: Se presentan las conclusiones del proyecto y las recomendaciones para el trabajo futuro en esta área.
- Capítulo VIII. Bibliografía: Se presenta la bibliografía empleada para la realización del proyecto

CAPITULO II. MARCO CONCEPTUAL

En los siguientes ítems se explican detalladamente los conceptos teóricos sobre DW, DSS, OLAP y DM, que se utilizaron para el desarrollo de este proyecto.

2.1. Programa de Computadores para Educar

El programa Computadores para Educar (CPE) se encuentra liderado por parte de la Presidencia de la República, además de la contribución del Ministerio de Comunicaciones, el Ministerio de Educación Nacional (MEN), el Servicio Nacional de Aprendizaje (SENA) y con el apoyo de varias empresas privadas. El programa adicionalmente cuenta con el apoyo del Gobierno de Canadá, quien desarrolla el programa “Computers for Schools” desde el año 1993 con gran desempeño; y ha ofrecido asesoramiento a Colombia con la implementación del programa CPE desde sus inicios.

CPE, “es un programa de reciclaje tecnológico cuyo objetivo es brindar acceso a las tecnologías de información y comunicaciones a instituciones educativas públicas del país, mediante el reacondicionamiento, ensamble y mantenimiento de equipos, y promover su uso y aprovechamiento significativo en los procesos educativos, a través de la implementación de estrategias de acompañamiento educativo y apropiación de TIC’s”[10].

Estrategia de Acompañamiento Educativo del Programa CPE

CPE además de entregar los equipos a los colegios y escuelas, realiza todo un proceso de acompañamiento a las instituciones educativas y comunidades que reciben estos recursos, brindando la formación y la asesoría educativa que permita hacer el uso efectivo de las herramientas y sacarles el mejor provecho. Para llevar a cabo esto, “*Computadores para Educar desarrolla en cada una de las instituciones educativas y comunidades que beneficia una Estrategia de Acompañamiento Educativo de dos años, dividida en dos Fases, la Fase Inicial y la Fase de Profundización. Durante dichas fases se desarrollan estrategias encaminadas a la incorporación y el uso efectivo de las TIC en la educación, fundamentado en tres ejes temáticos de acción⁷: Gestión Institución, Infraestructura Tecnológica y Pedagogía*”[11].

Fase de Profundización de la Estrategia de Acompañamiento Educativo

La Fase de Profundización (FP) es “*una fase complementaria que se realiza después de terminada la Fase Inicial; como su nombre lo indica pretende profundizar, con especial énfasis en los objetivos pedagógicos, además de los de gestión e infraestructura tecnológica, en la formación y la asesoría a las escuelas beneficiarias con un acompañamiento de 12 meses*” [11].

La FP es ejecutada a través de convenios interadministrativos de cooperación y cofinanciación, con varias universidades del país (Antioquia, Autónoma de Bucaramanga,

⁷ Los ejes son los componentes básicos alrededor de los cuales se articulan los objetivos y acciones específicas que el Programa desarrolla en su estrategia de acompañamiento [10].

Cauca, Distrital Francisco José de Caldas, Industrial de Santander y Pedagógica Nacional, entre otras) que están distribuidas en diferentes regiones del país [11].

Competencias definidas en la Fase de Profundización 2008

CPE persigue promover en los beneficiarios de la estrategia, el desarrollo de habilidades y desempeños, expresados en las siguientes competencias (excluyendo las competencias éticas, administrativas y de gestión, que no son de interés para el proyecto) [2]:

Competencias Pedagógicas

- Reconocer potencialidades de las TIC en los procesos de enseñanza, aprendizaje.
- Comprender las oportunidades y retos que ofrece el uso de tecnologías de información y comunicación (TIC) en contextos educativos.
- Aprovechar en el escenario pedagógico las potencialidades que le ofrecen las TIC de acuerdo con los recursos disponibles.

Competencias Comunicativas y Colaborativas

- Compartir los aprendizajes y las construcciones de su proceso de integración de las TIC a la práctica pedagógica.
- Interactuar con otros utilizando y compartiendo diversos recursos digitales, sincrónica y/o asincrónicamente.
- Comprender las implicaciones del trabajo colaborativo en redes y comunidades (presenciales y a distancia), haciendo uso de ellas en su vida personal y profesional.
- Participar en redes o comunidades virtuales y/o presenciales (en roles periféricos ó centrales) (El término virtual puede ser entendido como comunicación a distancia).

Competencias Técnicas y Tecnológicas

- Conocer y aplicar procedimientos básicos para el funcionamiento adecuado de los computadores.
- Configurar y mantener redes informáticas en el aula de cómputo.
- Dominar conceptos y funciones básicas que permitan usar fluidamente las TIC, según su nivel de apropiación.
- Seleccionar y utilizar aplicaciones fluidamente según su nivel de apropiación y producto requerido.
- Buscar y utilizar información en Web.

2.2. Inteligencia de Negocios

La Inteligencia de Negocios (Business Intelligence – BI, por sus siglas en inglés) es el conjunto de procesos y estructuras de datos usados para analizar los datos y la información usada en el soporte de toma de decisiones. Los componentes del BI son: DW, los Data Mart, la interfaz al DSS y los procesos para ingresar datos en el DW (conjunto de actividades para capturar datos de los sistemas operacionales y luego migrarlos al DW) y obtener información del DW (conjunto de actividades que entregan información del DW y los hace accesible a los usuarios finales) [12]. La inteligencia de negocios se puede aplicar a todos los niveles de decisión en una organización (Ver Figura 1): nivel operativo, nivel táctico y nivel estratégico [13].



Figura 1. Pirámide de los niveles de decisión en una organización, tomado de [13]

2.3. Sistema de Soporte a la Toma de Decisiones

Un Sistema de Soporte a la Toma de Decisiones (Decision-Making Support System – DSS, por sus siglas en inglés) es un sistema de BI que transforma los datos que tiene una organización en información, con el objetivo de tomar decisiones estratégicas basadas en un análisis dimensional [7].

El DSS traduce los requerimientos de los usuarios en sentencias de consulta correspondientes en el DW (núcleo central del DSS) y es el que interpreta los resultados devueltos para mostrarlos según lo solicitado por el usuario, por medio de herramientas OLAP que permiten navegar a través de los datos almacenados en el DW y analizarlos dinámicamente desde una perspectiva multidimensional [7].

Los DSS pueden ser complementados por la DM que constituye un paso más en el análisis de los datos para apoyar la toma de decisiones, permitiendo realizar un análisis más avanzado de los datos y extraer más información de ellos [7]. La Figura 2, muestra la arquitectura de un Sistema de Soporte a la Toma de Decisiones.

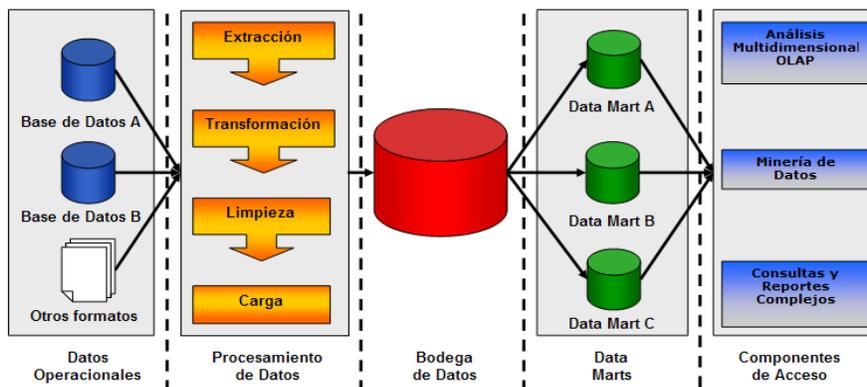


Figura 2. Arquitectura de un Sistema de Soporte a la Toma de Decisiones (DSS), adaptada de [7].

Datos Operacionales: Este componente hace referencia a las fuentes de los datos para la bodega de datos, que contienen los datos operacionales de una organización, estos datos generalmente se encuentran almacenados en sistemas de procesamiento

transaccional en línea (On-Line Transaction Processing -OLTP, por sus siglas en inglés), Hojas de cálculo y Archivos planos.

Procesamiento de Datos: Se encarga del proceso de extracción, transformación y carga (ETL) de los datos de las fuentes de datos operacionales a la bodega de datos. Este componente permite conectarse a diferentes fuentes de datos (bases de datos, archivos estructurados, archivos planos, etc.), extraer los datos de las mismas, transformarlos y cargarlos a una fuente de datos diferente.

Bodega de Datos: Hace referencia a la base de datos donde se almacenarán e integrarán la información de diferentes fuentes de datos relacionadas con los procesos del negocio, permitiendo realizar análisis multidimensional de dichos datos para obtener información relevante.

Componentes de Acceso: Estos componentes hacen referencia a las diferentes herramientas de acceso a los datos de la bodega de datos para obtener información y presentarla al usuario de una manera más amigable y le permite realizar un análisis visual de dicha información.

Análisis Multidimensional – OLAP: Este componente es una tecnología que permite la explotación de datos en diferentes puntos de vista o dimensiones, permitiendo el análisis de datos interactivo a través de una interfaz humano-computador.

2.4. Bodega de Datos

Definición de Bodega de Datos

Una Bodega de Datos “es una colección orientada a temas, integrada, variable en el tiempo y no volátil de datos para soportar las decisiones de gerencia” [14]. El DW es el punto central de integración de datos para la inteligencia de negocios y es la fuente de datos para los Data Mart, entregando una vista común de los datos de la empresa [12].

Las características del DW son [14]:

- *Orientada a temas*, significa que todos los datos relevantes sobre un tema son recogidos y guardados como un conjunto sencillo de datos en un formato práctico. La información es presentada de acuerdo a los temas específicos o áreas de interés.
- *Integrada*, hace referencia a los datos que están guardados globalmente en un modo compatible con las convenciones, las medidas, las estructuras de codificación y los atributos físicos con nombres consistentes, incluso cuando los sistemas operacionales almacenan los datos de manera diferente.
- *No volátil*, significa que la información es estable, que no cambia cada vez que un proceso operacional es ejecutado. La información es consistente sin considerar cuándo será accedido el DW.
- *Variable en el tiempo*, quiere decir que el DW contiene tanto la información historia del tema como la actual. Los datos del DW representan los datos a largo plazo de cinco a diez años.

Terminología y conceptos de Bodega de Datos

Atributo: Es el nivel más bajo de información relacionada a cualquier entidad, modela una propiedad de una entidad específica. El modelado dimensional tiene una definición más restrictiva; referencia la información que describe las características de una dimensión [12].

Medida: Es un término del modelado dimensional que se refiere a valores, que miden algún aspecto del negocio. Las medidas residen en las tablas de hechos. Un evento de medición siempre puede producir un registro en la tabla de hecho. Estos valores también son llamados *hechos* en los servicios de análisis [12].

Tabla de Dimensión: Una tabla de dimensión es un conjunto de tablas de referencias que proveen la base para las consultas de restricción y agrupación de la información en una tabla de hecho dentro de un modelo dimensional. Las tablas de dimensiones contienen información descriptiva y jerárquica. Una *tabla de dimensión conformada* es construida para ser usada por múltiples Data Mart [12].

Tabla de Hecho: Cada tabla de hecho contiene las mediciones y métricas asociadas con un proceso específico del negocio. Un registro en una tabla de hecho es una medición. La llave primaria para la tabla de hecho es una llave múltiple formada por un subconjunto de claves foráneas de cada dimensión involucrada en el evento del negocio [15].

Datos Departamentales: El Data Mart es un conjunto de datos personalizados y/o resumidos que son obtenidos del DW y adaptados para soportar los requerimientos analíticos específicos de una unidad o función del negocio. Utiliza una vista común de la empresa de los datos estratégicos y provee unidades del negocio con más flexibilidad, control y responsabilidad [12].

Granularidad: El nivel del detalle contenido en una tabla de hecho es llamado granularidad. Las tablas de hechos con el nivel más bajo de detalle igual al de la fuente original, generalmente es conocido como el nivel atómico. Cada tabla de hecho posee su propia granularidad [15].

Indicador de Desempeño Clave: Un Indicador de Desempeño Clave es una métrica que provee a los usuarios del negocio una indicación del desempeño actual e histórico de un aspecto del negocio [12].

Jerarquías: Las jerarquías son principalmente modos estándar para agrupar datos dentro de una dimensión. Comenzando desde un nivel alto de detalle hasta un nivel más bajo (drill down). Las jerarquías más comunes son las organizacionales (elementos primarios y secundarios), jerarquías de usuario, jerarquías geográficas basadas en una localización física, jerarquías de productos correspondientes a categorías y subcategorías, jerarquías de responsabilidades, jerarquías temporales, entre otras [8].

Jerarquías de Usuario: Las jerarquías definidas por el usuario son jerarquías de atributos que se utilizan en Microsoft SQL Server Analysis Services, para organizar los miembros de una dimensión en estructuras jerárquicas y proporcionar rutas de exploración en un cubo. Por ejemplo, en una dimensión de tiempo que admite tres atributos denominados Year, Quarter y Month, se utilizan para construir una jerarquía definida por el usuario

llamada Calendar. La relación entre los niveles y los miembros de la dimensión Calendar (una dimensión regular) se muestra en la siguiente figura [20].

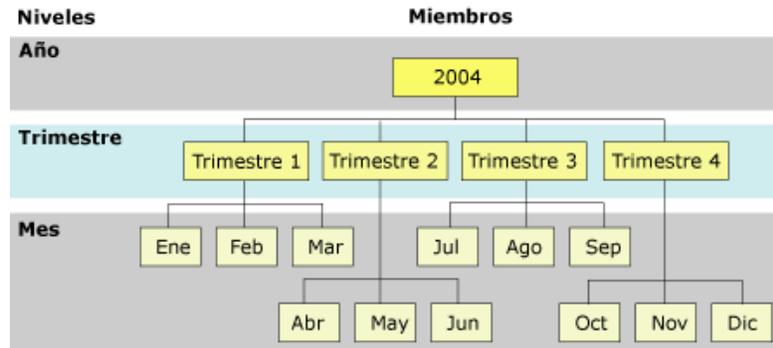


Figura 3. Relación niveles - miembros de la dimensión Calendar (tomado de [20])

Método de los cuatro pasos de Kimball: método utilizado para realizar el Modelado Inicial de Alto Nivel [7], compuesto por los siguientes pasos:

1. Escoger el proceso del negocio: Se escoge un proceso del negocio de la matriz bus, que obtenga los datos en lo posible de una fuente de datos y que no sea muy compleja.
2. Declarar la granularidad: Se declara la granularidad, que es equivalente a decir cuál es el registro individual de la tabla de hechos (transacciones, línea de elemento o foto).
3. Seleccionar las dimensiones: Se escogen las dimensiones para la tabla de hechos. Si una dimensión no concuerda con la granularidad, el equipo de diseño tiene dos caminos: borrar la dimensión ó cambiar la declaración del granularidad. Se deben identificar algunos atributos descriptivos de cada dimensión, hacer una lista de todos los atributos descriptivos conocidos. En este momento del diseño no se especifica la fuente de datos del atributo.
4. Escoger los hechos: En este paso se adicionan todas las medidas de acuerdo con el contexto de la granularidad declarada previamente.

4.1. Metodología para el desarrollo DW

La metodología propuesta por Ralph Kimball, es una de las metodologías más conocidas, empleadas y con bases conceptuales bien definidas [7]. En esta metodología, se aprecian diferentes etapas que deben ser llevadas a cabo en la construcción de un DW. En la Figura 4. **Ciclo de Vida Dimensional** se muestra una vista general del mapa de ruta de un proyecto de DW, y presenta la secuencialidad de tareas de alto nivel requeridas para el diseño, desarrollo e implementación del DW.

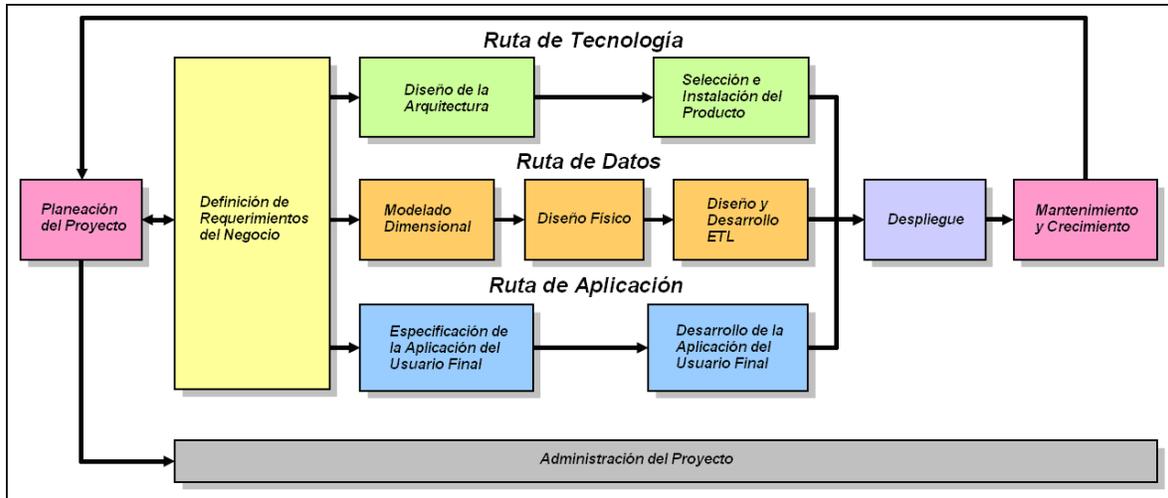


Figura 4. Ciclo de Vida Dimensional

Este ciclo de vida se caracteriza por ser cíclico y enfocarse en la identificación de los requerimientos del negocio, además se basa en incrementos que dan soporte a procesos de negocios específicos que tienen alto impacto analítico en el mismo.

Planeación del Proyecto: El objetivo de esta etapa fue identificar el escenario del proyecto para saber de dónde surgió la necesidad del DW y poder de esta manera construir el alcance del proyecto, se incluye el impacto y la evaluación de factibilidad. Esta etapa depende de los requerimientos, ya que a partir de estos se determina el alcance del proyecto y los recursos requeridos. Para llevar a cabo esta tarea de alto nivel fue necesario tener las especificaciones de las actividades y las tareas necesarias, para luego definir el alcance y la justificación del proyecto [7].

Definición de requerimientos del Negocio: Los requerimientos de los usuarios tienen un gran impacto en las decisiones que se toman durante el proyecto, ya que los requerimientos son parte fundamental y el centro de las bodegas de datos.

Modelamiento Dimensional: Es una metodología para el modelado lógico de datos, un modelo dimensional contiene una estructura simétrica propia que permite rendimiento de consultas, facilidad de uso y entendimiento, flexibilidad a los cambios. El modelo dimensional está compuesto por una tabla central llamada Tabla de Hechos y sus dimensiones asociadas por lo que recibe el nombre de Esquema Estrella [8].

Diseño Físico: etapa enfocada en la definición de las estructuras físicas necesarias para soportar el diseño lógico de la base de datos, se realiza la definición de convenciones estándares de nombres, como también las estrategias de particionamiento.

Diseño y Desarrollo ETL: Las actividades principales de este proceso son la Extracción que se utiliza para obtener los datos desde diferentes fuentes de datos; la Transformación que hace referencia a la limpieza, homogenización y conversión de los datos; y la Carga de datos que se encarga de poblar los datos en la bodega de datos [7].

Diseño de la Arquitectura: Dentro de un ambiente de DW/BI se presenta la integración de varias tecnologías, es por esto que para poder definir su arquitectura técnica es necesario tener en cuenta aspectos tales como: los requerimientos del negocio, los ambientes técnicos actuales y lineamientos técnicos de planeación [7]. En esta etapa se identifican las herramientas y tecnologías que se requieren para poder desarrollar el sistema.

Selección e Instalación del Producto: Basado en la arquitectura técnica del DSS, se evalúan y se seleccionan los componentes específicos para la arquitectura, tales como la plataforma hardware, el motor de base de datos, la herramienta de ETL, herramientas de acceso y otros. Una vez evaluados y seleccionados los productos, se procede con la instalación y prueba de los mismos en un ambiente integrado de DW [7].

Especificación de Aplicación de Usuario Final: Se definen una serie de herramientas estándares, ya que no todos los usuarios accederán de la misma manera a la bodega de datos, se definen plantillas de reportes, parámetros que maneja el usuario y cálculos requeridos.

Desarrollo de la Aplicación de Usuario Final: esta fase se enfoca básicamente en configurar la herramienta de metadata y construir los reportes específicos.

Despliegue del DW: En esta fase es donde convergen las tres rutas del ciclo de vida, la ruta de la tecnología, de datos y la de aplicaciones. En esta fase se realizan actividades que deben de ser hechas antes de poner en producción el DW, garantizando de esta manera un buen funcionamiento del producto. Se realizan actividades como: la configuración del hardware, configuraciones de red, conexiones a bases de datos, chequeos de seguridad, instalación de software, documentación, manuales y capacitaciones a los usuarios, etc. [7].

Mantenimiento y crecimiento: fase del ciclo de vida dimensional, es un proceso con etapas bien definidas y de naturaleza espiral que se desarrolla durante la evolución de la organización. Es de gran interés e importancia definir las prioridades para tener en cuenta los nuevos requerimientos de los usuarios para permitir que el sistema evolucione y crezca. Se debe tener en cuenta también varios aspectos para mantener el DW, tales como: el continuo soporte y constante capacitación de los usuarios del negocio, el manejo de la infraestructura (monitoreo de la base de datos, tráfico, etc.), mejorar el rendimiento sobre las consultas, mantenimiento de la metada y procesos ETL [7].

Administración del proyecto: En esta fase se realiza el monitoreo, control y administración de todas las actividades del proyecto que se llevan a cabo durante todo el Ciclo de Vida Dimensional. Todas las actividades que involucra se enfocan en verificar el estado del proyecto y evolución con el tiempo [7].

2.5. Procesamiento Analítico en Línea

Definición de Procesamiento Analítico en Línea

El Procesamiento Analítico en Línea (On-Line Analytical Processing – OLAP, por sus siglas en inglés), es un término establecido por E.F. Codd que se refiere a cualquier software que permite el análisis de datos interactivo a través de una interfaz humano-computador. Es comúnmente usado para etiquetar una categoría de tecnologías de software que permite a los usuarios llevar a cabo el acceso de datos ad hoc y al análisis basado en su dimensionalidad. Esta forma de análisis multidimensional provee un discernimiento del negocio a través del acceso rápido, consistente e interactivo a una gran variedad de vistas posibles de la información [12].

Terminología y conceptos de OLAP

Cubo OLAP: Codd propuso un conjunto de datos en arreglos o cubos para permitir el análisis rápido. El conjunto de datos dentro de cubos evita una limitación de las bases de datos relacionales que no son muy adecuadas para el análisis instantáneo de grandes cantidades de datos [12].

Jerarquía: A veces llamada *árbol*, es un tipo especial de relación *padre-hijo*. En una jerarquía, un hijo representa un nivel más bajo de detalle o granularidad del padre. Este crea un sentido de propiedad o control que una entidad padre tiene sobre una hijo [12].

2.6. Minería de Datos

Definición de Minería de Datos

La Minería de Datos (Data Mining – DM, por sus siglas en inglés) *"es el proceso de descubrir nuevas y significantes correlaciones, patrones y tendencias en grandes cantidades de datos almacenados en repositorios usando tecnologías de reconocimiento de patrones así como técnicas estadísticas y matemáticas"* [16].

Terminología y conceptos de Minería de Datos

Modelar: Habilidad o capacidad para aplicar algoritmos a un conjunto de datos para predecir atributos objetivos [9].

Tipos de Problemas de DM: Usualmente, los proyectos de DM involucran una combinación de diferentes tipos de problemas, que juntos solucionan el problema del negocio [9]. A continuación se presentan los tipos de problemas de DM más comunes [9]:

- *Descripción y Sumarización de Datos*: permite la descripción concisa de las características de los datos, básicamente en su forma primaria y agregada.
- *Segmentación*: separa los datos en subgrupos o clases significativas, todos los miembros de un subgrupo comparten características comunes. Algunas técnicas son: técnicas de agrupamiento, redes neuronales y visualización.

- *Descripción de Conceptos*: El propósito de esta técnica no es generar modelos con alta exactitud de predicción, si el de ofrecer una descripción más comprensible de los conceptos o clases. Algunas técnicas son: métodos de inducción de reglas, agrupamiento conceptual.
- *Clasificación*: se asume que hay un conjunto de objetos que pertenecen a clases diferentes. La etiqueta de clase es un valor (simbólico) discontinuo y es conocido para cada objeto. Algunas técnicas son: análisis de discriminante, métodos de inducción de reglas, aprendizaje de árboles de decisión, k-vecino más cercano, razonamiento basado en casos, redes neuronales y algoritmos genéticos.
- *Predicción*: es muy similar a la clasificación. La diferencia es que en la predicción el atributo objetivo (clase) no es un atributo discontinuo cualitativo es continuo. Algunas técnicas son: análisis de regresión, árboles de regresión, k-vecino más cercano, redes neuronales, métodos de box-jenkins y algoritmos genéticos.
- *Análisis de Dependencias*: consiste en encontrar un modelo que describe dependencias (o asociaciones) importantes entre ítems de datos. Las dependencias pueden ser usadas para predecir el valor de un ítem de datos dada la información sobre otros ítems. Algunas técnicas son: análisis de correlación, análisis de regresión, reglas de asociación, redes Bayesianas, programación lógica inductiva y técnicas de visualización.

CAPITULO III. DESCRIPCION DE LA METODOLOGIA PARA LA RECOLECCION DE DATOS

A continuación se presenta el desarrollo de la metodología utilizada para realizar el proceso de recolección de datos.

3.1 Metodología para el desarrollo del instrumento de recolección de información

La técnica de la encuesta nos permite obtener información sobre un problema o algún aspecto del mismo, a través de una serie de preguntas, previamente establecidas, dirigidas a las personas implicadas en el tema de estudio, en nuestro caso, a los docentes que participaron en las jornadas de formación de la Fase de Profundización del 2007 y estudiantes de las instituciones beneficiadas.

Para aplicar la técnica de la encuesta, se revisaron varias metodologías de las cuales se identificaron y se tuvieron en cuenta las recomendaciones, procesos y fases más importantes de ellas (Ver **Anexo 1**), tomando como principal metodología guía para el proceso de recolección de datos, la propuesta por Margarita Oncins de Frutos [17], la cual está dividida en las siguientes fases:

1. Definición de objetivos e hipótesis.
2. Elaboración del cuestionario.
3. Aplicación del cuestionario.
4. Tratamiento estadístico de los resultados.
5. Elaboración del informe.

A continuación se describen brevemente cada una estas fases con sus respectivas actividades.

1. Definición De Objetivos E Hipótesis

Formulación del problema

En esta actividad se hace un planteamiento claro y delimitado tanto como sea posible del problema, en el cual se describe qué es lo que se va a estudiar. El planteamiento del problema del proyecto de investigación es el siguiente:

Planteamiento del problema

Las universidades están enfocando el proceso de formación hacia el fortalecimiento del “saber hacer” sobre las comunidades educativas beneficiadas por el Programa CPE, persiguiendo promover sobre éstas, el desarrollo de saberes, habilidades y desempeños, expresados a través de las competencias (pedagógicas, comunicativas y colaborativas, técnicas y tecnológicas) de la Fase de Profundización [2]. Es por esto que es necesario identificar el nivel de las competencias adquiridas por las comunidades beneficiadas, para

determinar cómo es el uso realizado de las TIC en sus actividades y prácticas pedagógicas, comunicativas y colaborativas, y técnicas y tecnológicas, y poder llegar a evaluar el proceso de formación que se está realizando con los docentes de las instituciones educativas beneficiadas por el programa CPE.

Por lo anterior, el planteamiento del problema para la encuesta de los docentes: *“Identificar cómo están siendo utilizadas las TIC en las actividades y prácticas pedagógicas, comunicativas y colaborativas, y técnicas y tecnológicas; de los docentes que asistieron las Jornadas de formación de la Fase de Profundización del año 2007”*.

Mientras que el planteamiento del problema para la encuesta de los estudiantes es: *“Identificar como están siendo utilizadas las TIC en las actividades y prácticas de aprendizaje, comunicativas y colaborativas, y técnicas y tecnológicas; de estudiantes de las instituciones educativas a las cuales están vinculados los docentes beneficiados por las Jornadas de formación en el año 2007”*.

Recolección De Información Adicional

En esta actividad se debe recurrir a la búsqueda de bibliografía, o bien a técnicas de tipo exploratorio, para obtener más información sobre el tema, no cometer los errores que otros estudios hayan puesto de manifiesto, etc. Para ver en detalle el proceso realizado en esta etapa ver el **Anexo 2**.

Definición de las Variables

En esta actividad se identifican las variables a medir, estas son los aspectos de la realidad que influyen en el problema. Se trata pues, de especificar estos aspectos y ver cuáles de ellos son determinantes en cada situación.

Docentes

Las variables principales a medir en este estudio de los docentes, son las habilidades, actitudes y desempeños en el uso de las TIC, expresadas a través de las competencias pedagógicas, colaborativas y comunicativas, y técnicas y tecnológicas de la Fase de Profundización [2]; además de las variables relacionadas con la información general (identificación, demográfica, formación académica e información laboral, entre otra) y socio-económica del docente.

Estudiantes

Las variables a medir de los estudiantes de primaria y secundaria básica, son las habilidades, actitudes y desempeños en el uso de las TIC, expresadas a través de las competencias colaborativas y comunicativas, y técnicas y tecnológicas, además de las variables relacionadas con la información general del estudiante (identificación, demográfica, situación académica e información familiar).

Como las competencias para los estudiantes no estaban definidas por CPE, el grupo de trabajo se encargó de establecer un conjunto de habilidades y desempeños para estos (estudiantes de primaria básica y estudiantes de secundaria básica), acordes con lo requerido por el proyecto de investigación. Para esta tarea se tuvo como referencia principal la Evaluación e investigación educativa “Competencias básicas en las Tecnologías de la Información y la Comunicación (TIC)” [18], donde se especifica un conjunto de competencias muy amplio y completo para los estudiantes de primaria y

secundaria. Además las competencias definidas por CPE para los docentes, ya que se espera que se hayan transmitido las actitudes, aptitudes, habilidades y desempeños de ellos hacia los estudiantes, en particular, las Competencias de Colaboración y comunicación, y Técnicas y tecnológicas; complementándolas con las planteadas por el ICEC [18], obteniendo así un compendio de competencias y desempeños adecuadamente definidos para los estudiantes a encuestar (estudiantes de primaria básica y secundaria básica).

Una vez identificadas las variables a medir, se procede a definir las conceptualmente a cada una de ellas. En el **Anexo 3** se plantea la Definición de las Variables Docentes y en el **Anexo 4** se plantea la Definición de las Variables Estudiantes.

Definición de los objetivos

Teniendo en cuenta todos los aspectos mostrados en las actividades anteriores, especialmente la definición de las variables, se definieron los siguientes objetivos:

Objetivos Encuesta Docentes:

- Identificar los aspectos generales y socio-económicos de los docentes/directivos y docentes que participaron en las jornadas de formación de la Fase de Profundización realizadas durante el año 2007 al personal docente de las instituciones o centros educativos beneficiados por el programa Computadores para Educar.
- Identificar las habilidades y desempeños en el uso de las TIC de los docentes/directivos docentes que participaron en las jornadas de formación de la Fase de Profundización del año 2007, las cuales se brindaron al personal docente de las instituciones educativas beneficiadas por el CPE.

Objetivos Encuesta Estudiantes:

- Recopilar información sobre las habilidades, conocimientos y desempeños en el uso de las TIC, de estudiantes de los docentes que participaron en las jornadas de formación de la Fase de Profundización del año 2007, las cuales se brindaron al personal de las instituciones educativas beneficiadas por el CPE.

Definición de las hipótesis

En esta actividad se establece una relación a priori entre la población que se va a estudiar y las variables definidas. Teniendo en cuenta las variables definidas para los docentes, se definieron las siguientes dos hipótesis:

H1: “Los docentes que asistieron a las Jornadas de formación de la Fase de Profundización del año 2007, encuentran que es muy importante la integración de las TIC se fortalecen los procesos de enseñanza - aprendizaje”.

H2: “Los docentes que asistieron a las Jornadas de formación de de la Fase de Profundización del año 2007, están integrando las TIC en sus actividades y proyectos pedagógicos”.

Teniendo en cuenta las variables definidas para los estudiantes, se definieron las siguientes dos hipótesis:

H1: “Los estudiantes de los docentes que asistieron a las jornadas de formación de la Fase de Profundización del año 2007, tienen conocimiento básico del uso de las TIC.”

2. Elaboración del cuestionario

Elaboración de los indicadores

En esta actividad se elaboran los indicadores de las variables que en sí mismas no son directamente medibles. Los indicadores son hechos que ponen de manifiesto las distintas dimensiones que puede tener una variable.

Dado que las variables principales a medir en este estudio son las competencias definidas para los docentes y para los estudiantes, se tomaron los desempeños como las dimensiones y sub-dimensiones que puede tomar una variable. Se elaboraron los indicadores o hechos que ponen de manifiesto las distintas dimensiones que puede tener cada una de las variables. En el **Anexo 3** y **Anexo 4** se definen cada una de las dimensiones y sub-dimensiones de las variables con sus respectivos indicadores.

Diseño del cuestionario

El cuestionario es un conjunto de preguntas que tiene como finalidad la obtención de los datos necesarios para una investigación. Para su elaboración se parte de las variables y los indicadores establecidos en fases anteriores, construyendo un conjunto de preguntas para cada variable.

Docentes

Teniendo como referencia las competencias y desempeños definidos por CPE, se procedió a definir un primer conjunto de preguntas, de forma que las opciones de respuestas estuvieran estandarizadas o cerradas, conservando un número mínimo de preguntas abiertas.

Además se solicitó la asesoría de un experto en pedagogía (Doctor Miguel Corchuelo, quien tiene Doctorado en Educación y es docente de la Universidad del Cauca), quien realizó las siguientes recomendaciones a tener en cuenta en la elaboración de la encuesta:

- Incluir más preguntas abiertas para la prueba piloto para no condicionar la respuesta del encuestado.
- Estandarizar las preguntas con la información recogida en la prueba piloto, para tenerla en cuenta en la encuesta definitiva.
- Dar espacio en las respuestas del cuestionario, para recoger evidencia sobre las respuestas que daban los docentes en la encuesta.
- Modificar parte del vocabulario que se estaba utilizando en la redacción de las preguntas, de tal manera que fuera menos técnico, más claro y de mayor dominio para los encuestados.
- Modificar aspectos de forma y organización de la encuesta.

Después de realizar todas las recomendaciones del experto, se generó una segunda versión del cuestionario y se procedió a crear el formato de la encuesta, organizando las preguntas de acuerdo a la correlación de ellas con las variables definidas. Por lo tanto el cuestionario quedo compuesto por dos partes: la parte A que comprende las preguntas del modulo 1 y la parte B que comprende las preguntas de los módulos 2, 3, y 4. Esta organización se describe a continuación:

- **Parte A:**
 - **Modulo 1:** Información general y socio-económica del docente. Tiene como objetivo caracterizar información educativa y socio-económica del docente.
- **Parte B:**
 - **Modulo 2:** Competencias Pedagógicas. Tiene como objetivo obtener información relacionada a las habilidades y desempeños Pedagógicos desarrollados por el docente.
 - **Modulo 3:** Competencias Comunicativas y Colaborativas. El objetivo es obtener información relacionada a las habilidades y desempeños Comunicativos y Colaborativos desarrollados por el docente.
 - **Modulo 4:** Competencias Técnicas y Tecnológicas. Tiene como objetivo obtener información relacionada a las habilidades y desempeños Técnicos y Tecnológicos desarrollados por el docente.

En el **Anexo 5** Formulario para Docentes/Directivos Docentes -Parte A y el **Anexo 6** Formulario para Docentes/Directivos Docentes -Parte B, se detallan los objetivos y el conjunto de preguntas para la encuesta.

Estudiantes

Con las competencias y desempeños definidos por el grupo de trabajo del proyecto, se procedió a elaborar una primera versión del cuestionario (con preguntas cerradas y un mínimo de preguntas abiertas) que se realizaría a los estudiantes de las Instituciones educativas a las cuales estaban vinculados los docentes que participaron en las Jornadas de formación del año 2007. Con los cuestionarios de primaria y secundaria básica totalmente refinados se procedió a crear los formatos de las encuestas, organizando las preguntas de acuerdo a la correlación de ellas con las variables definidas.

Ambos cuestionarios quedaron compuestos por dos partes: una parte A que comprende las preguntas del modulo 1 y la parte B que comprende las preguntas de los módulos 2 y 3. Esta organización se describe a continuación:

- **Parte A:**
 - **Modulo 1:** Información general, académica y familiar del estudiante. Tiene como objetivo caracterizar información educativa del estudiante.
- **Parte B:**
 - **Modulo 2:** Competencias Comunicativas y Colaborativas. Tiene como objetivo obtener información relacionada a las habilidades y desempeños Comunicativos y Colaborativos desarrollados por el estudiante.
 - **Modulo 3:** Competencias Técnicas y Tecnológicas. Tiene como objetivo obtener información relacionada a las habilidades y desempeños Técnicos y Tecnológicos desarrollados por el estudiante.

En el **Anexo 7** Formulario para Estudiantes de Primaria Básica -Parte A, **Anexo 8** Formulario para Estudiantes de Primaria Básica -Parte B, **Anexo 9** Formulario para Estudiantes de Secundaria Básica -Parte A y el **Anexo 10** Formulario para Estudiantes de Secundaria Básica -Parte B, se detallan los objetivos y el conjunto de preguntas para cada una de las encuestas.

En el **Anexo 11** se encuentra el formulario desarrollado para la recolección de información general y socio-económica para los formadores, y en el **Anexo 12** se encuentra el formulario desarrollado para recolectar información sobre la infraestructura tecnológica de las Entidades Educativas.

Determinación de la población

Definición de la población

En esta actividad se define las características o rasgos físicos, socio-económicas y/o geográficos que delimitan la población de objeto de estudio.

Docentes

La población de estudio son los docentes o directivos docentes que participaron en las jornadas de formación de la fase de profundización que se desarrolló en las sedes-instituciones educativas de la región Sur Pacífico 2 (Cauca, Nariño y Putumayo) beneficiadas por el programa CPE en el año 2007.

Estudiantes

La población de estudio son los estudiantes que han recibido clases de los docentes mencionados previamente.

Determinación del tamaño de la población

En esta actividad se debe determinar si la población es finita o infinita, en el caso de ser finita se debe precisar el número de habitantes de la población de objeto de estudio.

Docentes

En el caso de los docentes la población es finita, después de un proceso dispendioso de recolección de información se determinó que aproximadamente son 745 docentes o directivos docentes distribuidos no uniformemente en 137 sedes educativas de la región Sur Pacífico 2 (Cauca, Nariño y Putumayo). En la siguiente tabla se muestra la distribución de la población por ruta, la convención utilizada es la siguiente:

- Nh: No. de docente del estrato h
- Uh: No. de sedes educativas del estrato h.

Tabla 1. Rutas de la región Sur-Pacífico definidas por CPE-UNICAUCA

Ruta	Nh	Uh	Depto.	Municipios
R1	147	17	Cauca	Tambo, Inzá, Puracé, Sotará, Timbío y Totoró
R2	101	17	Cauca	Almaguer, Argelia, Bolívar, La Vega, Patía, San Sebastián y Sucre
R3	102	17	Cauca	Santa Rosa
R4	57	17	Nariño	Alban, Colon, El Tablón, La Unión, San Bernardo, San Lorenzo, San Pablo, San Pedro de Cartago
R5	29	14	Nariño	Ancuya, El Peñol, El Rosario, La Florida, La Llanada, Leiva, Los Andes, Samaniego, Sandoná, Santacruz
R6	78	17	Nariño	Consacá, Contadero, Funes, Imues, Túquerres y Yacuanquer
R7	56	17	Nariño	Barbacoas, Cumbal, Ricaurte y Tumaco
R8	175	21	Putumayo	Mocoa, Orito, Puerto Asís, Puerto Caicedo, Valle del Guamuez y Villagarzón
8	745	137		

Estudiantes

En el caso de los estudiantes la población es finita, después de un proceso dispendioso de recolección de información se determinó que aproximadamente son 10.292 estudiantes distribuidos no uniformemente en 137 sedes educativas de la región Sur Pacífico 2 (Cauca, Nariño y Putumayo). En la

Tabla 2. Rutas de la región Sur-Pacífico definidas por CPE-UNICAUCA se muestra la distribución de la población por ruta, la convención utilizada es la siguiente:

- Neh: No. de estudiantes del estrato h
- Uh: No. de sedes educativas del estrato h.

Tabla 2. Rutas de la región Sur-Pacífico definidas por CPE-UNICAUCA

Ruta	Neh	Uh	Depto.	Municipios
R1	1131	17	Cauca	Tambo, Inzá, Puracé, Sotará, Timbío y Totoró
R2	1600	17	Cauca	Almaguer, Argelia, Bolívar, La Vega, Patía, San Sebastián y Sucre
R3	964	17	Cauca	Santa Rosa
R4	966	17	Nariño	Alban, Colon, El Tablón, La Unión, San Bernardo, San Lorenzo, San Pablo, San Pedro de Cartago
R5	1541	14	Nariño	Ancuya, El Peñol, El Rosario, La Florida, La Llanada, Leiva, Los Andes, Samaniego, Sandoná, Santacruz
R6	900	17	Nariño	Consacá, Contadero, Funes, Imues, Túquerres y Yacuanquer
R7	1634	17	Nariño	Barbacoas, Cumbal, Ricaurte y Tumaco
R8	1556	21	Putumayo	Mocoa, Orito, Puerto Asís, Puerto Caicedo, Valle del Guamuez y Villagarzón
8	10292	137		

La

Tabla 2. Rutas de la región Sur-Pacífico definidas por CPE-UNICAUCA hace referencia a los estudiantes por institución educativa y por zona, para el proyecto solo se encuestaron

a estudiantes de los últimos grados (cuarto y quinto para primaria, octavo a once para secundaria), debido a que el cuestionario se diseñó para estudiantes que estaban finalizando el proceso de aprendizaje y por lo tanto era posible que tuvieran un mayor conocimiento de los conceptos básicos de las TIC, en primaria o secundaria según sea el caso, y con una restricción adicional, que hayan recibido clase de los docentes que participaron en las Jornadas de Formación en el año 2007, de los cuales no se tiene ningún dato para el proceso de muestreo.

Diseño de la muestra

Una vez se conoce la población objeto del estudio, puede ocurrir que ésta sea tan grande que implique unos costos económicos y de energía muy elevados. Ello hará necesario que se seleccione un subconjunto de dicha población que reduzca los costos y permita generalizar los resultados obtenidos. A este subconjunto significativo de la población se le denomina muestra. En el **Anexo 13** se puede observar el proceso de diseño de la muestra detalladamente.

Prueba Piloto

El objetivo de esta fase es comprobar que el cuestionario funciona. Se selecciona un grupo reducido de personas y se les entrevista con el cuestionario inicial. Los objetivos de esta entrevista deben ser: ver qué le falta o qué le sobra al cuestionario (es decir, si el cuestionario es realista); comprobar la redacción de los enunciados (si el encuestado entiende claramente lo que se le pregunta); y comprobar si las respuestas múltiples se adecuan a lo que el encuestado responde de una forma espontánea.

Para la prueba de los cuestionarios, se seleccionaron Sedes Educativas teniendo en cuenta lo siguiente:

- Que fueran Instituciones cercanas a Popayán, para que se realizara de manera ágil y con un costo reducido.
- Que no estuvieran dentro de las Sedes Educativas que pertenecen a la muestra definida.

Aparte del cuestionario se les entregó una hoja adicional donde podían anotar las recomendaciones o comentarios sobre los inconvenientes que se les presentara durante el diligenciamiento del cuestionario. Los docentes y estudiantes involucrados, eran conscientes de que era una prueba y debían tener una posición crítica frente a la estructura y contenido de la cuestionario.

Docentes

La prueba piloto para el cuestionario para docentes, se realizó en las siguientes Instituciones Educativas:

- Timbío: Centro Educativo Barro Blanco Sede Centro Docente Las Guacas (3 Docentes).
- Coconúco: Institución Educativa Guillermo León Valencia (18 Docentes)

Teniendo en cuenta las recomendaciones y comentarios realizados en la prueba piloto, se realizaron algunas modificaciones en el cuestionario para evitar los mismos inconvenientes y tener nuestro cuestionario definitivo para aplicarla a la muestra de Sedes Educativas definida. Estas modificaciones fueron a nivel de estructuración del contenido del cuestionario (hacer el diligenciamiento más fácil y ágil) y en algunas preguntas fue

necesario cambiar la redacción (el encuestado se mostraba confundido con el planteamiento de algunas preguntas).

Estudiantes

La prueba piloto para el cuestionario de estudiantes, se realizó en las siguientes instituciones educativas:

- Timbío: Centro Educativo Barro Blanco Sede Centro Docente Las Guacas, 20 estudiantes de los grados tercero, cuarto y quinto (Primaria básica).
- Coconúco: Institución Educativa Guillermo León Valencia, 20 estudiantes de los grados octavo, noveno, decimo y once (Secundaria básica).

Se presentaron inconvenientes en cuanto a forma y presentación de algunas preguntas, ya que resultaban confusas y no las comprendieron. Por lo tanto se realizaron las respectivas correcciones de forma y de fondo sobre las preguntas.

3. Aplicación del cuestionario

En esta fase, el investigador decide si aplicará personalmente los cuestionarios o a través de encuestadores. Esta decisión está condicionada, en gran parte, por el tamaño de la muestra. Si ésta es pequeña, lo mejor es que la realice el propio investigador en aras de una mayor efectividad. Cuando la muestra es grande (cientos o miles de personas) es conveniente realizarla con encuestadores.

Los encuestadores encargados fueron los autores de este proyecto, cada uno fue responsable de encuestar la mitad de las Sedes Educativas de la muestra definida (37 Sedes Educativas). No se contrataron encuestadores debido a que era muy importante que el encuestador tuviera un conocimiento muy profundo del proyecto y del objetivo de la encuesta al momento de resolver dudas a los encuestados.

En primera instancia se contactó a los rectores de 24 Sedes educativas, informándole sobre la visita a realizar para el diligenciamiento del cuestionario y para establecer la fecha en la cual se realizaría la visita y el horario de reunión con los docentes de la Sede Educativa certificados para el diligenciamiento (aproximadamente una hora), de esta manera se estableció el cronograma de visitas. La programación de las visitas a las entidades educativas se presenta en el **Anexo 14**.

En el momento de aplicar las encuestas en las Instituciones Educativas se presentaron algunos inconvenientes externos, como los siguientes:

- En algunas Sedes Educativas se encontraron varios computadores fuera de servicio o en espera de recibirlos por Reposición, lo que limitó de alguna forma que los estudiantes, docentes y directivos pudieran desarrollar adecuadamente sus competencias, razón por la cual se rehusaron a contestar la encuesta.

Por causa de disturbios sociales y por ello problemas con el servicio de transporte, no fue posible realizar la totalidad de las visitas en el mes de noviembre, en particular a las Sedes ubicadas en los departamentos de Nariño y Putumayo, lo cual impactó en un retraso en la ejecución de las visitas a las instituciones educativas hasta el mes de abril del año 2009, debido a la disposición de los recursos para viáticos.

- Se tuvieron contratiempos en el cumplimiento del cronograma planteado de la segunda visita, tales como la aplicación de las pruebas “Saber” para estudiantes, entrega de boletines de notas de los estudiantes y varias festividades (Día de la madre, del niño y del maestro).

Para futuras aplicaciones de encuestas dentro del mismo contexto, se recomienda:

- Aplicar el módulo Información general y Socio-económica, al final o durante el transcurso de las Jornadas de Formación, dado que esta información está disponible en aquel momento y no resulta dispendiosa su recolección.
- Aplicar algunas de las preguntas del módulo de las Competencias Técnicas y Tecnológicas, al finalizar las Jornadas de Formación, debido a que parte de los desempeños y habilidades técnicas y tecnológicas, se refieren al conocimiento teórico adquirido recientemente durante las jornadas.
- Aplicar los módulos de las Competencias Pedagógicas, y Comunicativas y Colaborativas al año siguiente de haber finalizado las Jornadas de Formación, ya que luego de este periodo de tiempo los docentes ya han desarrollado y aplicado el conocimiento práctico adquirido en las jornadas de formación.

4. Tratamiento Estadístico de los Resultados

Los resultados de las fases de Tratamiento Estadístico de los Resultados y de la fase de Elaboración del Informe, se presentan el **Anexo 15**, los cuales hacen parte de los informes técnicos presentados a CPE Bogotá, junto con los resultados de todos los pasos anteriores. A continuación se presentan algunos de los resultados estadísticos obtenidos:

En la siguiente figura se muestra la distribución de la población docente que asistió a un porcentaje de las jornadas de formación de la FP, de la cual se puede concluir que, la mitad de la población asistió a la mayoría (más o igual de 80%) de las jornadas de formación.

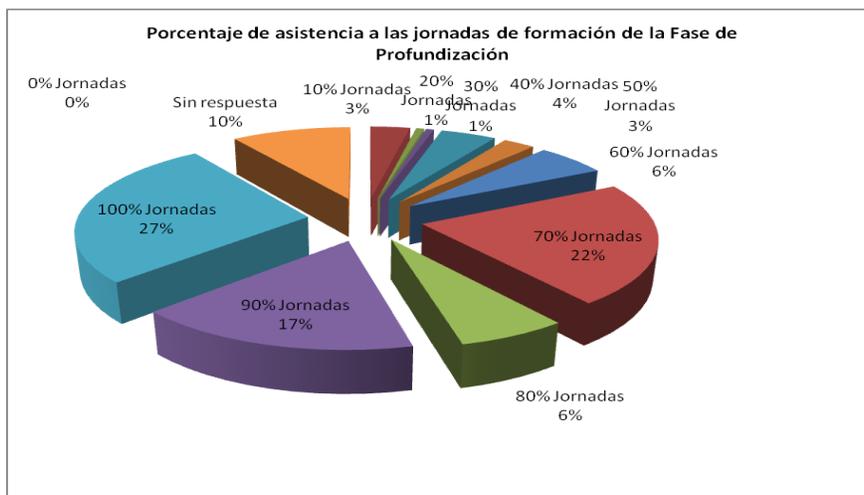


Figura 5. Porcentaje de asistencia del docente a las jornadas de formación de la Fase de Profundización

En la siguiente figura se muestra la distribución de la población docente por su formación académica, de la cual se puede concluir que, casi la mitad (con un 47%) de la población esta licenciado en educación, y que la mayoría (con un 81%) de la población como mínimo ha realizado un postgrado.

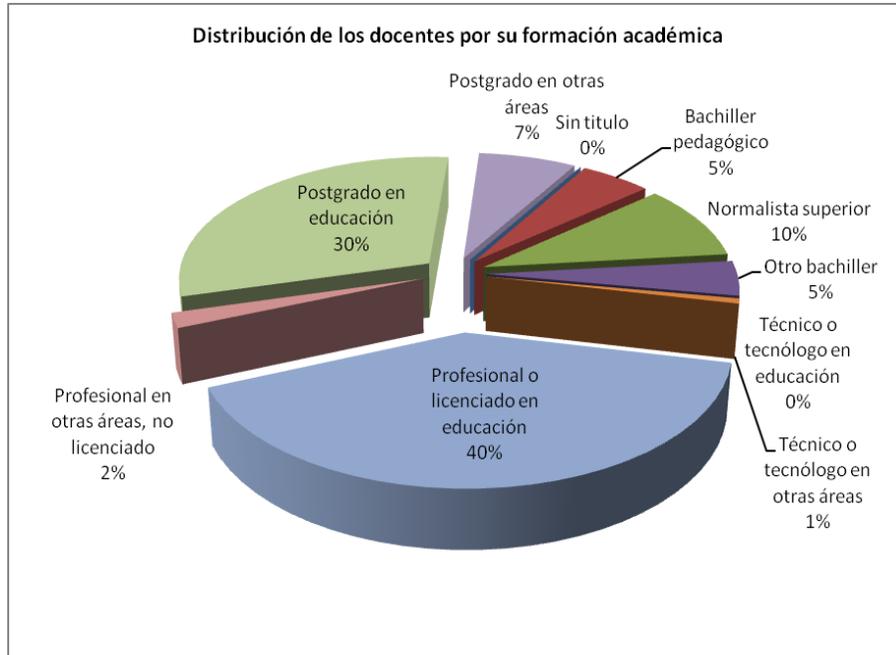


Figura 6. Distribución de la población docente por su formación académica

En la siguiente figura se muestra la distribución de la población docente que ha identificado problemas en el aula o sede educativa que pueden ser resueltos utilizando las TIC, de la cual se puede concluir que, casi los tres cuartos de la población lo hacen.

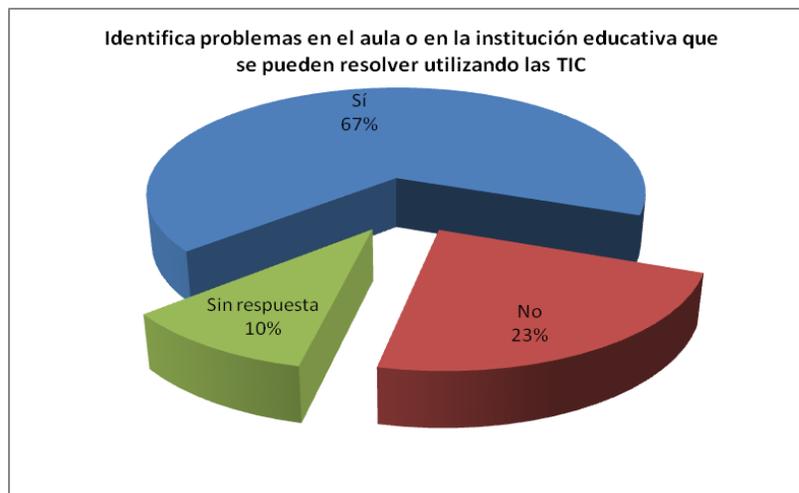


Figura 7. Docente que ha identificado problemas en el aula o en la institución educativa y se pueden resolver utilizando las TIC

3.2. Herramienta de Recolección de Datos

Se construyó un prototipo de herramienta de recolección de datos (OLTP), llamado “Sistema de Información de la Fase de Profundización (SIFP)”. Este sistema se implementó en Microsoft Office Access 2007, y se desarrolló especialmente para almacenar información relacionada con la Fase de Profundización, además de la información recolectada durante las visitas realizadas a las diferentes instituciones educativas de la muestra. A continuación se detalla la información que se registra en este sistema:

- Información general del docente/directivo docente como: información de identificación, demográfica, localización, socio-económica, formación académica y profesional, laboral, formación en informática. Dicha información se obtiene de los formatos *FM-A Docente-Directivo Docente* (Ver **Anexo 5**) aplicados a los docentes de la muestra.
- Información sobre las competencias (habilidades, conocimientos, aptitudes, actitudes y valores) en el uso de las TIC de los docentes/directivos docentes a través de las encuestas. Esta información se obtiene de los formatos *FM-B Docente-Directivo Docente* (Ver **Anexo 6**) aplicados a los docentes de la muestra.
- Información general del estudiante de primaria como: información de identificación, demográfica, localización, socio-económica, formación académica, núcleo familiar. Dicha información se obtiene de los formatos *FM-A Estudiante Primaria Básica* (Ver **Anexo 7**) aplicados a los estudiantes de la muestra.
- Información sobre las competencias (habilidades, conocimientos, aptitudes, actitudes y valores) en el uso de las TIC de los estudiantes de primaria a través de las encuestas. Esta información se obtiene de los formatos *FM-B Estudiante Primaria Básica* (Ver **Anexo 8**) aplicados a los estudiantes de la muestra.
- Información general del estudiante de secundaria como: información de identificación, demográfica, localización, socio-económica, formación académica, núcleo familiar. Dicha información se obtiene de los formatos *FM-A Estudiante Secundaria Básica* (Ver **Anexo 9**) aplicados a los estudiantes de la muestra.
- Información sobre las competencias (habilidades, conocimientos, aptitudes, actitudes y valores) en el uso de las TIC de los estudiantes de secundaria a través de las encuestas. Esta información se obtiene de los formatos *FM-B Estudiante Primaria Secundaria* (Ver **Anexo 10**) aplicados a los estudiantes de la muestra.
- Información general del formador como: información de identificación, demográfica, localización, socio-económica, formación académica, laboral. Dicha información se obtiene de los formatos *FM-A Formador* (Ver **Anexo 11**) aplicados a los formadores.
- Información de infraestructura tecnológica de las sedes educativas. Esta información se obtiene de los formatos *FM-B Entidad Educativa* (Ver **Anexo 12**) aplicados a las sedes educativas de la muestra.
- Información del modelo de competencias de la Fase de Profundización. Información obtenida del documento Términos de Referencia Fase de Profundización 2008.
- Información sobre las jornadas de formación; contenidos o temas, tiempos, rutas, formadores, entre otros. Información obtenida de diferentes archivos digitales semi-estructurados manejados por CPE-UNICAUCA.
- Información sobre los departamentos, municipios, zonas, niveles educativos, grados de enseñanza, áreas de enseñanza, especialidades del sistema colombiano. Información obtenida de los diferentes formatos manejados por el DANE. De dicha

información se obtuvo su estructura e información de los datos (códigos, nombres, etc.).

- Información sobre las zonas, regiones, convenios manejados por CPE. Información obtenida de diferentes documentos digitales no estructurados del portal Web de CPE (<http://www.computadoresparaeducar.com>).
- Información de configuración y relación entre las diferentes entidades de la base de datos.

El SIFP se construyó en Microsoft Office Access porque es portable, liviano y provee un mecanismo fácil y rápido para crear las estructuras y los componentes de gestión de los datos, como las tablas y los formularios. En el **Anexo 16** se muestran las figuras de un formulario ejemplo para el registro de los datos de los docentes, los formularios base para la gestión de datos en el Sistema de Información de la Fase de Profundización y el modelo relacional del SIFP.

CAPITULO IV. DESCRIPCIÓN DEL PROCESO DE DESARROLLO DEL DSS

A continuación se presentan las etapas que conforman el ciclo de vida dimensional, realizando una breve descripción de su objetivo general, seguido de los respectivos resultados obtenidos en cada una de ellas.

4.1.1. PLANEACIÓN DEL PROYECTO

El objetivo de esta etapa fue identificar el escenario del proyecto para saber de dónde surgió la necesidad del DW y poder de esta manera construir el alcance del proyecto, se incluye el impacto y la evaluación de factibilidad. Esta etapa depende de los requerimientos, ya que a partir de estos se determina el alcance del proyecto y los recursos requeridos. Para llevar a cabo esta tarea de alto nivel fue necesario tener las especificaciones de las actividades y las tareas necesarias, para luego definir el alcance y la justificación del proyecto [7].

DEFINICIÓN DEL PROYECTO

Para el desarrollo del DSS en el marco de CPE en la FP, se identifiqué que por parte de los directivos del convenio entre CPE y la Universidad del Cauca, se tiene un gran interés por poder realizar análisis en cuanto a las competencias TIC adquiridas por los beneficiarios del programa, además de poder relacionar información del tipo socio económico de los mismos, ruta, formador, recursos disponibles, entre otras.

Se debe tener en cuenta que para el desarrollo del proyecto se tuvo que realizar un estudio detallado del entorno de aplicación, ya que la educación (en particular el tema de competencias TIC), no es el entorno tradicional en el que se enfocan este tipo de sistemas (los cuales están dirigidos a necesidades empresariales) y para su aplicación al entorno de este proyecto se tuvieron que realizar ajustes conceptuales (Sucursales -> Entidades Educativas, Clientes -> Beneficiarios, Negocio -> Convenio CPE-Unicauca, entre otros), terminológicos y de implementación (Ganancias en pesos y cantidades -> Indicadores, donde estos no se comportan como simples agregaciones, este fue un gran ajuste de implementación, el cual se detalla en el CAPITULO V. CASOS DE DISEÑO DE DW PARA EL MODELADO DIMENSIONAL DE HABILIDADES Y DESEMPEÑOS EN EL USO DE LAS TIC).

Evaluación de la disponibilidad organizacional para la ejecución del proyecto

Dentro de las tareas propuestas por Kimball [7], se presentan algunos factores que permiten evaluar la disposición que tiene una organización para el desarrollo de un DW, para así tener una mayor probabilidad de éxito del proyecto DW; pero teniendo en cuenta que algunos de estos factores no pueden ser aplicados completamente según el entorno de desarrollo del proyecto, sin embargo se analizaron y se adaptaron para su aplicación.

Para realizar esta actividad se una prueba de Disponibilidad organizacional, la cual se encuentra detallada en el **Anexo 17**. A continuación se muestran las conclusiones obtenidas después realizar el análisis con la prueba mencionada.

Influencia del Sponsor en la organización: El DSS se desarrolla para dar soporte a la toma de decisiones a los directivos del convenio CPE–Universidad del Cauca, en cuanto a las estrategias de acompañamiento educativo, análisis de las competencias TIC adquiridas por los beneficiarios teniendo en cuenta información relacionada con la Fase de profundización (formador, ruta, institución, entre otras). El Ingeniero Jair Moreno, es el director del Convenio 2008, en el cual se encuentra enmarcado este proyecto de investigación, y tiene el rol Sponsor dentro del proyecto, ya que es el directivo más interesado dentro del convenio en utilizar este tipo de soluciones.

Como el Sponsor es el mismo director del Convenio CPE–Universidad del Cauca, es la persona que autoriza los gastos y recursos que tienen que ver con el desarrollo del DSS. Además como permanece en la ciudad de Popayán, es fácil para el equipo de trabajo del DSS acceder a él en busca de soluciones.

Motivación del proyecto: Hasta el momento el convenio CPE–Universidad del Cauca no cuenta con un sistema que les permite realizar seguimiento de las jornadas de formación realizadas por los formadores, ni de las competencias TIC adquiridas por los beneficiarios formados, tampoco un sistema que les brinde un soporte en el establecimiento de estrategias de acompañamientos educativo.

Por lo tanto el DSS es una alternativa de solución, en la cual se encuentran interesados el grupo de trabajo del Convenio y además el programa de CPE con sede en la ciudad de Bogotá.

Relación entre el equipo del convenio (Directivos – Formadores) y los del Sistema de Información (Área de tecnologías de la información): La relación entre el grupo de trabajo del convenio CPE–Universidad del Cauca y la persona encargada del registro en el Sistema de Monitoreo y Control CPE (SIMEC) es muy buena y existe mucha cordialidad. Por lo tanto la persona encargada de SIMEC brindará la información de cada actividad realizada por los formadores sin ningún inconveniente.

Presencia de cultura analítica: El grupo de trabajo del convenio CPE–Universidad del Cauca, es un grupo comprometido a cumplir las metas establecidas por el convenio, en particular para la Fase de Profundización se realiza un seguimiento de: cantidad de beneficiarios formados, cantidad de recursos donados, cantidad de instituciones beneficiadas; cifras manejas por formador encargado de una ruta, a través del diligenciamiento de los formatos que son registrados en el SIMEC. Esto evidencia que para el grupo de trabajo del convenio CPE–Universidad del Cauca es muy importante registrar los datos y analizarlos.

Factibilidad: Como este proyecto se encuentra enmarcado dentro del convenio CPE–Universidad del Cauca, se cuenta con los recursos tanto de parte de la universidad como del programa CPE, además se cuenta con el acceso de los datos pertenecientes a este convenio.

Para concluir se observa una alta disponibilidad para poder desarrollar satisfactoriamente el proyecto, teniendo en cuenta que los factores evaluados con el test de Litmus dieron como resultado una disponibilidad alta equivalente a un porcentaje mayor al 80%.

PERSONAS RELACIONADAS CON EL PROYECTO

Directivos región sur pacifico 2: Son las personas que colaboraran al entendimiento de la información proporcionada para el desarrollo del DSS, como también proporcionar los recursos necesarios para poder ejecutar las actividades programadas en el proyecto.

Sponsor: Ingeniero Jair Moreno quien hace parte de los directivos del convenio CPE – Universidad del Cauca, es el vínculo entre el área de aplicación del proyecto y el equipo de desarrollo.

Gerente del proyecto: Magister Martha Eliana Mendoza es la directora del proyecto, y es la persona experta en el tema de las bodegas de datos, es quien asesorara y supervisara todas las actividades establecidas para el desarrollo del proyecto.

Analista del sistema: Deyro Enrique Zúñiga Sánchez y Jaime Alexander Mendoza Galviz, estudiantes en decimo semestre de Ingeniería de Sistemas.

Modeladores de datos: Martha Eliana Mendoza directora del proyecto, Deyro Enrique Zúñiga Sánchez y Jaime Alexander Mendoza Galviz, estudiantes en decimo semestre de Ingeniería de Sistemas.

Diseñadores ETL: Deyro Enrique Zúñiga Sánchez y Jaime Alexander Mendoza Galviz, estudiantes en decimo semestre de Ingeniería de Sistemas.

Programadores ETL: Deyro Enrique Zúñiga Sánchez y Jaime Alexander Mendoza Galviz, estudiantes en decimo semestre de Ingeniería de Sistemas.

Administradores de la bodega de datos: Deyro Enrique Zúñiga Sánchez y Jaime Alexander Mendoza Galviz, estudiantes en decimo semestre de Ingeniería de Sistemas.

Diseñadores del sistema temporal de almacenamiento de los datos: Deyro Enrique Zúñiga Sánchez y Jaime Alexander Mendoza Galviz, estudiantes en decimo semestre de Ingeniería de Sistemas.

Desarrolladores de las aplicaciones de usuario final: Deyro Enrique Zúñiga Sánchez y Jaime Alexander Mendoza Galviz, estudiantes en decimo semestre de Ingeniería de Sistemas.

DESARROLLO DEL PLAN DEL PROYECTO

En el **Anexo 20** se puede encontrar definido el plan detallado que conduce el desarrollo del DSS, donde se describen tiempos, actores y responsabilidades.

Alcance Preliminar

- **Respuesta Encuesta Beneficiario:** El objetivo de este Data Mart es producir información que permita realizar consultas analíticas relacionadas con las respuestas de las encuestas aplicadas a las personas beneficiadas por el programa CPE (docentes, directivos/docentes, estudiantes de primaria básica y secundaria básica).
- **Indicador Beneficiario:** El objetivo de este Data Mart es producir información relevante que permita satisfacer necesidades de consulta analíticas con relación a los modelos de competencias en TIC definidos para los beneficiarios del programa CPE
- **Acompañamiento Entidad Educativa:** La implementación de este Data Mart permitirá generar información relacionada con las fases; Inicial y de Profundización que se realizan en las entidades educativas beneficiadas por el programa CPE, información detallada de duración de las jornadas de formación, cantidad de recursos tecnológicos disponibles en la Entidad Educativa, entre otros.
- **Jornadas de Formación:** La implementación de este Data Mart permitirá generar información relacionada con las Jornada de Formación realizadas por los formadores para los beneficiarios del programa CPE, información sobre la duración de cada jornada de formación, cantidad de horas en los ejes pedagógico, de gestión y de infraestructura dictadas, entre otra.

PLAN DE COMUNICACIONES

Este plan busca establecer una comunicación adecuada y constante entre las personas relacionadas con el desarrollo del DSS. Si es necesario se definen los tipos de mensajes, contenido, formato y la frecuencia de los mensajes de comunicación entre los actores del proyecto.

Las estrategias establecidas para la comunicación entre las personas involucradas en el proyecto son:

Equipo del proyecto: Reuniones semanales entre el equipo y la directora del proyecto, para analizar y realizar el respectivo seguimiento del estado del proyecto, observar si las actividades se han realizado satisfactoriamente por los responsables, como también para asignar las nuevas actividades.

Formadores del convenio: Reuniones con los formadores del convenio CPE– Universidad del Cauca para obtener requerimientos generales del proyecto por parte de estos usuarios.

Directivos y formadores del convenio: Reuniones con los directivos y formadores del convenio CPE–Universidad del Cauca para presentar los avances del desarrollo del proyecto y poder retroalimentar el proceso llevado.

4.1.2. Definición de Requerimientos del Negocio

Los requerimientos de los usuarios tienen un gran impacto en las decisiones que se toman durante el proyecto, ya que los requerimientos son parte fundamental y el centro de las bodegas de datos.

La captura de requerimientos permitió establecer cuáles eran los datos que deberían estar disponibles en la bodega de datos, su organización, la frecuencia de actualización, quienes accederían a la bodega de datos, también el cómo se quería que se presentaran los datos; terminando con la planeación del desarrollo, mantenimiento y crecimiento de la bodega de datos teniendo en cuenta las opiniones del cliente. Los requerimientos del negocio guían al equipo de desarrollo para tomar las mejores opciones de estrategia, diseño y arquitectura técnica y establecen la base para continuar el ciclo de desarrollo sobre las tres rutas paralelas: Tecnología, Datos y Aplicaciones.

Para realizar la captura de requerimientos para el sistema se siguieron algunos pasos como: conocimiento detallado del contexto del convenio, entrevistas y reuniones con los responsables del grupo de trabajo del convenio CPE–Universidad del Cauca.

Por medio de las reuniones se identificaron varias necesidades analíticas que fueron agrupadas en los llamados “Temas analíticos”, los cuales son clasificados luego en procesos de negocio que darán inicio a la implementación de la bodega de datos, teniendo en cuenta su respectiva priorización y nivel de impacto en el negocio. A continuación se presentan los pasos realizados para la captura de los requerimientos del sistema:

Paso 1. Personal Administrativo y técnico del convenio: Se identificó el personal administrativo y técnico del grupo de trabajo del convenio CPE–Universidad del Cauca y el equipo de desarrollo del DSS. Los resultados se muestran en la Tabla 3. **Personal administrativo, Técnico y de Desarrollo** se presentan los participantes con sus respectivos roles.

Tabla 3. Personal administrativo, Técnico y de Desarrollo

Participantes	Rol en el modelado de procesos
Msc. Miguel Niño Msc. Jair Moreno	Ejecutivos y analistas del negocio
María de los Ángeles Anaya	Sistema de información - SIMEC

Paso 2. Entrevistas: Luego de haber identificado roles y participantes del convenio, se formularon las preguntas de los cuestionarios para realizar las entrevistas al personal administrativo, analista y técnico del negocio. Como se observa en la Tabla 3. **Personal administrativo, Técnico y de Desarrollo** el rol de administrativo y analista era ejecutado por las mismas dos personas. Los cuestionarios para las entrevistas desarrollados se encuentran en el **Anexo 18**.

Paso 3. Identificación de los Procesos del Negocio: Se realizó un resumen de los resultados obtenidos en el paso anterior para agrupar los requerimientos similares en temas analíticos comunes. Luego se identificaron los procesos de negocio que dan

soporte a cada uno de ellos. La Tabla 4 muestra el proceso de negocio Indicado Beneficiario junto con los temas analíticos asociados a éste proceso, los demás procesos de negocio con los temas analíticos asociados se encuentran en el **Anexo 19**.

Tabla 4. Temas Analíticos soportados por Indicador Beneficiario

Tema Analítico	Solicitud de Análisis	Proceso de negocio que lo soporta
Competencias Pedagógicas	<ul style="list-style-type: none"> • Saber en qué medida los docentes beneficiarios reconocen las potencialidades de las TIC en los procesos de enseñanza, aprendizaje. • Conocer en qué medida los docentes beneficiarios comprenden las oportunidades y retos que ofrece el uso de las TIC en los contextos educativos. • Conocer si se están aprovechando las potencialidades que ofrecen las TIC en el escenario pedagógico, de acuerdo con los recursos disponibles. 	Indicador Beneficiario
Competencias Comunicativas y Colaborativas	<ul style="list-style-type: none"> • Poder conocer si los docentes comparten, los aprendizajes y las construcciones de su proceso de integración de las TIC a la práctica pedagógica. • Conocer en qué medida los docentes interactúan con otros utilizando y compartiendo diversos recursos digitales, sincrónica y/o asincrónicamente. • Conocer en qué medida los docentes comprenden las implicaciones del trabajo colaborativo en redes y comunidades (presenciales y a distancia), haciendo uso de ellas en su vida personal y profesional. • Conocer en qué medida los docentes participan en redes o comunidades virtuales y/o presenciales (en roles periféricos ó centrales). • Analizar en qué medida los estudiantes beneficiarios interactúan con otras personas, sincrónica y/o asincrónicamente. 	Indicador Beneficiario
Competencias Técnicas y Tecnológicas	<ul style="list-style-type: none"> • Analizar si los docentes conocen y aplican procedimientos básicos para el funcionamiento adecuado de los computadores. • Analizar si los docentes configuran y mantienen las redes informáticas en el aula de cómputo. • Analizar en qué medida los docentes dominan conceptos y funciones básicas que permitan usar fluidamente las TIC, según su nivel de apropiación. 	Indicador Beneficiario

	<ul style="list-style-type: none"> • Analizar en qué medida los docentes seleccionan y utilizan aplicaciones fluidamente según su nivel de apropiación y producto requerido. • Conocer en qué medida los docentes buscan y utilizan información en Web. 	
	<ul style="list-style-type: none"> • Conocer en qué medida los estudiantes beneficiarios identifican los elementos básicos del ordenador y sus funciones. • Conocer en qué medida los estudiantes beneficiarios realizan un uso básico del sistema operativo. 	

Paso 4. Construcción de la Matriz Bus: En este paso se construye la matriz bus de la bodega de datos en la cual básicamente se hacen relaciones entre los DM (procesos del negocio) y las dimensiones de la bodega de datos, las cuales nos ayudarán más adelante a definir los hechos y los diagramas estrellas de la bodega de datos.

LISTAR LOS DATA MARTS DE LA BODEGA DE DATOS

En esta actividad se listan los DM de una bodega de datos que hacen referencia a los procesos principales del negocio. A continuación se muestra una breve descripción de los DM que se identificaron:

- **Respuesta Encuesta Beneficiario:** Almacena las respuestas de la respectiva encuesta realizada a los docentes y estudiantes beneficiarios por el programa CPE.
- **Indicador Beneficiario:** Almacena información relacionada con los modelos de competencias (habilidades, conocimientos, aptitudes, actitudes y valores) en el uso de las TIC que han adquirido los beneficiarios en las jornadas de formación, teniendo en cuenta los modelos establecidos para docentes y estudiantes.
- **Acompañamiento Entidad Educativa:** Almacena información relacionada con las fases; Inicial y de Profundización que se realizan en las instituciones educativas beneficiadas por el programa CPE, información detallada de duración de las jornadas de formación, cantidad de recursos tecnológicos donados, entre otros.
- **Jornadas de Formación:** Almacena la información relacionada con las jornadas de formación efectuadas por los formadores para los beneficiarios durante la Fase de Profundización, información tal como la cantidad de horas dictadas para el eje pedagógico, para el eje de infraestructura, el eje de gestión, asistencia de los beneficiarios, entre otra información.

LISTAR LAS DIMENSIONES DE LA BODEGA DE DATOS

En esta actividad se listan todas las dimensiones de una bodega de datos que hacen referencia a los objetos y sujetos principales que intervienen en los procesos del negocio. Las dimensiones que se identificaron fueron:

- **Área de Enseñanza:** Almacenan todas las áreas de conocimiento para la enseñanza de los diferentes niveles educativos que rigen en el sistema educativo colombiano.
- **Auditoría:** Se almacena el estado de todos los procesos relacionados con el cargue y actualización de los registros de la bodega de datos.

- **Beneficiario:** Almacena la información genérica de las personas beneficiadas por el programa CPE, con información detallada de identificación, demográfica, socio-económica, formación en informática, entre otras.
- **Convenio:** Se almacena la información de los convenios realizados entre CPE y las universidades, con la información de identificación y descriptiva de cada convenio.
- **Cuestionario:** Almacena la información general de todas las encuestas, con información detallada como: título, introducción, objetivo, versión, fecha de creación y autor del mismo.
- **Docente:** Se almacena información detallada de cada docente como: información laboral, formación profesional, grados y áreas de enseñanza que orienta, entre otras.
- **Empleado CPE:** Almacena la información de las personas que pertenecen al grupo de trabajo del convenio de Universidad del Cauca–CPE, quienes realizaron el proceso de formación y encuesta a las personas beneficiadas; con información detallada de identificación, demográfica, formación profesional, entre otras.
- **Entidad Educativa:** Se almacena la información de las sedes educativas que fueron beneficiadas con el programa CPE, con información detallada de cada sede como: la identificación, el nombre, la dirección, la localización, entre otros.
- **Estudiante:** Almacena información detallada de los estudiantes, adicional a la especificada en beneficiario, tal como su Información académica.
- **Fecha:** Se almacenan los días del calendario relacionados con las operaciones del sistema, con información detallada de cada fecha como: mes, día, año, día del año, entre otros.
- **Grado de Enseñanza:** Almacenan la información de los grados de de enseñanza que rigen en el sistema educativo colombiano.
- **Indicador:** Se almacena la información relacionada con los modelos de competencias en TIC establecidos para los beneficiarios por el programa (competencias, desempeños, indicadores y preguntas de las encuestas).
- **Localización:** Almacenan las localizaciones geográficas del territorio colombiano, con información detallada de cada localización como: el municipio, el departamento, la zona, entre otros.
- **Pregunta:** Se almacena información relacionada con las preguntas, que especifica si es una pregunta de única o múltiple respuesta.
- **Respuesta:** Almacena todas las categorías de las respuestas y de las opciones de respuesta para una pregunta, incluye información detallada como: etiqueta de la respuesta, su descripción, orden, peso, entre otros.
- **Ruta CPE:** Se almacena toda la información relacionada con las rutas definidas por CPE para ejecutar las Jornada de Formación y Acompañamiento educativo, información tal como: nombre de la ruta, la región CPE, la zona CPE, entre otras.
- **Universidad:** Almacenan datos relacionados con las universidades que ejecutaron la Fase de Profundización del programa CPE, incluyendo información de identificación y contacto de cada una de ellas.

MARCAR LAS INTERCEPCIONES

En esta actividad se marcan las intercepciones donde una dimensión existe para un proceso del negocio. Se debe verificar que la dimensión es importante y que debe estar cuidadosamente conformada a través de los procesos del negocio. A continuación en la

Tabla 5. **Matriz de Arquitectura Bus de la Bodega de Datos.** Se detallan las relaciones de los procesos del negocio con las dimensiones de la bodega de datos.

Tabla 5. Matriz de Arquitectura Bus de la Bodega de Datos.

	Fecha	Beneficiario	Entidad Educativa	Localización	Formador	Encuestador	Encuesta	Pregunta	Respuesta	Universidad	Convenio	Indicador	Auditoría	Área de Enseñanza	Grado de Enseñanza
Jornadas Formación	X	X	X	X	X					X	X		X	X	X
Acompañamiento Entidad Educativa	X	X	X	X	X					X	X		X		
Respuesta Encuesta Beneficiario	X	X	X	X	X	X	X	X	X	X	X		X	X	X
Indicador Beneficiario	X	X	X	X	X	X	X	X		X	X	X	X	X	X

Paso 5. Priorización de los Procesos del Negocio: Para priorizar los procesos del negocio se realizaron reuniones entre el equipo de desarrollo del DSS y el equipo del convenio CPE–Universidad del Cauca, teniendo en cuenta su impacto y viabilidad para el proyecto, además se pudo establecer si existía un correcto entendimiento de los requerimientos y de los procesos de negocio identificados. La Figura 8. **Impacto vs Viabilidad** muestra la priorización de los procesos de negocio realizada.

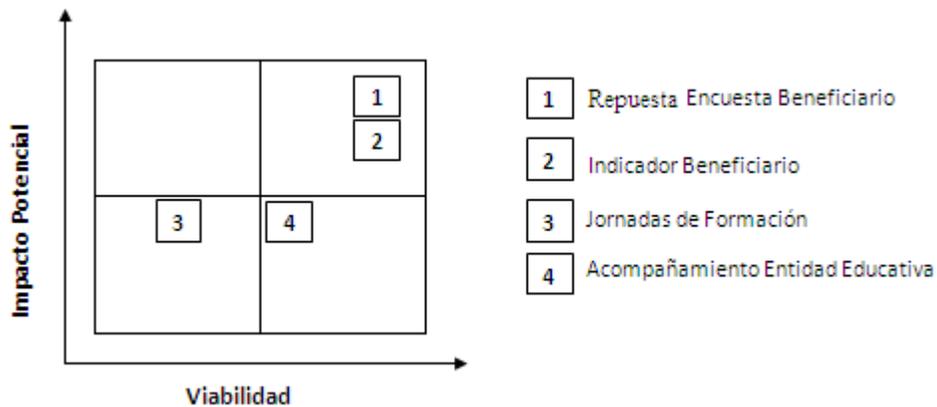


Figura 8. Impacto vs Viabilidad

Con la priorización de los procesos del negocio podemos observar lo siguiente:

1. **Respuesta Encuesta Beneficiario:** Tiene un alto impacto ya que permite gestionar la información recolectada por medio de las encuestas realizadas a los beneficiarios del programa CPE, para poder realizar análisis de la información socio-económica de los beneficiarios e incluyendo la información de infraestructura disponible en las entidades educativas, que puede influir en las competencias TIC adquiridas. Presenta una

viabilidad Alta, puesto que se obtendrá información detallada de los beneficiarios y de las entidades educativas gracias al instrumento de recolección de datos diseñado (Encuesta).

2. **Indicador Beneficiario:** Este proceso presenta un alto impacto pues gracias a este se podrán medir las competencias en TIC de los beneficiarios y se realizarán sus respectivos análisis por parte de los directivos y formadores del convenio CPE– Universidad del Cauca. Su viabilidad es alta porque se cuenta con un modelo de competencias en TIC definido por CPE para los docentes beneficiados por el programa y el modelo definido para los estudiantes beneficiarios, los cuales tienen relación directa con el instrumento de recolección de datos diseñado, para poder realizar el proceso de medición de las competencias de un beneficiado en particular.
3. **Jornadas de Formación:** Tiene un impacto medio, ya que este proceso involucra información relacionada con los temas tratadas en las diferentes Jornadas de Formación impartidas por los formadores a los beneficiarios, y también los recursos educativos utilizados para este fin. Estas temáticas tienen relación directa con el modelo de competencias en TIC definido por CPE. Su viabilidad es baja porque no se cuenta con información detallada sobre todo este proceso de formación.
4. **Acompañamiento Entidad Educativa:** Tienen un impacto medio, se maneja información cuantitativa sobre el estado de los recursos tecnológicos donados, disponibles en buen y en mal estado, etc. Su viabilidad es alta gracias a que se cuenta con información detallada por cada entidad educativa beneficiada por el programa.

Luego de realizar la priorización de los procesos de negocio, se realizan reuniones nuevamente con el fin de obtener un mayor entendimiento de los procesos de negocio a tener en cuenta para el desarrollo (Respuesta Encuesta Beneficiario, Indicador Beneficiario y Acompañamiento Entidad Educativa) los cuales serán el enfoque inicial del proyecto. El proceso de Negocio Jornadas de Formación presenta una baja viabilidad, debido a que no se posee la información representativa ni estructurada para poder llevar a cabo la implementación de este proceso de negocio en su totalidad, sin embargo se tendrá en cuenta durante las fases del desarrollo del proyecto.

Alcance del proyecto

A continuación se expone el alcance del proyecto, el cual es una refinación del alcance definido anteriormente, ya que se han priorizado y definido a un nivel más detallado todos los procesos del negocio junto con sus dimensiones respectivas.

- **Respuesta Encuesta Beneficiario:** El objetivo de este DM es producir información relevante que permita satisfacer necesidades de consulta analíticas con relación a las respuestas de las encuestas aplicadas a las personas beneficiadas por el programa CPE (docentes, directivos/docentes, estudiantes primaria básica y secundaria básica), permitiendo análisis de información desde múltiples perspectivas dimensionales (beneficiarios, las entidades educativas, formador, ruta, encuestador, pregunta, respuesta, entre otros). Este DM debe responder a preguntas como: ¿Qué cantidad de Docentes piensan que con la integración de las TIC se fortalecen los procesos de

enseñanza - aprendizaje?, ¿Qué cantidad de los docentes reflexiona con sus estudiantes sobre el aporte de las TIC en los procesos de enseñanza - aprendizaje?, ¿Qué cantidad de Entidades Educativas cuentan con servicio de internet?

- **Indicador Beneficiario:** El objetivo de este DM es producir información relevante que permita satisfacer necesidades de consulta analíticas con relación a los modelos de competencias en TIC definidos para los beneficiarios del programa CPE, permitiendo análisis de información desde múltiples perspectivas dimensionales (beneficiario, ruta, formador, encuesta, entidad educativa, entre otros). Este DM debe responder a preguntas como: ¿Qué calificación obtuvo un docente en particular, para las Competencias Pedagógicas?, ¿Cuál fue el promedio de calificación obtenido por los docentes de una institución en particular, para las Competencias Técnicas y Tecnológicas?, ¿Cuál es el promedio de la calificación obtenida en las Entidades Educativas que pertenecen a una ruta particular?, ¿Promedio de los resultados obtenidos por los beneficiarios para un indicador particular, de una entidad educativa ubicada en zona rural?.
- **Acompañamiento Entidad Educativa:** La implementación de este DM permite generar información relacionada con las fases; Inicial y de Profundización que se realizan en las entidades educativas beneficiadas por el programa CPE, información detallada de duración de las jornadas de formación, cantidad de recursos tecnológicos disponibles en la Entidad Educativa, entre otros. También debe permitir el análisis de información desde múltiples perspectivas dimensionales, que involucran obtener y cruzar atributos de convenio, fechas, ruta, formador y entidad educativa. Permitiendo responder a inquietudes con respecto al tiempo de duración de las Jornadas de Formación, promedio de: recursos educativos donados, beneficiarios formados y cantidad de jornadas de formación por institución educativa, entre otras.
- **Jornadas de Formación:** La implementación de este DM permite generar información relacionada con las Jornada de Formación realizadas por los formadores para los beneficiarios del programa CPE, información sobre la duración de cada jornada de formación, cantidad de horas en los ejes pedagógico, de gestión y de infraestructura dictadas, entre otra. También debe permitir el análisis de información desde múltiples perspectivas dimensionales (convenio, universidad, ruta CPE y formador). Las preguntas a responder están relacionadas con la cantidad de las horas dictadas en los diferentes ejes por formador, promedio de duración de las jornadas de formación, cantidad de jornadas de formación dictadas por entidad educativa, entre otras.

4.1.3. Modelado Dimensional

En esta fase se definieron los modelos dimensionales teniendo en cuenta los requerimientos obtenidos en la fase anterior.

Partiendo de la Matriz bus, se realizó el proceso de modelado dimensional (diagrama de flujo mostrado en la Figura 9. **Diagrama de Flujo del proceso de modelado dimensional (Adaptado de [8])**, que inicia con el desarrollo del modelo inicial de alto nivel junto con la lista de atributos, continuando con la fase de desarrollo del modelo dimensional detallado que es de donde se obtienen los campos para la tabla de hechos,

la descripción de los atributos y el diagrama de las tablas de las dimensiones, luego se continua con revisiones y validaciones del modelo dimensional hasta obtener el modelo dimensional definitivo.

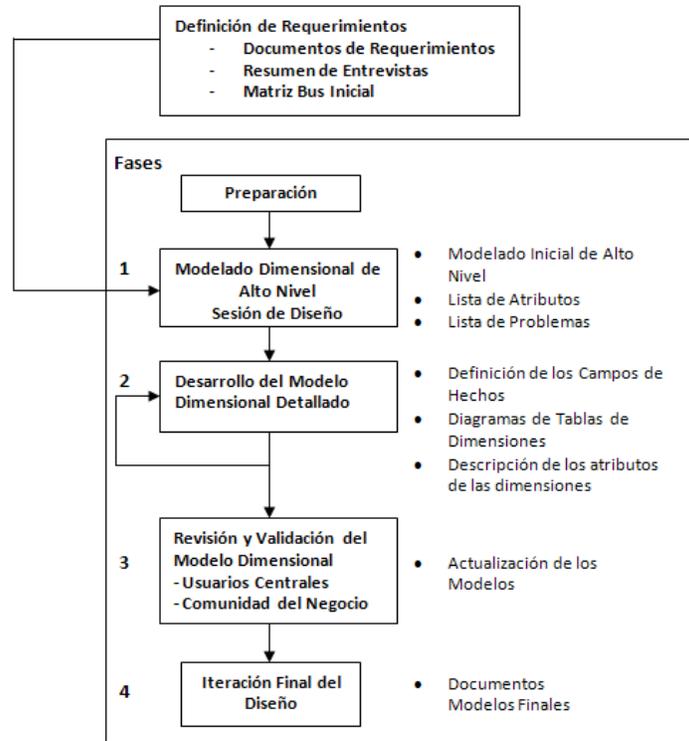


Figura 9. Diagrama de Flujo del proceso de modelado dimensional (Adaptado de [8])

MODELADO INICIAL DE ALTO NIVEL

Para desarrollar esta fase, los modelos iniciales de alto nivel para el DSS, se siguieron los cuatro pasos definidos por Kimball [7], realizados para cada uno de los procesos del negocio de la bodega de datos, se obtuvieron los siguientes resultados como se muestra a continuación en la Tabla 6. **Resultado del método de cuatro pasos de Ralph Kimball** (para ver en más detalle cada una los procesos del negocio revisar el **Anexo 21**).

Tabla 6. Resultado del método de cuatro pasos de Ralph Kimball

Proceso del Negocio	Dimensiones	Hechos
Acompañamiento Entidad Educativa	Convenio Fecha Inicio Fecha Final	Cantidad Computadores Donados Cantidad Impresoras Donados Cantidad Beneficiados
Granularidad: Línea de Elementos (cada registro en la tabla de hechos hace referencia a una acción en el tiempo)	Ruta CPE Formador Entidad Educativa	Formados Cantidad Jornadas Formación Cantidad Días Formación Cantidad Horas Formación
Repuesta Encuesta	Fecha	Texto Respuesta

<p>Beneficiario</p> <p>Granularidad: Línea de Elementos (cada registro en la tabla de hechos hace referencia a una respuesta de una pregunta)</p>	<p>Convenio Cuestionario Encuestador Ruta CPE Formador Entidad Educativa Beneficiario Número Encuesta (DD) Pregunta Respuesta</p>	<p>Valor Respuesta</p>
<p>Indicador Beneficiario</p> <p>Granularidad: Línea de Elementos (cada registro en la tabla de hechos hace referencia a una pregunta de un indicador)</p>	<p>Fecha Convenio Cuestionario Encuestador Ruta CPE Formador Entidad Educativa Beneficiario Número Encuesta (DD) Indicador</p>	<p>Valor Indicador</p>
<p>Jornada de Formación</p> <p>Granularidad: Transacción (cada registro en la tabla de hechos hace referencia a una acción en el tiempo)</p>	<p>Convenio Formador Fecha Inicio Fecha Final Ruta CPE</p>	<p>Cantidad de horas eje Pedagógico Cantidad de horas eje Gestión Cantidad de horas eje Infraestructura</p>

DESARROLLO DETALLADO DEL MODELO DIMENSIONAL

En esta fase se examina en detalle cada una de las tablas, profundizando en las descripciones, las fuentes, relaciones, tipos de datos, tipos de atributos SCD (atributos que cambian lentamente), problemas de calidad en los datos y transformaciones requeridas para poblar los modelos [8]. La descripción detallada de los atributos de las dimensiones y de las tablas de hechos de cada proceso de negocio del DSS se encuentra en el **Anexo 22**.

DIAGRAMAS DE LOS Data Mart

A continuación se muestran los diagramas de las tablas de hechos con sus respectivas dimensiones relacionadas. Para ver más detalladamente los atributos de las dimensiones y las medidas de la tabla de hechos de cada uno de los diagramas vea el **Anexo 23**.

- **Acompañamiento Entidad Educativa:** Almacena información relacionada con las fases; Inicial y de Profundización que se realizan en las instituciones educativas beneficiadas por el programa CPE, información detallada de duración de las jornadas de formación, cantidad de recursos tecnológicos donados, entre otros. El modelo de

alto nivel construido es mostrado en la Figura 10. **Diagrama Data Mart Acompañamiento Entidad Educativa.**

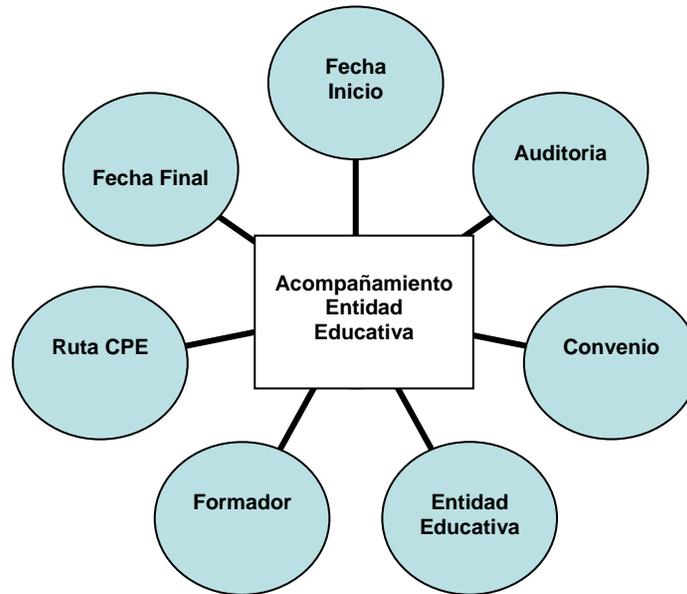


Figura 10. Diagrama Data Mart Acompañamiento Entidad Educativa

- **Respuesta Encuesta Beneficiario:** Almacena las respuestas de las encuestas aplicadas a las personas beneficiadas por el programa CPE (docentes, directivos/docentes, estudiantes primaria básica y secundaria básica), el cual permitirá analizar estas respuestas por las diferentes dimensiones que se relacionan con este hecho. El modelo de alto nivel construido es mostrado en la Figura 11.
- **Indicador Beneficiario:** Almacena información del modelo de competencias en TIC (Competencias, Desempeños e Indicadores) definidas por CPE para los docentes beneficiados del programa [2], y del modelo establecido para los estudiantes; relacionándolas con las dimensiones: Beneficiario, Encuesta, Entidad Educativa, Formador, Ruta CPE, Cuestionario, Convenio, Universidad, Fecha y la medida *Valor Indicador*, la cual representa al valor obtenido para cierto indicador, desempeño o Competencia de un beneficiario según sus respuestas al Cuestionario aplicado. El modelo de alto nivel es mostrado en la Figura 12. **Diagrama Data Mart Indicador Beneficiario.**

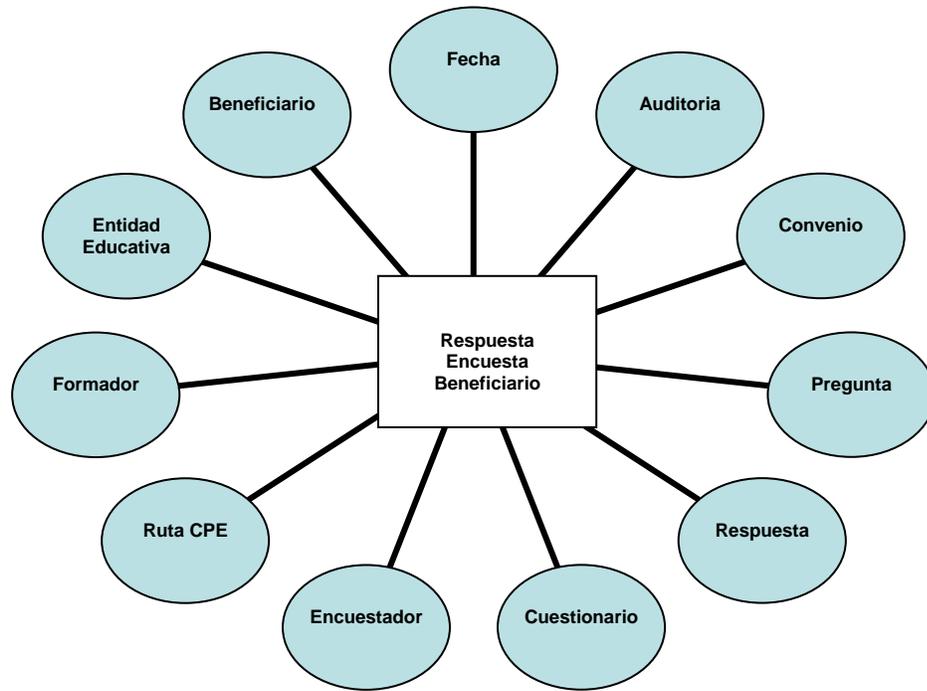


Figura 11. Diagrama Data Mart Respuesta Encuesta Beneficiario

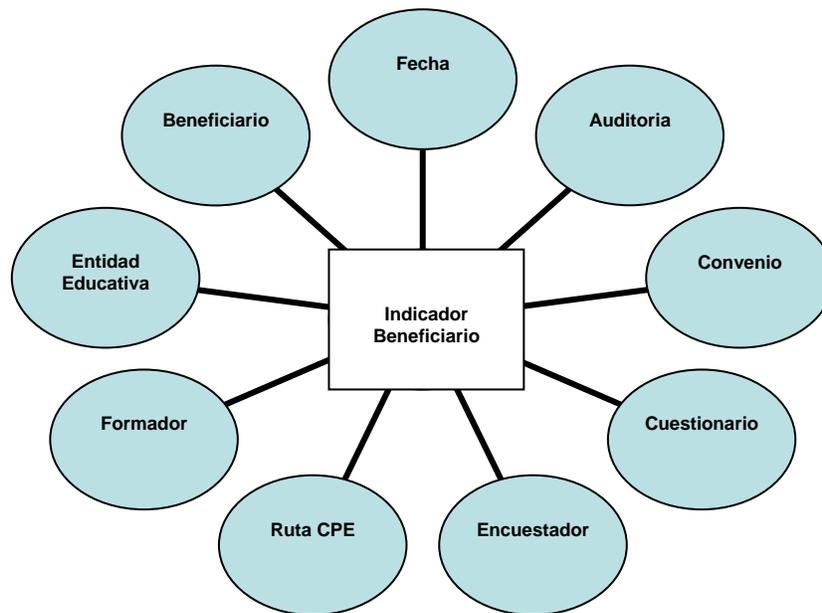


Figura 12. Diagrama Data Mart Indicador Beneficiario

- **Data Mart Jornada de Formación:** En este proceso del negocio se almacena información general de las Jornada de Formación impartidas durante la fase de Profundización, tales como, el nombre de la jornada, cantidad de horas dictadas

en los ejes pedagógico, infraestructura y de gestión, temas y recursos educativos utilizados por Jornada de Formación. El modelo de alto nivel construido es mostrado en la siguiente figura.

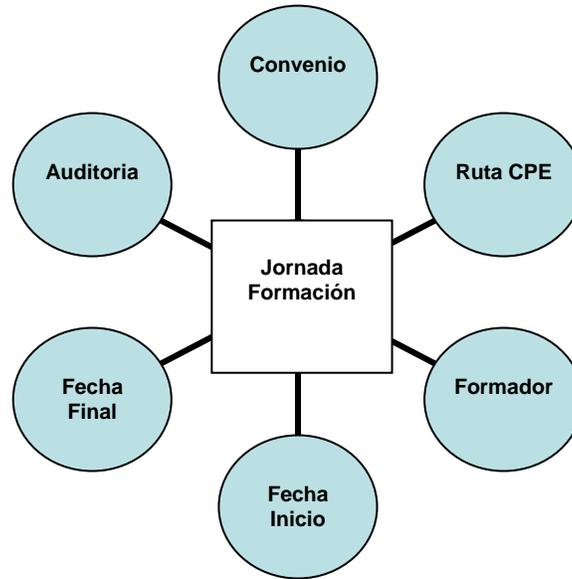


Figura 13. Diagrama Data Mart Jornada de Formación

Revisión y validación

Esta etapa involucra revisar los modelos dimensionales obtenidos en reuniones conjuntas: expertos técnicos, administrativos del negocio y equipo de desarrollo del sistema de DSS. De la misma manera se identifican algunas necesidades analíticas extras que deberían soportar los modelos, lo que implica modificar la estructura de los mismos [7].

En el Capítulo V de este documento se presentan los casos de diseño que se presentaron durante el desarrollo del proyecto, para los cuales se tuvieron en cuenta los sugeridos por Kimball[8], aplicándolos o adaptándolos según fuera el caso de modelado presentado durante el diseño del DSS.

Finalmente, luego de haber realizado las últimas revisiones y validaciones se construyen los modelos dimensionales finales. Cada modelo dimensional representa un proceso de negocio o DM y la unión de estos representa el modelo completo del DW.

A continuación se presentan los cuatro modelos dimensionales finales que representan el DW para el DSS. Estos cuatro modelos dimensionales dan origen a la construcción del diseño físico del DW y a su posterior proceso de extracción, transformación y carga

JERARQUÍAS

Jerarquías de Usuario

A continuación se presentan algunas de las jerarquías de usuario que se diseñaron e implementaron en el DSS.

Jerarquía - Dimensión Área de Enseñanza

Sobre la dimensión Área de Enseñanza se creó la jerarquía *Especialidad Área de Enseñanza* (Figura 14), la cual está constituida por la siguiente jerarquía de atributos: Carácter Especialidad->Especialidad->Área Enseñanza.

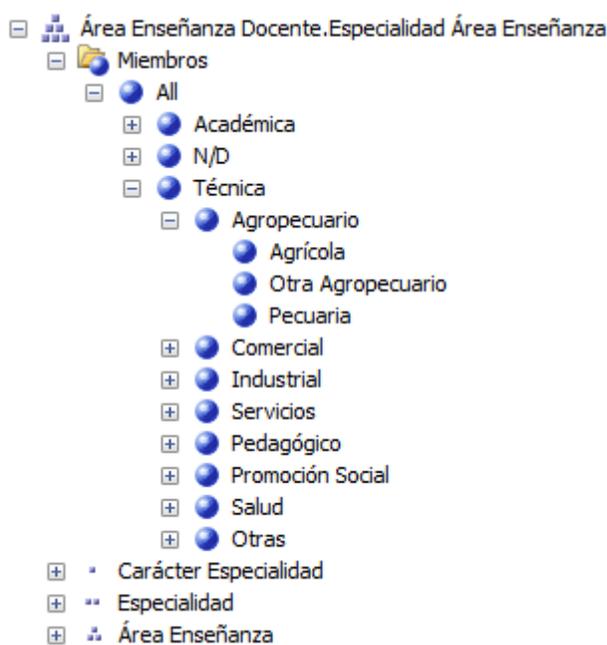


Figura 14. Jerarquía – Dimensión Área de Enseñanza

Jerarquía - Dimensión Beneficiario

Sobre la dimensión Beneficiario se crearon las siguientes jerarquías:

- *Lugar Nacimiento Beneficiario* (Tipo Beneficiario->Localización Nacimiento->Beneficiario),
- *Lugar Residencia Beneficiario* (Tipo Beneficiario->Localización Residencia->Beneficiario),
- Género – Estrato Residencia Beneficiario (Tipo Beneficiario->Género Beneficiario->Beneficiario),

A través de ésta última jerarquía se le permite al usuario navegar los datos del beneficiario por Tipo Beneficiario (Docente/Directivo Docente, Directivo Docente, Docente, Estudiante Primaria Básica, Estudiante Secundaria Básica), género y estrato de residencia.

Jerarquía – Dimensión Docente

En la creación de la dimensión Docente se establecieron las jerarquías para manejar la localización del lugar de residencia y para el lugar de nacimiento, tal como se especifica a continuación.

- Género – Estrato Residencia Docente (Tipo Docente -> Género Docente -> Estrato Residencia Docente -> Docente)
- Localización Nacimiento Docente (Tipo Docente -> Localización Nacimiento Docente -> Docente)
- Género – Grado Escalafón Docente
- Cargo – Nivel Educativo Docente
- Cargo – Tipo Vinculación Docente
- Localización Residencia Docente

Jerarquía – Dimensión Entidad Educativa

Para la jerarquía de la dimensión Entidad Educativa, se establecieron los siguientes niveles: Región, Departamento, Clase Municipio (Capital, Normal), Zona (Rural, Urbana), Localidad y Entidad Educativa, para detallar la localización de las Entidades Educativas. Y para el Carácter de la Entidad Educativa se definieron los siguientes niveles: Carácter y Entidad Educativa.

Jerarquía – Dimensión Fecha

Se crearon dos jerarquías *Día Mes Año* y *Día Semana Año*, para manejar la exploración de las diferentes registros de fechas de las dimensiones con las cuales está relacionada esta dimensión (los diferentes roles de la dimensión), tales como: Auditoría, fecha inicio Convenio, fecha inicio EAE, fecha inicio Jornada Formación, fecha final Convenio, fecha final EAE, fecha final Jornada Formación.

Jerarquía – Dimensión Grado Enseñanza

En la dimensión Grado Enseñanza se estableció la jerarquía Nivel Educativo para agrupar adecuadamente el grado de enseñanza de un docente, por los siguientes niveles: Carácter (Académico, Técnico), Nivel Educativo (Pre-Escolar, Primaria Básica, Secundaria Básica, etc.) y finalmente el grado de enseñanza (Primero, Cuarto, Décimo, Once, etc.).

Jerarquía – Dimensión Ruta CPE

El programa CPE es ejecutado por varios grupos de trabajo en diferentes partes del país, los cuales son los encargados de beneficiar a las instituciones educativas ubicadas en su región y ejecutar todas las actividades del programa (Región: sur Pacífica, Centro-Islands, Nororiente, entre otras). En la dimensión Ruta CPE se tiene una jerarquía que involucra esta clasificación de los diferentes convenios de las universidades con CPE (Jerarquía *Región Ruta CPE*), el cual está hecho por Regiones a lo largo del país, adicionalmente incluye otro nivel, Ruta CPE, el cual especifica las diferentes rutas de acción y recorrido dentro de una región específica, las cuales son asignadas a los formadores para realizar sus respectivas visitas a las instituciones educativas beneficiadas.

Jerarquía – Dimensión Universidad

En la dimensión Universidad se creó una jerarquía para manejar la localización de la universidad encargada de cierta región por convenio con CPE, la cual tiene los siguientes niveles: Región, Departamento, Clase (Capital, Normal), Municipio, Zona (Urbana, Rural) y finalmente la Universidad.

Jerarquía – Dimensión Respuesta Beneficiario

Esta dimensión (puente Respuesta Cuestionario) es la que nos permite relacionar la encuesta con las preguntas y las respuestas hechas por un beneficiario en particular; para organizar esta información y poder explorarla de manera adecuada esta la jerarquía *Respuesta Pregunta Encuesta*, la cual nos agrupa las respuestas de una pregunta, y las preguntas de una encuestas respectivamente, de un docente o estudiante encuestado; usando los niveles que tiene la jerarquía creada.

Para ver en detalle las jerarquías creadas, ver el **Anexo 24**.

4.1.4. Diseño Físico del Data Warehouse

Teniendo en cuenta que el convenio CPE–Universidad del Cauca solo cuenta con un servidor, se definió como arquitectura física la denominada “ALL IN ONE”[7]. El proceso de cargue no tiene requerimientos grandes de almacenamiento y procesamiento, debido a que la cantidad de registros cargados en la bodega de datos son pocos. La Figura 15 muestra esta configuración.

En el **Anexo 25** se presentan los estándares de nombramiento utilizados.



Figura 15. Configuración ALL IN ONE para un Sistema de Inteligencia de Negocios

CONSTRUCCION DE LA BASE DE DATOS MULTIDIMENSIONAL

Posterior a la construcción y cargar del DW relacional, se continua con la construcción de la base de datos multidimensional con la herramienta Analysis Services de Microsoft Business Intelligence. Esta fase se enmarca dentro del camino de datos del ciclo de vida dimensional, donde se realiza el modelado dimensional, diseño físico y el proceso ETL para poblar la base de datos.

Para la construcción de la base de datos multidimensional en Analysis Services se tuvieron los siguientes pasos [8]:

- Preparar el diseño y ambiente de desarrollo
- Crear una Vista de Origen de datos
- Crear y refinar las dimensiones
- Crear y Editar el cubo
- Procesar el cubo
- Crear cálculos
- Realizar iteraciones

Al tener creado el cubo, podemos acceder a su interfaz de diseño dentro del Business Intelligence Development Studio como se ve en la Figura 16. **Interfaz de diseño para Cubos Multidimensionales.**

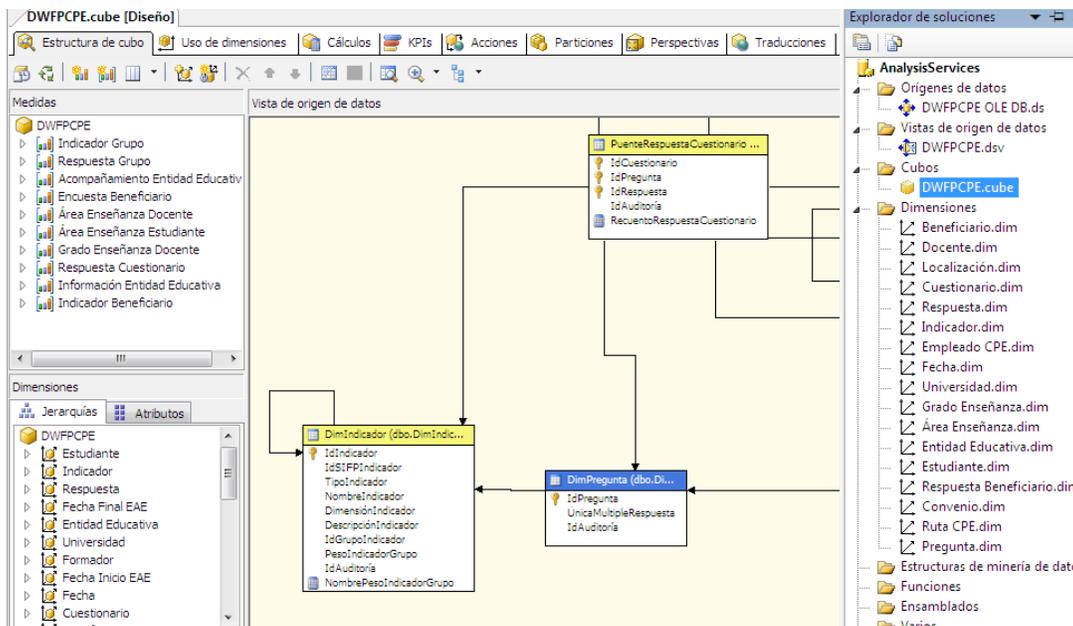


Figura 16. Interfaz de diseño para Cubos Multidimensionales

A continuación se explicara de manera muy general lo que nos permite realizar la interfaz de diseño del cubo.

ESTRUCTURA DEL CUBO

Es el lugar de edición principal para la estructura del cubo (propiedades de los grupos de medida, medidas y dimensiones del cubo). En la siguiente figura se observa la estructura del cubo, junto con las medidas y dimensiones creadas.

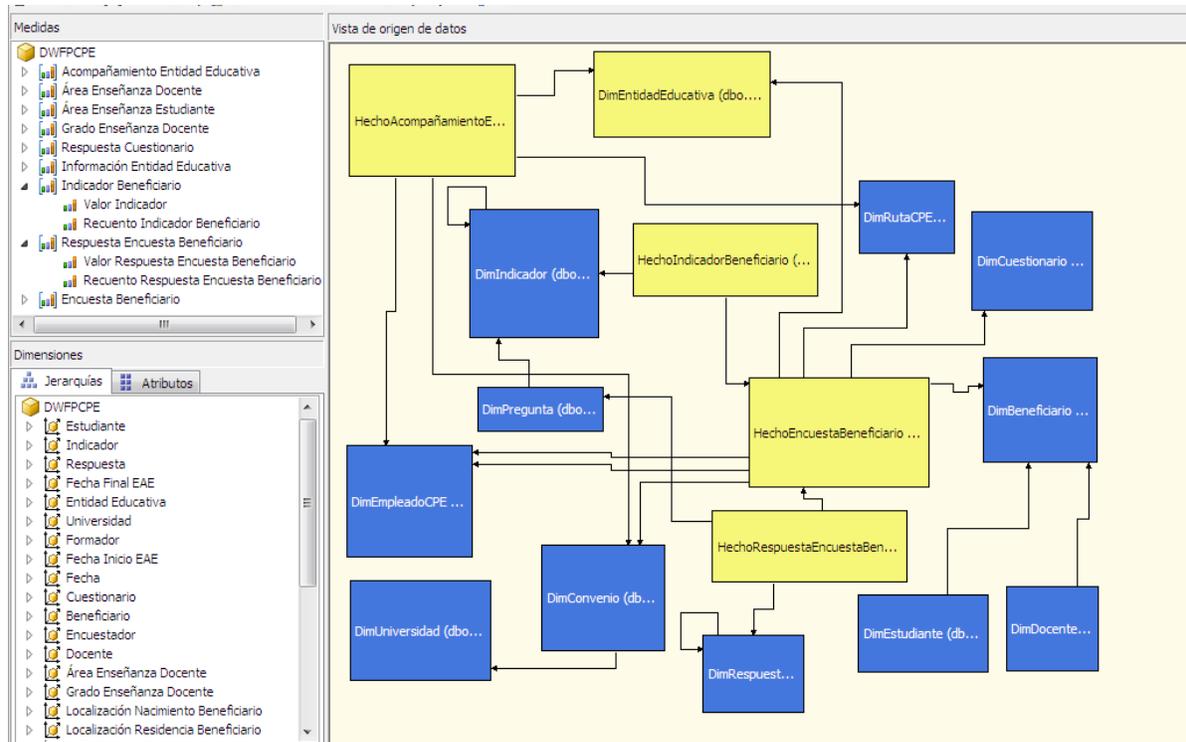


Figura 17. Estructura del Cubo, Medidas y Dimensiones

Un aspecto importante en el diseño, es el juego de roles de las dimensiones, en el desarrollo del proyecto se presentó para las dimensiones Localización, Empleado CPE y Fecha, para las cuales se generan dimensiones de cubo por cada rol existente entre la dimensión y la tabla de hechos u otra dimensión.

USO DE DIMENSIONES

Es el lugar donde podemos revisar los tipos de relaciones de las dimensiones con los grupos de medidas y configurar las relaciones que aun no están definidas. En los grupos de medida se tienen relaciones de tipo normal, a continuación se describen relaciones especiales entre las dimensiones existentes:

- Sub-dimensión Universidad: se establece una relación de referencia con el grupo de medida Acompañamiento Entidad Educativa y Encuesta Beneficiario (tablas de hechos), por medio de la dimensión intermedia Convenio.
- Sub-dimensión Estudiante: se establece una relación de referencia con los grupos de medida Encuesta Beneficiario e Indicador Beneficiario, por medio de la dimensión intermedia Beneficiario.

- Sub-dimensión Docente: se establece una relación de referencia con los grupos de medida Encuesta Beneficiario e Indicador Beneficiario, por medio de la dimensión intermedia Beneficiario.
- Sub-dimensión Localización (Rol Localización Nacimiento Beneficiario): se establece una relación de referencia con el grupo de medida Encuesta Beneficiario por medio de la dimensión Beneficiario; con el grupo de medida Área de enseñanza Docente, por medio de la dimensión Docente; con el grupo de medida Área de enseñanza Estudiante, por medio de la dimensión Estudiante; y con el grupo de medida Grado de enseñanza Docente por medio de la dimensión Docente.
- Sub-dimensión Localización (Rol Localización Residencia Beneficiario): se establece una relación de referencia con el grupo de medida Encuesta Beneficiario por medio de la dimensión Beneficiario; con el grupo de medida Área de enseñanza Docente, por medio de la dimensión Docente; con el grupo de medida Área de enseñanza Estudiante, por medio de la dimensión Estudiante; y con el grupo de medida Grado de enseñanza Docente por medio de la dimensión Docente.
- Sub-dimensión Localización (Rol Localización Entidad Educativa): se establece una relación de referencia con el grupo de medida Encuesta Beneficiario por medio de la dimensión Entidad Educativa y con el grupo de medida Acompañamiento Educativo por medio de la dimensión Entidad Educativa.
- Sub-Dimensión Localización (Rol Localización Nacimiento Formador): tiene una relación de referencia con el grupo de medida Acompañamiento Entidad Educativa y Encuesta Beneficiario por medio de la dimensión Formador.
- Sub-Dimensión Localización (Rol Localización Residencia Formador): tiene una relación de referencia con el grupo de medida Acompañamiento Entidad Educativa y Encuesta Beneficiario por medio de la dimensión Formador.
- Sub-Dimensión Localización (Rol Localización Nacimiento Encuestador): tiene una relación de referencia con el grupo de medida Acompañamiento Entidad Educativa y Encuesta Beneficiario por medio de la dimensión Encuestador.
- Sub-Dimensión Localización (Rol Localización Residencia Encuestador): tiene una relación de referencia con el grupo de medida Acompañamiento Entidad Educativa y Encuesta Beneficiario por medio de la dimensión Encuestador.
- Dimensión Indicador: presenta una relación reflexiva, por lo que se debe crear una relación padre – hijo para poder visualizar el árbol que representa el modelo de competencias en TIC establecido para los beneficiarios del programa CPE.
- Dimensión Respuesta: presenta una relación reflexiva, por lo que se debe crear una relación padre – hijo para poder visualizar el árbol que representa la estructura definida para todas las posibles respuestas de un cuestionario.

En la siguiente figura se muestra la interfaz de configuración del uso de dimensiones, donde se observan las dimensiones y el tipo de relaciones con los grupos de medida.

Grupos de medida			
Dimensiones	Indicador Beneficiario	Respuesta Encuesta Beneficiario	Encuesta Beneficiario
Estudiante	Encuesta Beneficiario	Encuesta Beneficiario	Beneficiario
Indicador	Indicador		Indicador Beneficiario
Respuesta		Respuesta	Respuesta Encuesta Beneficiario
Fecha (Fecha Final EAE)			
Entidad Educativa	Encuesta Beneficiario	Encuesta Beneficiario	Entidad Educativa
Universidad	Encuesta Beneficiario	Encuesta Beneficiario	Convenio
Empleado CPE (Formador)	Encuesta Beneficiario	Encuesta Beneficiario	Empleado CPE
Fecha (Fecha Inicio EAE)			
Fecha	Encuesta Beneficiario	Encuesta Beneficiario	Fecha
Cuestionario	Encuesta Beneficiario	Encuesta Beneficiario	Cuestionario
Beneficiario	Encuesta Beneficiario	Encuesta Beneficiario	Beneficiario
Empleado CPE (Encuestador)	Encuesta Beneficiario	Encuesta Beneficiario	Empleado CPE
Docente	Encuesta Beneficiario	Encuesta Beneficiario	Beneficiario
Área Enseñanza (Área Enseñanza Docente)	Encuesta Beneficiario	Encuesta Beneficiario	Área Enseñanza Docente
Grado Enseñanza (Grado Enseñanza Docente)	Encuesta Beneficiario	Encuesta Beneficiario	Grado Enseñanza Docente
Localización (Localización Nacimiento Beneficiario)	Encuesta Beneficiario	Encuesta Beneficiario	Beneficiario
Localización (Localización Residencia Beneficiario)	Encuesta Beneficiario	Encuesta Beneficiario	Beneficiario
Fecha (Fecha Final Convenio)	Encuesta Beneficiario	Encuesta Beneficiario	Convenio
Fecha (Fecha Inicio Convenio)	Encuesta Beneficiario	Encuesta Beneficiario	Convenio
Convenio	Encuesta Beneficiario	Encuesta Beneficiario	Convenio
Ruta CPE	Encuesta Beneficiario	Encuesta Beneficiario	Ruta CPE
Localización (Localización Nacimiento Formador)	Encuesta Beneficiario	Encuesta Beneficiario	Formador
Localización (Localización Residencia Formador)	Encuesta Beneficiario	Encuesta Beneficiario	Formador
Localización (Localización Nacimiento Encuesta...)	Encuesta Beneficiario	Encuesta Beneficiario	Encuestador
Localización (Localización Residencia Encuesta...)	Encuesta Beneficiario	Encuesta Beneficiario	Encuestador
Fecha (Fecha Inicio Jornada Formación)			
Fecha (Fecha Final Jornada Formación)			
Pregunta		Pregunta	Respuesta Encuesta Beneficiario
Encuesta	Encuesta	Encuesta	Encuesta
Área Enseñanza (Área Enseñanza Estudiante)	Encuesta Beneficiario	Encuesta Beneficiario	Área Enseñanza Estudiante

Figura 18. Uso de Dimensiones – Tipos de relaciones entre Dimensiones y Grupos de Medida

Por último se procede a procesar el cubo, cuando este se ha completado en su totalidad y sin errores, se procede a examinar los datos a través de cruces dimensionales para resolver posibles inconsistencias. Cuando el cubo esta correctamente generado, se puede realizar sobre él consultas OLAP a través de una herramienta para Análisis de Datos Multidimensionales.

4.1.5. Diseño del Sistema ETL

Las actividades principales de este proceso son la Extracción que se utiliza para obtener los datos desde diferentes fuentes de datos; la Transformación que hace referencia a la limpieza, homogenización y conversión de los datos; y la Carga de datos que se encarga de poblar los datos en la bodega de datos [7]. Además se diseñó y se construyó el sistema de Auditoría que tiene como función principal realizar un seguimiento completo de los datos que son cargados y actualizados dentro de la bodega de datos.

El primer paso del desarrollo del sistema ETL es empezar con los diagramas de Mapas de Alto Nivel de ETL, para lo cual se deben identificar de ante mano las fuentes de datos, las posibles transformaciones para finalmente llegar al destino de los datos (DW relacional)[8].

El grupo de trabajo del convenio CPE–Universidad del Cauca proporcionó información de las entidades educativas, los formadores, tipo de recursos educativos utilizados en las jornadas de formación y datos relacionados con el modelo de competencias en TIC, manejado por CPE para los docentes beneficiarios por el programa; toda esta información en su mayoría presentaba homogeneidad en valores y formatos de los datos (archivos de MS Excel). En cuanto a la información detallada de docentes y estudiantes beneficiarios, fue necesario el diseño de un instrumento para su recolección (encuestas), para luego ser almacenada en el prototipo OLTP (base de datos hecha en MS Access, descrita en la sección de Metodología para la Recolección de datos); lo anterior resultó ser un factor positivo para el desarrollo del proceso de ETL, ya que se estructuró la información de manera adecuada pensando en el desarrollo futuro de la bodega de datos.

Luego de identificar las fuentes de datos se construyeron los diagramas de Alto Nivel de ETL para cada una de las dimensiones y tablas de hechos del DW relacional, identificando sus respectivos orígenes, transformación y destinos.

En la Figura 19 se muestra el flujo realizado para la carga de los datos que pertenecen a la dimensión Indicador, estos datos se encuentran en el archivo de MS Access: DBFPCPE, donde se almacena toda la información de los indicadores (Nombre, Tipo, Dimensión, Descripción, entre otros). Los demás diagramas se encuentran en el **Anexo 26**.



Figura 19. Diagrama de Alto Nivel ETL

Para el desarrollo del sistema ETL se utilizó la herramienta Microsoft SQL Server 2005 Integration Services (SSIS), la cual es una plataforma orientada a la creación de datos de alto rendimiento, incluyendo paquetes de extracción, transformación y carga (ETL) de datos. Esta herramienta incluye utilidades gráficas y asistentes para la creación y depuración de los paquetes; tareas para realizar funciones de flujo de datos, como operaciones FTP, ejecución de instrucciones SQL, transformaciones para borrar, agregar y copiar datos, entre otras funciones que permiten una serie de cambios sobre los datos [8].

Para la implementación del proceso ETL se hace uso de los paquetes de Integration Services, estos son archivos XML (con extensión .dtsx). Un paquete contiene una o más tareas que permiten hacer manipulaciones sobre los datos.

Se tuvieron en cuenta las prácticas de diseño y construcción propuestas por Kimball [8], creando paquetes hijos para la carga de cada una de las tablas, y paquetes maestros que contienen las tareas de ejecución de paquetes, haciendo que la secuencia de carga de tablas sea clara y adecuada. Estas técnicas de construcción de paquetes incrementan la flexibilidad y el entendimiento del sistema.

Para el sistema ETL se crearon un total de 23 paquetes hijo y 1 paquete maestro. En la Figura 20. **Paquete ETL de la Dimensión Pregunta** se muestra un paquete (DimPregunta.dtsx) y su flujo de tareas que hacen la extracción, transformación y carga de datos en la Dimensión Pregunta. Los demás paquetes con sus flujos y tareas se encuentran en el **Anexo 27**.

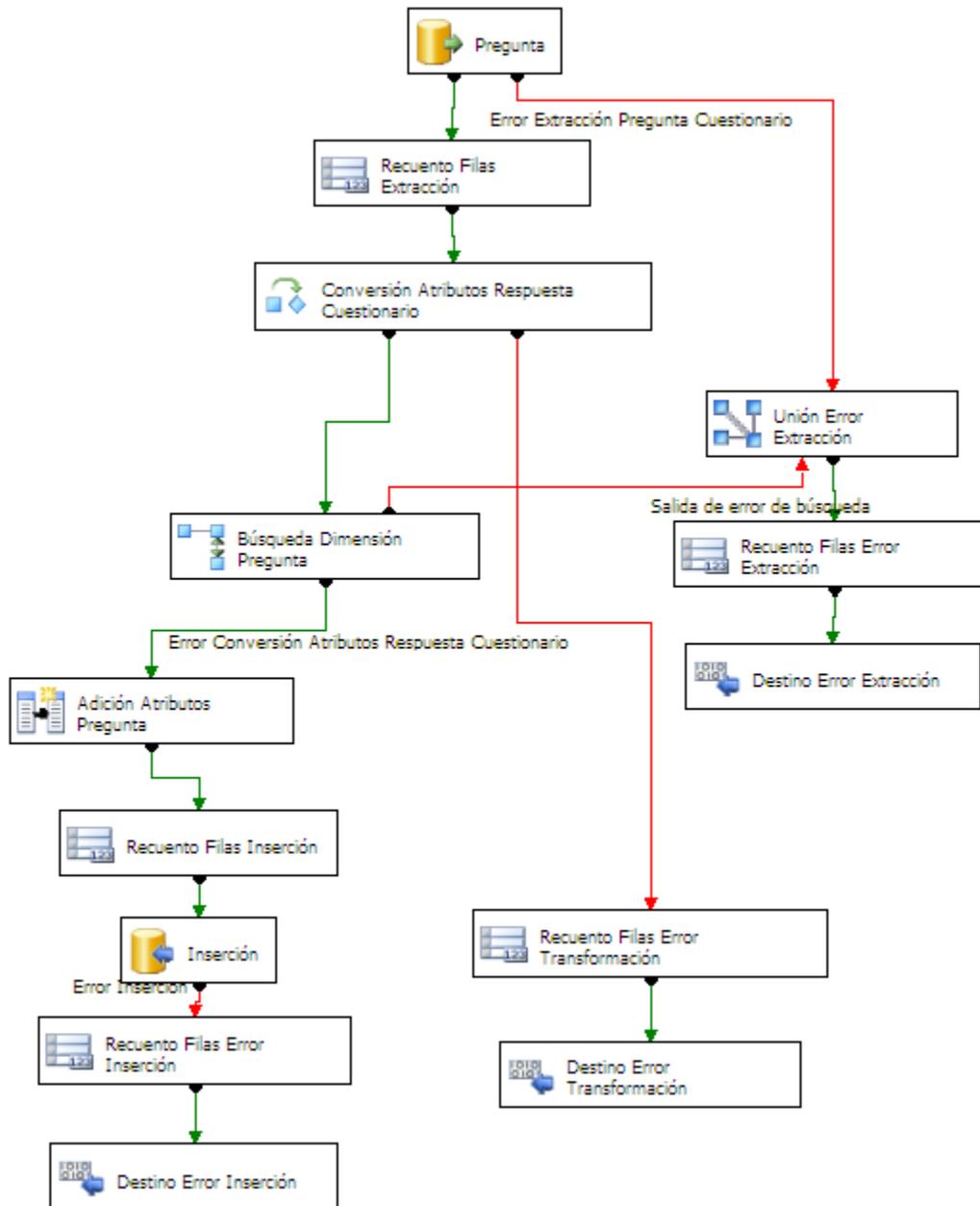


Figura 20. Paquete ETL de la Dimensión Pregunta

Para la implementación del proceso ETL fue necesario crear dos tablas temporales (PuenteRespuestaCuestionario y PuentePreguntaCuestionario, ver Figura 21.), que permitieran realizar el proceso ETL de aquellas dimensiones que presentaban relaciones reflexivas, ya que en el prototipo OLTP creado en MS Access no soporta este tipo de relaciones.

4.1.6. Diseño de la Arquitectura Técnica

Dentro de un ambiente de DW/BI se presenta la integración de varias tecnologías, es por esto que para poder definir su arquitectura técnica es necesario tener en cuenta aspectos tales como: los requerimientos del negocio, los ambientes técnicos actuales y lineamientos técnicos de planeación [7]. En esta etapa se identifican las herramientas y tecnologías que se requieren para poder desarrollar el sistema.

La arquitectura se divide en 2 grandes partes conocidas como Back Room (parte interna) donde se describen los orígenes de datos, el flujo de datos, consulta de servicios, metadata, entre otros; y el Front Room (parte externa) donde se describe las herramientas estáticas y dinámicas para acceder y consultar los datos del DW [7]. Además está conformada por cuatro componentes (Datos operacionales, Procesamiento de datos, Bodega de datos y DM, y Componentes de acceso).

4.1.7. SELECCIÓN E INSTALACIÓN DEL PRODUCTO

Basado en la arquitectura técnica del DSS presentado anteriormente, se evalúan y se seleccionan los componentes específicos para la arquitectura, tales como la plataforma hardware, el motor de base de datos, la herramienta de ETL, herramientas de acceso y otros. Una vez evaluados y seleccionados los productos, se procede con la instalación y prueba de los mismos en un ambiente integrado de DW [7].

Se presenta en el **Anexo 29** los cuadros de selección de las herramientas para el desarrollo del proyecto. Teniendo en cuenta adicionalmente, la experiencia del equipo de desarrollo del proyecto en el uso de las herramientas de SQL Server Enterprise Edition, y que al ser este un proyecto de investigación se tienen disponibles las Licencias de Alianza Académica para estas herramientas durante el proceso de desarrollo.

Por lo mencionado anteriormente se selecciono SQL Server enterprise Edition como motor de base de datos relacional y multidimensional junto con sus respectivos ambientes de desarrollo y de administración integrados, estos son: Business Intelligence Development Studio (Herramienta de diseño y construcción) y SQL Server Management Studio (Herramienta primaria de administración). Estas herramientas proveen todas las funcionalidades requeridas para desarrollar completamente un sistema de DW como producto para el desarrollo del proyecto.

Teniendo en cuenta la arquitectura técnica se instalaron las siguientes herramientas de desarrollo y administración de SQL Server 2005 [8] (Ver Figura 23. **Arquitectura un sistema Microsoft de DW/BI**):

SQL Server Management Studio: se usa para administrar el sistema de DW. A través de esta herramienta se puede conectar y administrar los servidores de base de datos relacionales, de Analysis Services (Bases de Datos Multidimensionales), Integration Services (Servicios de ETL) y Reporting Services (Servidor de Reportes OLAP). También permite realizar consultas de tipo MDX, DMX y XMLA para bases de datos de análisis dimensional como por ejemplo cubos y modelos de minería de datos.

Business Intelligence Development Studio: se utiliza para desarrollar y diseñar sistemas de DW, esta herramienta proporciona interfaces de usuario que facilitan el diseño, desarrollo e implementación de bases de datos de Analysis Services, modelos de minería de datos, paquetes de Integration Services y reportes de Reporting Services.

Analysis Services (AS): Es un motor OLAP que permite almacenar, gestionar y consultar bases de datos OLAP (multidimensionales), diseñadas y construidas para usos de Inteligencia de Negocios. AS tiene las siguientes características:

- **Metadata orientada al usuario:** La estructura de las bases de datos de AS permite definir nombre de los objetos (dimensiones, hechos, jerarquías) orientados al usuario, que le permite navegar fácilmente sobre los datos.
- **Análisis Complejos almacenado en la BD:** Las bases de datos de AS almacenan información sobre los cálculos, desde cálculos muy simples hasta muy complejos, como miembros calculados, indicadores de desempeño, cálculos inteligentes, perspectivas específicas, medidas semi-aditivas, traducciones y otros que permiten construir una base de datos OLAP con mejores características.

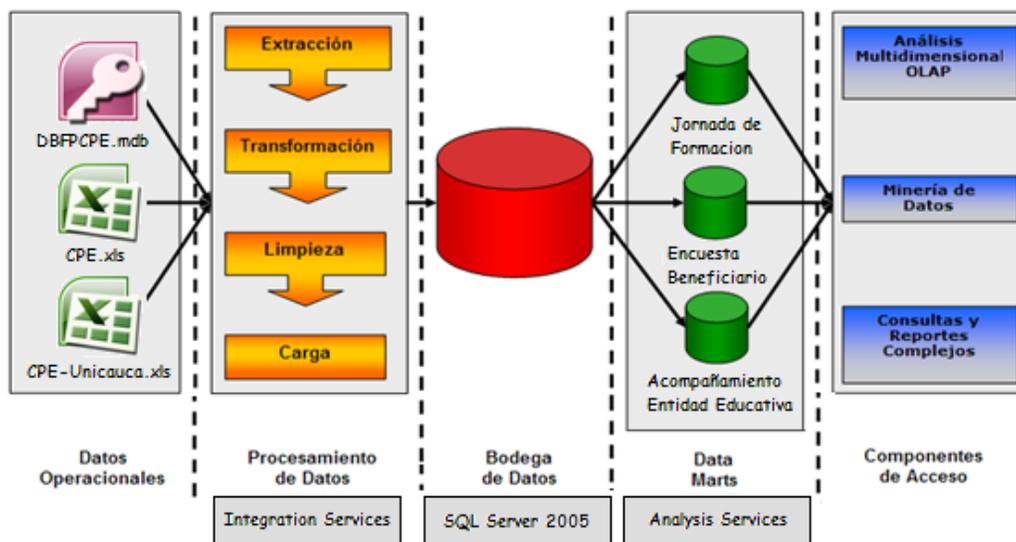


Figura 23. Arquitectura un sistema Microsoft de DW/BI, adaptada de [7]

- **Riqueza del lenguaje Analítico:** SQL es un lenguaje orientado a las consultas, no es un lenguaje analítico, las herramientas de consulta para Analysis Services generan MDX (Expresiones Multidimensionales) en vez de SQL. La fortaleza más grande del MDX es el entendimiento de la metadata dimensional: Hechos, dimensiones, atributos, jerarquías, miembros, etc.
- **Desempeño de Consulta:** El motor OLAP ha sido diseñado como un servidor de alto desempeño. AS tiene características de almacenamiento y gestión de tablas resúmenes o agregadas que mejoran significativamente el desempeño de las consultas.

- **Minería de Datos:** Provee las herramientas y funcionalidades necesarias para desarrollar modelos de minería de Datos.

Integration Services (IS): Es una herramienta de diseño y desarrollo de paquetes que son usados para hacer extracción, transformación y carga del DW relacional. Sin embargo IS es más que una herramienta de ETL, esta puede ser usada para gestionar y mantener bases de datos e integrar datos entre aplicaciones o realizar cálculos complejos en tiempo real.

Reporting Services: Es una herramienta de reportes que permite el almacenamiento, la gestión y consulta de reportes predefinidos. Igualmente provee a los usuarios la posibilidad de construir informas ad hoc (personalizados) simples, por medio de una herramienta llamada el “Report Builder” (constructor de Reportes).

Para la selección de las herramientas del Front Room de la arquitectura técnica mencionadas anteriormente, en la Tabla 7. **Comparación Herramientas OLAP** se presenta la comparación de herramientas para la selección de la herramienta OLAP, se selecciono Dundas OLAP principalmente por su fácil integración con SQL Server Analysis Services, utilizada en el desarrollo del DW, y por permitir las consultas Off Line utilizando cubos locales.

Tabla 7. Comparación Herramientas OLAP

Característica	Peso	Pentaho Report Designer	Wabit	Palo Web Client	Jmagallanes	Dundas OLAP
Documentación	10	3	4	3	2	4
Facilidad de instalación	5	5	5	4	4	4
Open Source	5	5	5	5	5	1
Permite crear reportes ad hoc	15	5	5	5	5	5
Proporciona una interfaz web de usuario final	15	5	0	0	0	5
Facilidad de Uso	15	3	4	3	3	3
Integración con SQL Server Analysis Services	20	1	1	1	1	5
Consultas Off Line - Cubos Locales	15	1	1	1	1	5
Puntaje Total	100	310	260	230	220	435

En la columna Peso se especifica el porcentaje de importancia para la selección de la herramienta, y en las siguientes columnas se asigna un valor de 1 a 5 según se cumpla con la característica evaluada. Para ver en más detalle el proceso de integración de la herramienta OLAP con el DSS ver el apartado **Especificación de Aplicaciones de Usuario Final.**

4.1.8. Especificación de Aplicaciones de Usuario Final

Reportes Estándar

En esta etapa se construyen los reportes estándares, que presenten información interesante para analizar, por parte de los usuarios del sistema. En la siguiente figura se presenta la plantilla estándar para los reportes.



COMPUTADORES PARA EDUCAR

[NOMBRE DEL CONVENIO]

[NOMBRE DEL REPORTE]

[DATOS DEL REPORTE]			

Figura 24. Plantilla Estándar

En la Tabla 8 se presentan los reportes a ser construidos y publicados, revisando las necesidades analíticas identificadas en la etapa de recolección de requerimientos del ciclo de vida dimensional.

Tabla 8. Reportes Estándar

Nombre Reporte	Descripción	Categoría del Reporte	Tipo de Reporte	Comentarios
Indicador por Entidad Educativa	Presenta el promedio del valor del indicador agrupando el resultado por Entidad Educativa, por Tipo de Beneficiario y por Género.	Indicador Beneficiario	Matriz	El indicador es un parámetro del reporte
Indicador por Nivel Educativo Docente	Presenta el promedio del valor del indicador agrupando el resultado por Entidad Educativa, por Cargo del docente y por su Nivel Educativo.	Indicador Beneficiario	Matriz	El indicador es un parámetro del reporte
Indicador por Género Estudiante	Presenta el promedio del valor del indicador agrupando el resultado por Entidad Educativa y por el Género de los Estudiantes	Indicador Beneficiario	Matriz	El indicador es un parámetro del reporte

En la siguiente figura se presenta el resultado del reporte estándar Indicador por Entidad Educativa, donde se puede apreciar los valores promedio del indicador “Utiliza los procesadores de texto”, agrupados por Entidad Educativa, por Tipo de Beneficiario y por Genero del Beneficiario. En el **Anexo 30** se encuentran los resultados de todos los reportes estándar creados.



COMPUTADORES PARA EDUCAR

CONVENIO CPE-UNICAUCA

Indicador por Entidad Educativa

			Indicador
CENTRO EDUCATIVO EL EMPALME SEDE CENTRO DE EDUCACIÓN BÁSICA EL EMPALME	Estudiante de Secundaria Básica	Femenino	0,279
		Masculino	0,309
CENTRO EDUCATIVO GUAPUSCAL ALTO	Docente	Femenino	0,316
		Masculino	0,470
CENTRO EDUCATIVO RURAL ALTO AFAN	Estudiante de Secundaria Básica	Femenino	0,276
		Masculino	0,391
	Docente	Femenino	0,107
		Masculino	0,127
CENTRO EDUCATIVO RURAL COCAYÁ SEDE ESCUELA LA PLAYA HONG KONG	Docente	Femenino	0,196
		Masculino	0,131
CENTRO EDUCATIVO SANTA ROSA DE LIMA SEDE PRINCIPAL	Docente	Femenino	0,208
		Masculino	0,321
ESCUELA RURAL MIXTA LOMA DE ZURAS	Docente	Femenino	0,219
		Masculino	0,136
	Directivo Docente	Masculino	0,414

Figura 25. Reporte Estándar – Indicador por Entidad Educativa

4.1.9. DESARROLLO DE LA APLICACIÓN DE USUARIO FINAL

Integración Herramienta OLAP

Según los criterios presentados en el apartado **Herramientas OLAP**, la herramienta seleccionada para realizar la integración con el DSS fue Dundas OLAP Services for Windows Forms.

Las herramientas OLAP son ampliamente utilizadas en inteligencia de negocios, minería de datos y presentación de informes, son aplicaciones que permiten realizar análisis de gran cantidad de datos en un corto período de tiempo. Dundas OLAP Services amplía el Enterprise de Dundas Chart for .NET para ayudar a las empresas en el análisis. Fácil de usar, permiten a los usuarios construir nuevas consultas de datos con operaciones simples de drag-and-drop, a continuación se presentan algunas de las principales características de este producto:

- Cliente OLAP: Establecer la conexión al cubo OLAP para iniciar el análisis inmediatamente, usando el visor gráfico o la grilla.
- Simple, drag and drop, funcionalidad de construir consultas sin escribir ningún código.
- Interfaz de usuario simple que permite fácilmente crear, manejar y serializar reportes.
- Despliegue claro de estructuras de metadatos complejas, organizadas en estructuras de árboles.
- Filtros avanzados, ordenamiento y capacidades drilldown.
- Integración completa para todos los controles de usuario en Visual Studio 2005
- Drill up y drill down para varianzas de abstracción y nivel de detalle.
- Función Pivot, que da al usuario diferentes perspectivas de un mismo grupo de datos.

- Múltiple análisis de cubos que residen el mismo servidor, con un simple control.
- Soporte de KPI, ofrece la habilidad a los usuarios de desplegar los indicadores claves de desempeño directamente desde el cubo OLAP.
- Permite personalizar el lenguaje de los controles OLAP.

El instalador de Dundas OLAP Services for Windows Forms se encuentra disponible en la página Web oficial (<http://www.dundas.com>), versión de evaluación.

Para esta fase del proyecto se desarrollo un prototipo de aplicación de consultas OLAP teniendo en cuenta la documentación, ejemplos y demos incluidos en la herramienta descargada, y utilizando los componentes ofrecidos por la misma. A continuación se presenta el proceso para utilizar la herramienta y realizar los reportes sobre la bodega de datos.

En la siguiente figura se presenta la primera interfaz del prototipo, donde se especifica:

- El tipo de proveedor (Provider type)
- Tipo de conexión (Connection Type)
- El nombre del servidor (Server Name)
- Catalogo (Catalog)

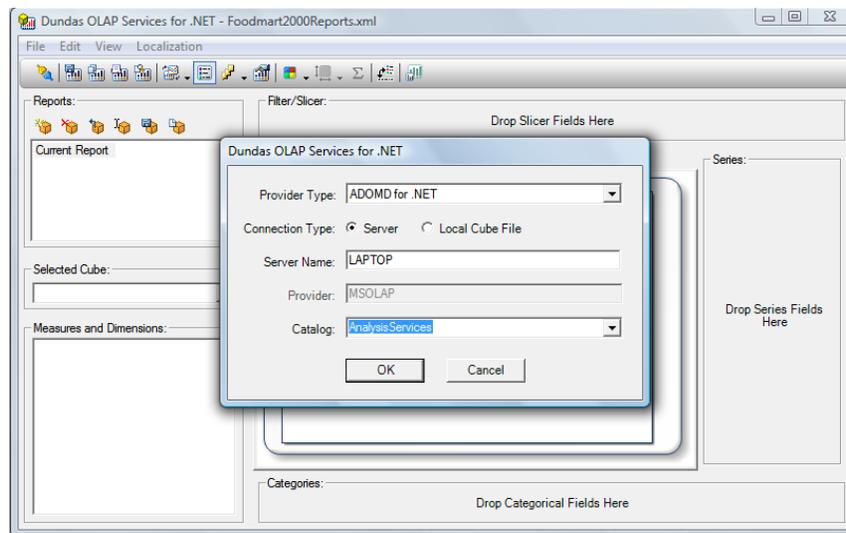


Figura 26. Interfaz de selección del cubo

Luego de establecer correctamente la conexión con el Catalogo, entramos a la interfaz donde podemos empezar a crear las consultas Ad Hoc sobre el cubo que seleccionemos.

En la herramienta se encuentra dividida en cuatro áreas principales (Ver Figura 27):

1. Seleccionar el cubo sobre el cual se van a realizar las consultas
2. Seleccionar las Medidas, KPI's o Dimensiones que pertenecen al cubo sobre el cual se está trabajando.

3. Área para crear las consultas Ad Hoc, que bien puede ser en la pestaña de Grilla o en la pestaña de Grafica.
4. Área para crear, abrir, guardar las consultas en un formato XML, entre otras tareas.

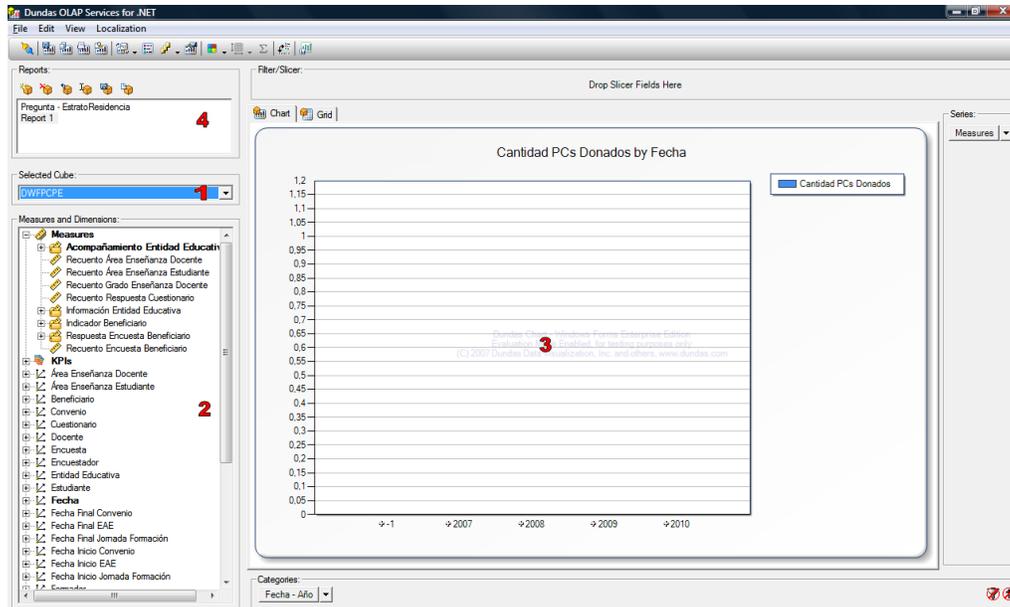


Figura 27. Área de trabajo Dundas OLAP Services

A continuación se presenta un ejemplo de la consulta realizado, mostrando el resultado de la consulta *Respuesta por Nivel Educativo*, donde se filtran las respuestas de la pregunta “¿Considera que con la integración de las TIC se fortalecen los procesos de enseñanza-aprendizaje?” por el Nivel Educativo de los docentes beneficiarios encuestados. En la siguiente figura podemos observar que en la mayoría de los niveles educativos la respuesta de los docentes fue afirmativa (88,89%), mientras que una minoría de docentes respondieron negativamente, especialmente los docentes del nivel educativo Normalista Superior (2,78%), y otro pequeño grupo que no contestaron a la pregunta (8,33%).

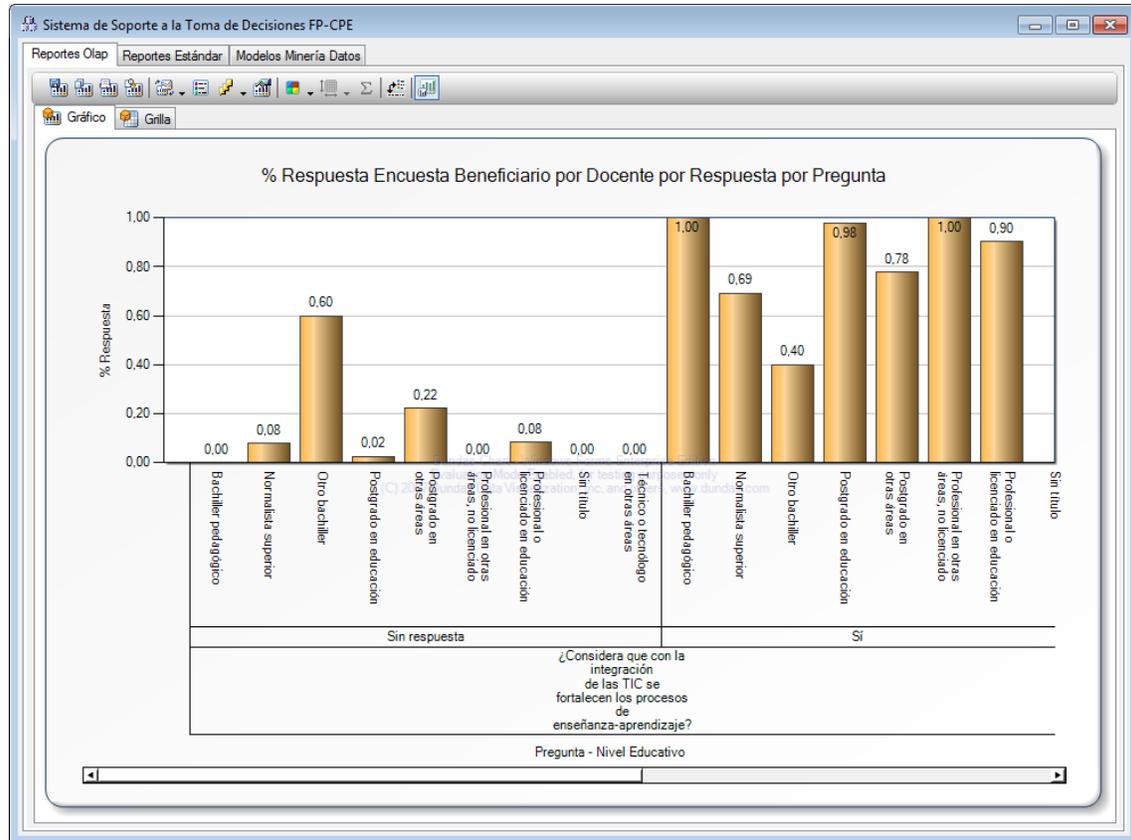


Figura 28. Consulta – Respuesta por Nivel Educativo (Grafica)

En la siguiente figura se presenta la visualización de la misma consulta comentada anteriormente, en la representación de Grilla.

			% Respuesta Encuesta Beneficiario
¿Considera que con la integración de las TIC se fortalecen los procesos de enseñanza-aprendizaje?	Sin respuesta	Normalista superior	7,69%
		Otro bachiller	60,00%
		Postgrado en educación	2,33%
		Postgrado en otras áreas	22,22%
		Profesional o licenciado en educación	8,20%
		Total	8,33%
	Sí	Bachiller pedagógico	100,00%
		Normalista superior	69,23%
		Otro bachiller	40,00%
		Postgrado en educación	97,67%
		Postgrado en otras áreas	77,78%
		Profesional en otras áreas, no licenciado	100,00%
	Profesional o licenciado en educación	90,16%	
Técnico o tecnólogo en otras áreas	100,00%		
Total	88,89%		
No	Normalista superior	23,08%	
	Profesional o licenciado en educación	1,64%	
Total	2,78%		
Total		100,00%	

Figura 29. Consulta – Respuesta por Nivel Educativo (Grilla)

En la siguiente figura se muestra un reporte en la vista avanzada, donde se aprecian los controles de selección de cubo, selección de dimensiones y medidas.

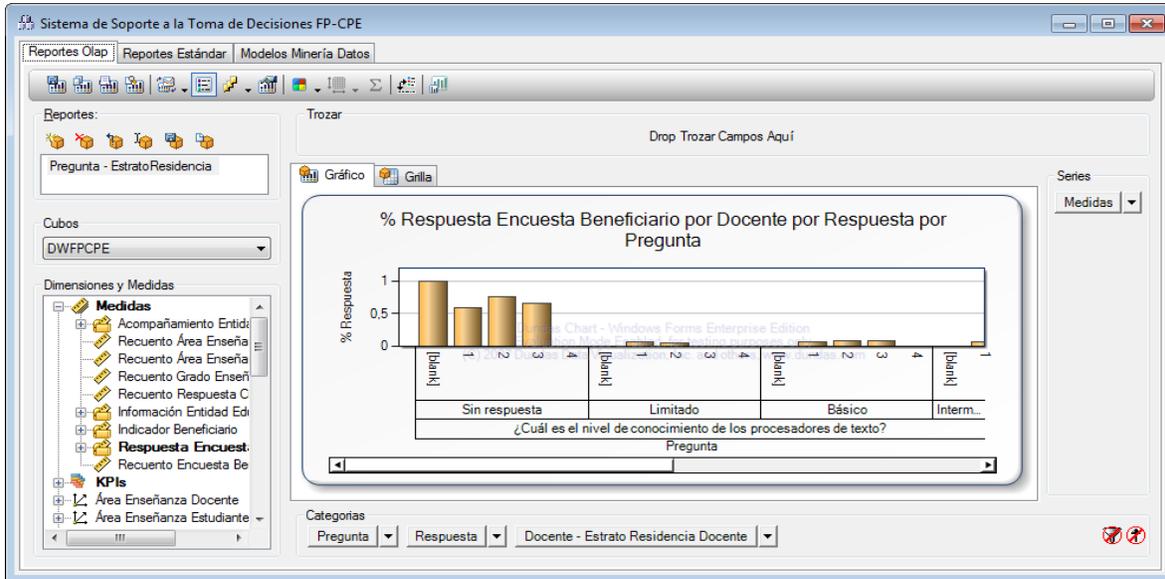


Figura 30. Vista Avanzada de Reportes OLAP

La herramienta permite presentar los reportes en diferentes tipos de graficas: barras, tortas, puntos de dispersión, entre otros.

En el **Anexo 31** se presentan otro tipo de reportes realizados, agrupando, relacionando y representando otro tipo de información importante dentro del cubo.

4.1.10. Despliegue del DW

Para poder realizar el despliegue del sistema fue necesario verificar que los productos estuvieran correctamente instalados, confirmar el buen funcionamiento del servidor (hardware, conexiones a bases de datos), correcto funcionamiento del servidor ETL, del servidor de análisis, el acceso a los datos, como también realizar capacitaciones a los usuarios administradores, hacer la entrega de manuales y documentación necesaria para el uso y manejo del sistema.

Todas estas actividades se realizan con los principales clientes directos de nuestro proyecto, en nuestro caso, los coordinadores y formadores del convenio CPE-Unicauca, instalando localmente el DSS para el uso efectivo del personal del convenio.

Como estrategia de transición a nivel nacional, para que el DSS se utilice por todos los convenios de CPE en Colombia, sería necesario implantar el sistema en un entorno Web, teniendo en cuenta además la adquisición de las licencias de las herramientas especificadas en etapas anteriores de la metodología.

Por otra parte, se debe tener en cuenta que para poder utilizar el DSS y realizar análisis en un convenio en particular del país, se debe realizar todo el proceso de recolección de información (encuestas para los beneficiarios, diligenciamiento de los formatos de las entidades educativas y de los formadores), y hacer el respectivo registro en el prototipo OLTP desarrollado, para luego actualizar la información de la bodega de datos.

Con el registro de información relacionada de mas convenios en el país, se podrán realizar análisis más interesantes al tener disponible información de diferentes convenios, lo que implica identificar aspectos interesantes de las distintas Estrategias de Acompañamiento Educativo utilizadas en cada uno de ellos, permitiendo mejorar cada proceso realizado por el programa CPE, obteniendo un impacto positivo sobre los beneficiarios de las Jornadas de Formación realizadas, ya que se mejoraran aspectos necesarios para que estos sean más competentes en el campo del uso de las TIC en la educación.

4.1.11. Mantenimiento y crecimiento

Las actividades de estas dos etapas (continuo soporte, el manejo de la infraestructura, mejoramiento del rendimiento sobre las consultas, mantenimiento de la metada y procesos ETL) están fuera del alcance de este proyecto.

4.1.12. Administración del proyecto

La respectiva administración del presente proyecto se realizó a lo largo de todo el desarrollo del sistema, contrastando los resultados obtenidos en cada etapa con el plan del proyecto establecido inicialmente. Se conto con la guía durante todo el proceso de desarrollo del proyecto del director, quien es experta en el tema de los DSS, además de contribuir con la verificación de que la construcción de la solución cumpliera con todas las fases necesarias para el desarrollo del sistema.

CAPITULO V. CASOS DE DISEÑO DE DW PARA EL MODELADO DIMENSIONAL DE HABILIDADES Y DESEMPEÑOS EN EL USO DE LAS TIC

Durante el proceso de modelamiento dimensional para la construcción de la bodega de datos dentro del contexto Educativo, en particular sobre las Competencias en TIC para Docentes y Estudiantes según los modelos planteados, se presentaron varios casos especiales para el modelado multidimensional de este proyecto.

A continuación se presentaran los casos de diseño, expuestos por Kimball [8], que se tuvieron en cuenta para el proceso de modelado dimensional realizado, exponiendo las situaciones encontradas y explicando su respectiva solución, aplicación o adaptación del respectivo caso en cuestión. Empezando con las relaciones de hecho, relaciones Muchos a Muchos, Relaciones Referenciadas, Jerarquías y terminando con el sistema de Auditoría modelado.

5.1. DIMENSION DE HECHO

En ciertas ocasiones los usuarios desean poder dimensionar las medidas según los elementos de datos que se encuentran en la tabla de hechos o realizar consultas en la tabla de hechos sobre determinada información relacionada adicional (números de factura o números de pedidos de compra relacionados con hechos de venta específicos). Cuando se define una dimensión basada en un elemento de tabla de hechos de este tipo, la dimensión se conoce como *dimensión de hechos* [19].

Las dimensiones de hechos también se denominan dimensiones *degeneradas*. Las dimensiones de hechos son útiles para agrupar filas de tablas de hechos relacionadas, como en el ejemplo presentado en [7], donde el número de factura nos permite agrupar hechos que están asociados a la transacción de compra hecha por un cliente. Aunque esta información puede colocarse en una tabla de dimensiones independiente de la base de datos relacional, crear una tabla de dimensiones independiente para la información no ofrece ninguna ventaja, ya que la tabla de dimensiones crecerá al mismo ritmo que la tabla de hechos, y simplemente crearía datos duplicados y una complejidad innecesaria.

Las dimensiones de hechos deben actualizarse de forma incremental después de cada actualización realizada en el grupo de medida al que hace referencia la relación de hechos [19].

Las relaciones de hechos presentan estas limitaciones:

- El atributo de granularidad debe encontrarse en la columna de clave de la dimensión, que crea una relación uno a uno entre la dimensión y los hechos de la tabla de hechos.
- Una dimensión puede tener una relación de hechos con un solo grupo de medida.

Dimensión de Hecho – Dimensión Encuesta

El caso de diseño Dimensión de Hecho se presento en el hecho encuesta beneficiario (Figura 31. Dimensión Encuesta) donde el identificador de la encuesta utilizado en el modelo de datos del negocio se convierte en la columna clave de la dimensión, generándose así la dimensión *Encuesta*.

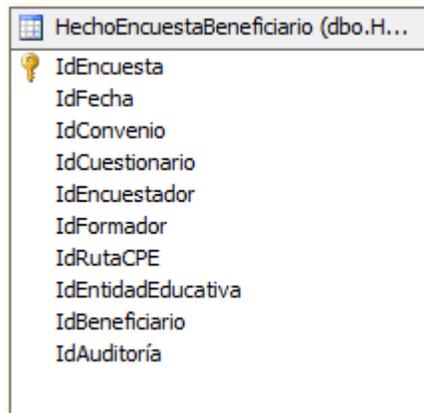


Figura 31. Dimensión Encuesta

Dimensión de Hecho – Dimensión Respuesta

Un caso diseño muy similar al caso de diseño Dimensión de Hecho se presentó en la dimensión Respuesta (Figura 32. Dimensión Respuesta), en este caso se parte de una dimensión la cual contiene atributos pero algunos de sus atributos se desean tomar como medidas.

En este caso el atributo *PesoRespuestaGrupo* de la dimensión respuesta representa el valor cuantitativo del peso de la respuesta con respecto a su grupo de respuestas, tomando los siguientes valores: 0, 1, 2, 3, 4 ó 5.

Así mismo el atributo *NombrePesoRespuestaGrupo* de la dimensión Respuesta representa su valor cualitativo (utilizado en las aplicaciones cliente), tomando respectivamente los siguientes valores: 'Nulo', 'Muy Bajo', 'Bajo', 'Medio', 'Alto' ó 'Muy Alto'.

El atributo *Peso Respuesta Grupo* de la dimensión Respuesta combinado con el atributo *Única Múltiple Respuesta* de la dimensión Pregunta permiten definir el valor cuantitativo de la respuesta, y este a su vez permite definir los valores cuantitativos de los niveles de la jerarquía: Indicador -> Pregunta -> Grupo Respuesta -> Respuesta.

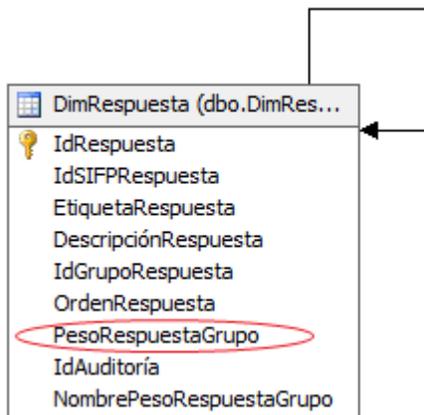


Figura 32. Dimensión Respuesta

Dimensión de Hecho – Dimensión Indicador

El caso de diseño presentado en la dimensión Respuesta se presenta en la dimensión indicador (Figura 33. Dimensión Indicador) ya que era necesario almacenar información adicional a sus atributos, el atributo *PesoIndicadorGrupo* se utiliza como medida para calcular el valor del peso ponderado en relación al grupo de indicadores, los cuales a su vez se utilizan para calcular los valores de los indicadores que articulan los modelos de Competencias en TIC definidos para estudiantes y docentes.

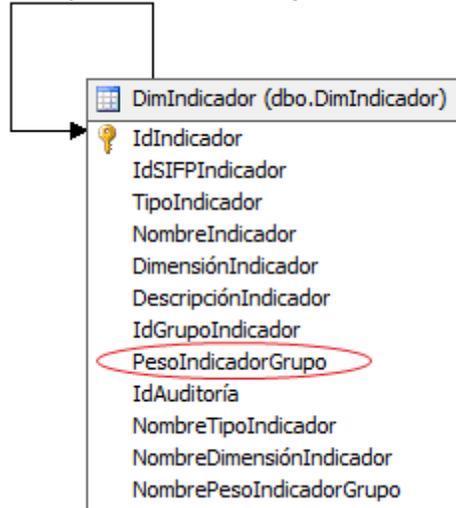


Figura 33. Dimensión Indicador

Dimensión de Hecho – Dimensión Entidad Educativa

Para la dimensión Entidad Educativa sucedió el mismo caso de *Dimensión de Hecho*, fue necesario almacenar dentro de ella medidas tales como la cantidad de computadores en red, cantidad de aulas de informática, cantidad de computadores para el uso administrativo y para el uso educativo, entre otras. Valores importantes para realizar un mejor análisis con relación a las Competencias, tenidos en cuenta como valores externos que pueden afectar directa o indirectamente a un docente o estudiante en ser o no competente en cierta dimensión (Pedagógica, Comunicativa y Colaborativa o Técnica y Tecnológica). (Ver Figura 34. Dimensión Entidad Educativa).

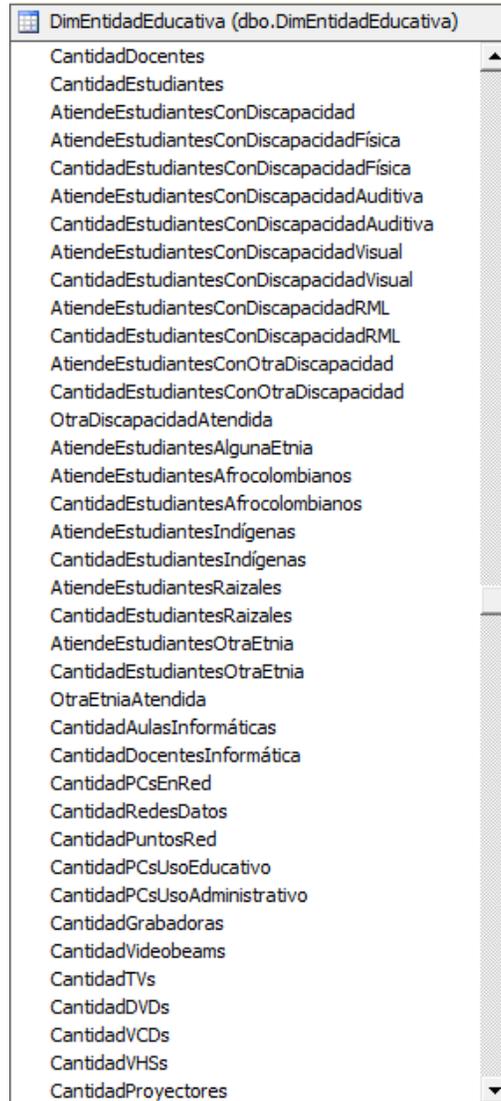


Figura 34. Dimensión Entidad Educativa

5.2. RELACIONES MUCHOS A MUCHOS

Generalmente cuando se define una dimensión, se relaciona cada hecho de la tabla de hechos con un único miembro de la dimensión, aunque un miembro de la dimensión puede estar relacionado con varios hechos diferentes de la tabla de hechos (relación uno a muchos). Sin embargo, existen ocasiones en el que un hecho puede estar relacionado con varios miembros de una dimensión (relación muchos a muchos), este mismo caso también suele suceder entre dimensiones, tal como sucedió en el desarrollo del modelado dimensional del proyecto, ampliando el modelo dimensional más allá del esquema estrella [19].

Para manejar este tipo de relaciones en Microsoft SQL Server Analysis Services, entre una dimensión y un grupo de medida u otra dimensión, se especifica una tabla de hechos intermedia que esta combinada con la dimensión, esta tabla intermedia es la que permite

establecer las relaciones de muchos a muchos entre los miembros de la dimensión y los hechos de la tabla de hechos (o los miembros de otra dimensión), comúnmente conocida como una tabla puente [19].

RELACION MUCHOS A MUCHOS ENTRE DIMENSIONES

Puente Dimensión Área de Enseñanza - Dimensión Docente

En el contexto del proyecto se encontró con la situación en la que un docente puede dictar clases en más de un área de enseñanza (ciencias naturales, matemáticas, etc.) y un área de enseñanza puede ser dictada por más de un docente, lo cual es el típico caso de relación muchos a muchos, para lo cual se optó por tomar la solución planteada para este tipo de relaciones de crear una tabla intermedia *PuenteÁreaEnseñanzaDocente* como se muestra a continuación en la **Figura 35**.

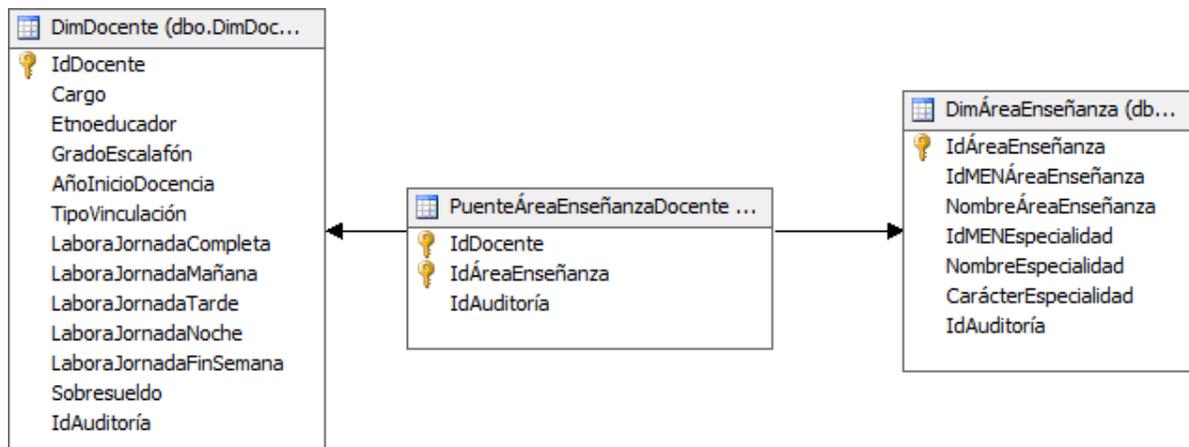


Figura 35. Puente Área de Enseñanza – Docente

Puente Dimensión Área Enseñanza - Dimensión Estudiante

Entre las dimensiones *Estudiante* y *Área de enseñanza* se presentó el mismo caso de relación muchos a muchos, el cual fue modelado con una tabla intermedia *PuenteÁreaEnseñanzaDocente* (ver **Figura 36**) entre las dimensiones, para manejar las relaciones de un estudiante con las Áreas de enseñanza que estudia, y los estudiantes que estudian un área de enseñanza en particular.

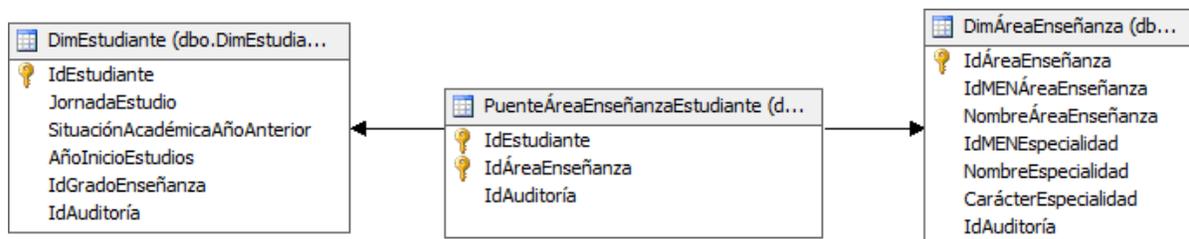


Figura 36. Puente Área de Enseñanza – Estudiante

Puente Dimensión Grado Enseñanza – Dimensión Docente

Este caso es muy similar al presentado anteriormente entre los Docentes y las Áreas de Enseñanza, ahora se maneja las relaciones entre un docente y los Grados de enseñanza que maneja (primero, tercer, noveno, once técnico, etc.), ya que un docente puede manejar varios Grados de enseñanza y un Grado de enseñanza puede ser manejado por varios Docentes, para manejar esta relación de muchos a muchos se creó la tabla intermedia *PuenteGradoEnseñanzaDocente* como se muestra a continuación en la **Figura 37**.

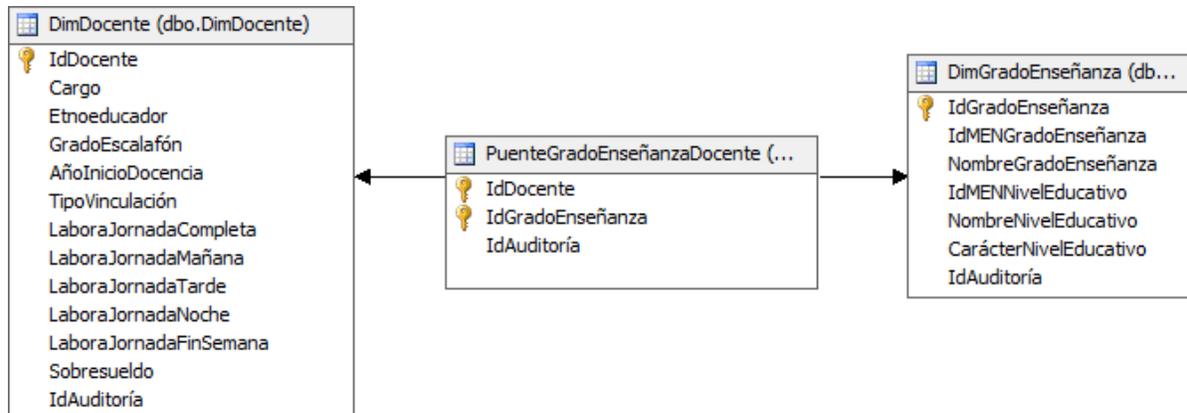


Figura 37. Puente Grado Enseñanza – Docente

Puente Dimensión Respuesta – Dimensión Cuestionario

En este caso se presentó una relación muchos a muchos entre tres dimensiones: *Respuesta*, *Pregunta* y *Cuestionario*. En la cual una pregunta tiene asociado unas posibles respuestas (primera relación muchos a muchos) y a su vez un cuestionario tiene asociado unas preguntas (segunda relación muchos a muchos). Dado que las respuestas, preguntas y cuestionarios son tratadas como entidades autónomas (dimensiones), estas entidades por si solas no tienen relación entre ellas, por lo tanto fue necesario el diseño de una tabla intermedia *PuenteRespuestaCuestionario* (ver **Figura 38**) donde se maneje y almacene las relaciones históricas entre estas entidades.

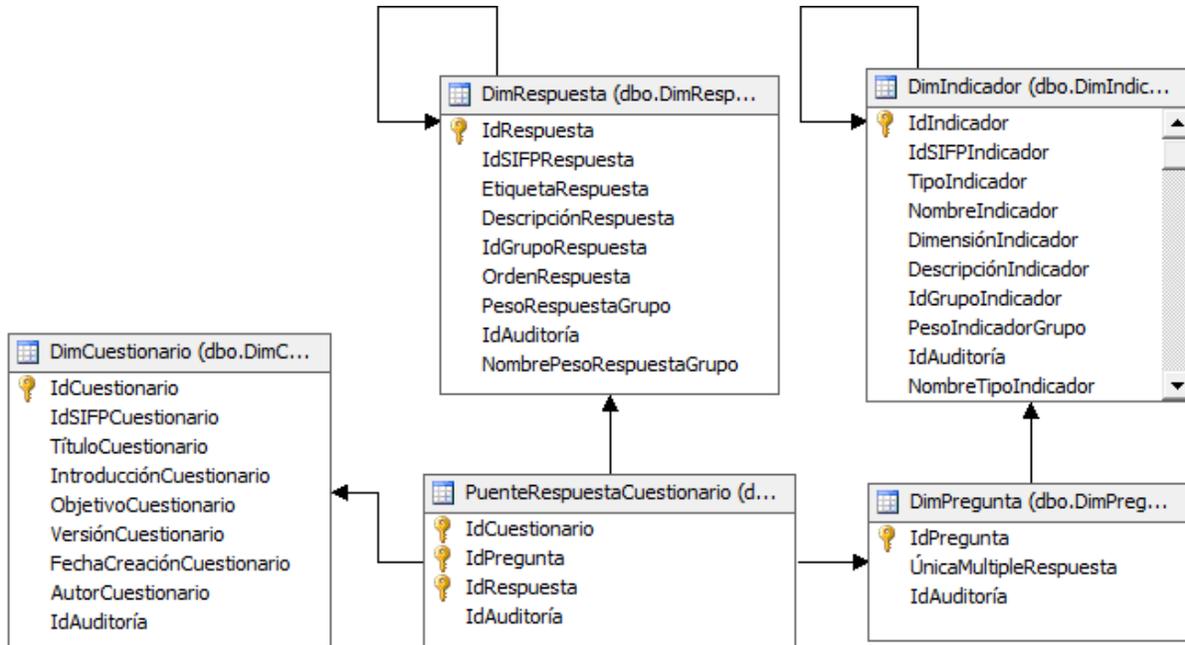


Figura 38. Puente Respuesta – Cuestionario

Dado que sobre el hecho *HechoRespuestaEncuestaBeneficiario* solo se registran las respuestas seleccionadas de una pregunta, a partir de éste puente se pueden identificar todas las posibles respuestas definidas para una pregunta, gran utilidad a la hora de visualizar la información en las aplicaciones cliente.

RELACION MUCHOS A MUCHOS ENTRE TABLA DE HECHOS - DIMENSIÓN

Puente Dimensión Encuesta – Dimensión Pregunta - Dimensión Respuesta

Entre las tablas de hechos *HechoRespuestaEncuestaBeneficiario* y *HechoEncuestaBeneficiario*, existe una relación de tipo Muchos a Muchos, ya que las respuestas se encuentran asociadas a las preguntas de una encuesta. En este caso la tabla de hechos *HechoRespuestaEncuestaBeneficiario* (ver **Figura 39**) actúa como puente entre las dimensiones encuesta, pregunta y respuesta.

Conceptualmente Kimball propone [8], crear un puente entre la dimensión y la tabla de hechos, la cual tendrá el atributo clave de la dimensión y la clave alternativa de la tabla de hechos. Aunque en esencia un registro de la tabla de hechos se identifica por todas las claves de las dimensiones que se relacionan al hecho, este puede ser identificado por las dimensiones o claves del modelo del negocio que identifican más al hecho. En nuestro caso al hecho *HechoEncuestaBeneficiario* lo identifica la clave de la encuesta utilizado en el Sistema de Información de la Fase de Profundización (SIFP).

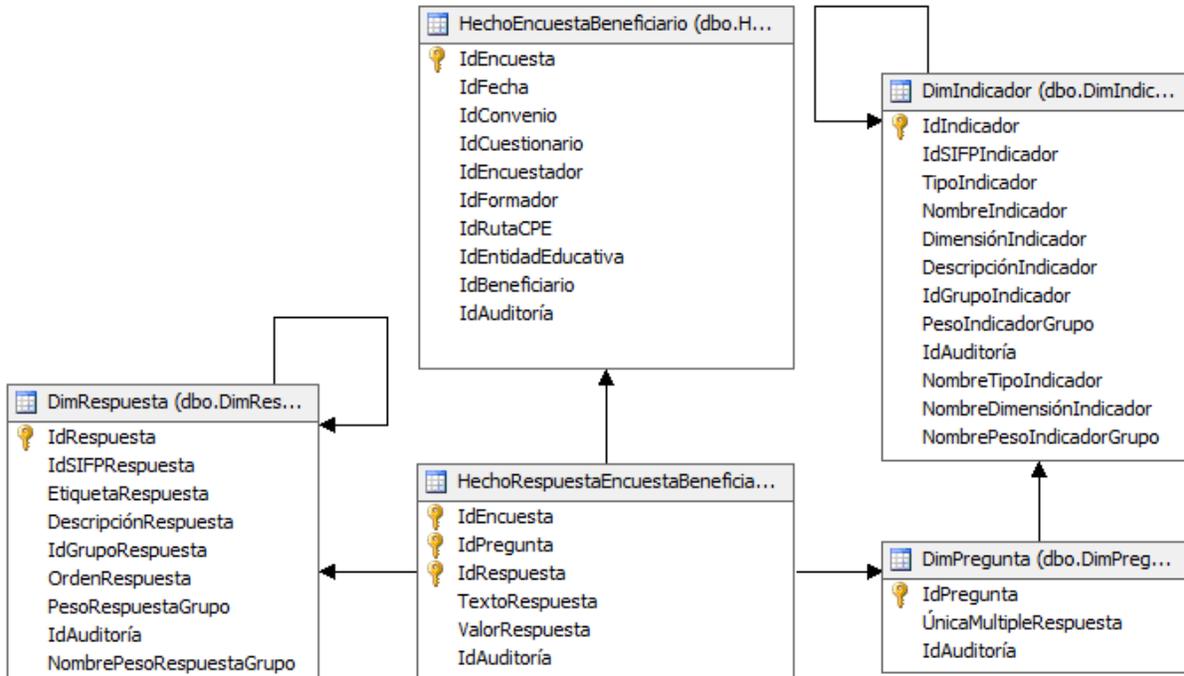


Figura 39. Puente Respuesta – Pregunta - Encuesta

Puente Dimensión Encuesta – Dimensión Indicador

Igual al caso presentado entre las respuestas de una encuesta, entre los indicadores y las encuestas, existe una relación de tipo Muchos a Muchos, ya que un Indicador puede estar relacionado con varias encuestas, y una encuesta puede relacionarse con varios indicadores, para lo cual se creó una tabla intermedia *HechoIndicadorBeneficiario* como se muestra a continuación en la Figura 40.

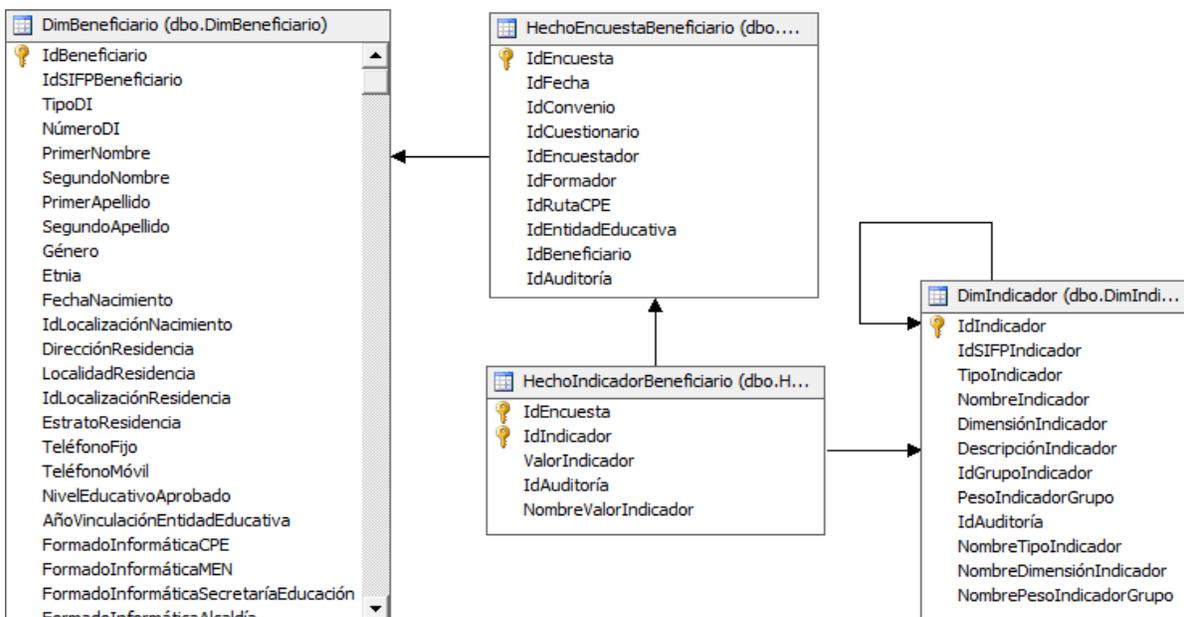


Figura 40. Puente Indicador - Encuesta

5.3. SUBDIMENSION

La mayoría de las dimensiones se relacionan directamente con la tabla de hechos para un grupo de medida, utilizando la clave principal de la dimensión en la tabla de hechos, pero en algunos casos es necesario relacionar una dimensión A (Subdimensión) con la tabla de hechos por medio de una dimensión intermedia B (Dimensión de referencia), de esta manera se podrá dimensionar los hechos de la tabla de hechos por medio de la dimensión A, por ejemplo: tenemos la tabla de hechos Ventas y la dimensión Distribuidor relacionada directamente con la tabla de hechos, si incluimos la dimensión Geografía, relacionándola con la tabla de hechos por medio de la dimensión Distribuidor, podremos dimensionar las ventas de distribuidor por geografía, las subdimensiones son conocidas en [19] como relación de referencia.

La relación de dimensión de referencia se define al especificar lo siguiente:

- La dimensión intermedia con la que se va a combinar. Puede ser una dimensión normal u otra dimensión de referencia.
- Un atributo de dimensión de referencia que defina el nivel mínimo en el que la dimensión está disponible para la agregación con respecto al grupo de medida.
- El atributo (clave externa) de la dimensión intermedia que corresponde al atributo de dimensión de referencia.

Subdimensiones del Hecho Encuesta Beneficiario

Para la tabla de hechos *HechoEncuestaBeneficiario* se establecieron las relaciones de Referencia para las dimensiones Docente y Estudiante, utilizando como dimensión Intermedia la dimensión Beneficiario, como se muestra en la **Figura 41** de esta manera es posible realizar consultas sobre los hechos de las encuestas hechas para los docentes en particular o para las estudiantes, tal como se explicaba anteriormente.

En el modelado dimensional se representa una generalización de la dimensión Beneficiario hacia las dimensiones Docente y Estudiante, estando en beneficiario toda la información común existente entre un docente y un estudiante; este tipo de relaciones no tiene representación puesto que no se tienen en cuenta en los modelos dimensionales, ya que este tipo de relaciones se realizan para disminuir la redundancia de datos, lo cual no es relevante en modelos dimensionales.

Se represento la generalización identificada por medio de la utilización jerarquías, en la siguiente figura se puede apreciar que en la dimensión Beneficiario se encuentra un atributo *TipoBeneficiario*, el cual identifica los registros de información que pertenecen a un docente o a un estudiante, gracias a este tipo de relación las dimensiones Docente y Estudiante se relacionan con las tablas de hecho y las demás dimensiones a través de la dimensión Beneficiario.

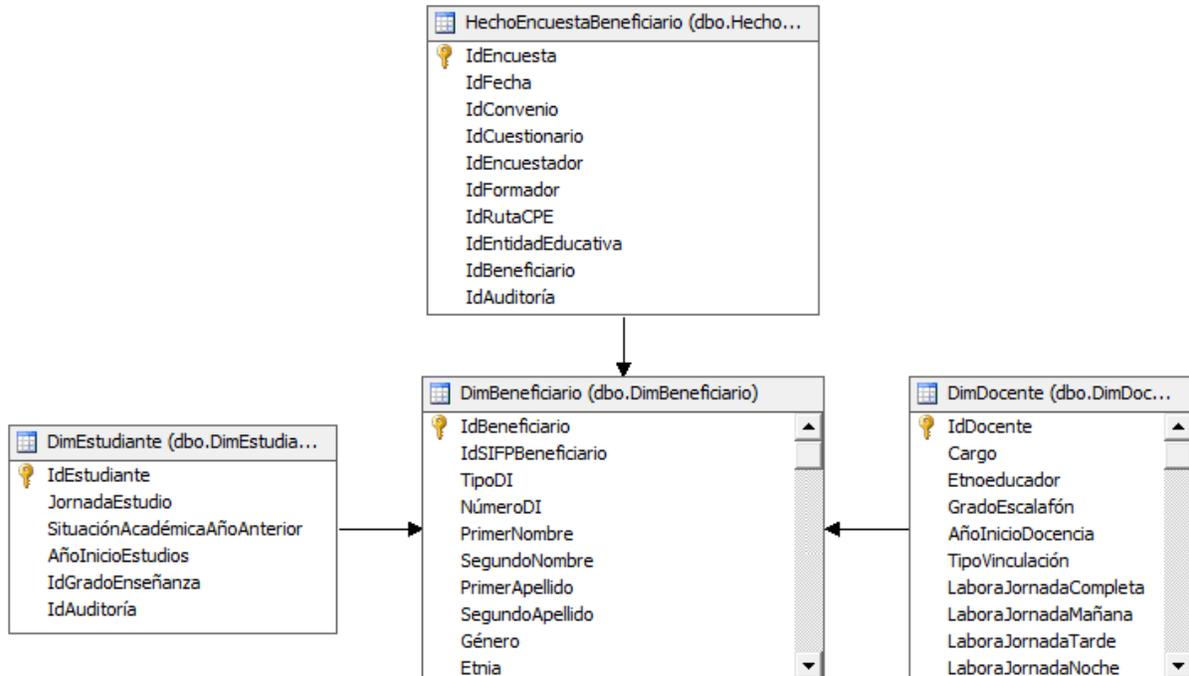


Figura 41. Subdimensiones Docente y Estudiante

Subdimensiones de la Dimensión Localización

La dimensión Localización tiene 6 roles (Localización residencia Beneficiario, localización nacimiento beneficiario, localización Entidad educativa, localización residencia empleado CPE, localización nacimiento empleado, localización universidad), para poder de esta manera dimensionar por localización los hechos de las tablas de hechos que están relacionadas con las dimensiones involucradas en los roles de la dimensión Localización, mencionados anteriormente (Ver Figura 42).

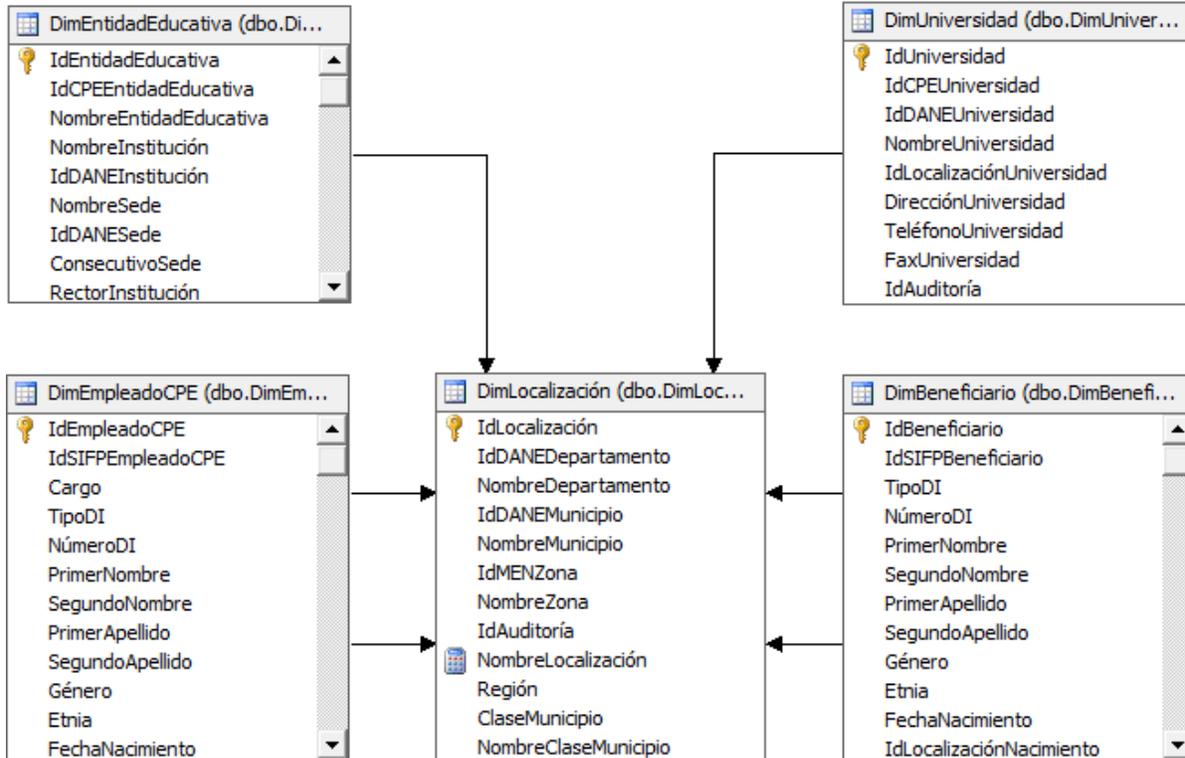


Figura 42. Subdimensiones de la Dimensión Localización

JERARQUÍAS DE ORGANIZACIÓN

Las jerarquías de Organización o de elementos primarios y secundarios, como se conocen en [19], estas dimensiones se definen mediante un atributo especial, llamado *atributo primario*, para determinar cómo se relacionan entre sí los miembros. Un *atributo primario* describe una *relación que hace referencia a sí misma* o una *autocombinación* dentro de una tabla principal de dimensiones. Las jerarquías de elementos primarios y secundarios se construyen a partir de un único *atributo primario*. A una jerarquía de elementos primarios y secundarios sólo se le asigna un nivel, puesto que los niveles presentes en la jerarquía se extraen de las relaciones de elementos primarios y secundarios entre los miembros asociados al atributo primario. El esquema de dimensiones de una jerarquía de elementos primarios y secundarios depende de la relación que hace referencia a sí misma presente en la tabla principal de dimensiones.

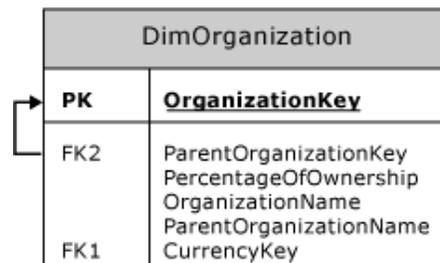


Figura 43. Dimensión Organization, tomada de [19]

En la **Figura 43** se muestra la dimensión *Organization*, la columna *ParentOrganizationKey* tiene una relación de clave externa con la columna de clave principal *OrganizationKey*. En otras palabras, cada registro de esta tabla puede relacionarse a través de una relación de elementos primarios y secundarios con otro registro de la tabla. Este tipo de *autocombinación* se utiliza generalmente para representar los datos de entidad de la organización, como la estructura de administración de los empleados de un departamento.

Cuando se crea una jerarquía de elementos primarios y secundarios, las columnas representadas por ambos atributos deben tener el mismo tipo de datos. Además, los dos atributos deben estar en la misma tabla. De forma predeterminada, se considera que cualquier miembro cuya clave principal sea igual a su propia clave de miembro, valor NULL, 0 (cero) o un valor no presente en la columna de las claves de miembro es un miembro del nivel superior (excepto el nivel (All)).

La profundidad de una jerarquía de elementos primarios y secundarios puede variar entre sus ramas jerárquicas. En otras palabras, una jerarquía de elementos primarios y secundarios se considera una jerarquía desequilibrada (las ramas descienden hasta niveles diferentes).

A diferencia de las jerarquías definidas por el usuario, en las que el número de niveles de la jerarquía determina el número de niveles que se mostrarán a los usuarios finales, una jerarquía de elementos primarios y secundarios se define con el único nivel de una jerarquía de atributo, y los valores de este único nivel generan los diversos niveles que ven los usuarios finales. El número de niveles mostrados depende del contenido de las columnas de la tabla de dimensiones, que almacenan las claves de miembro y las claves principales. El número de niveles puede variar cuando cambian los datos de las tablas de dimensiones [19].

Jerarquía – Dimensión Indicador

Esta jerarquía es la representación del modelo de Competencias en TIC para los docentes y estudiantes beneficiarios por el programa CPE, la cual es formada por la relación recursiva que existe en esta dimensión, estableciendo las relaciones entre los diferentes niveles de la jerarquía (Competencia, Desempeño, Indicador, Pregunta), se puede observar en la **Figura 44** el atributo *Tipo Indicador* es el que nos identifica el nivel de la jerarquía en el que estamos o si estamos hablando de una Competencia, un Desempeño, un Indicador o una Pregunta. También se aprecia el atributo *Peso Indicador Grupo*, el cual es un valor importante para calcular el valor del peso ponderado del indicador sobre el grupo de indicadores como se explico anteriormente.



Figura 44. Jerarquía – Dimensión Indicador

Jerarquía – Dimensión Respuesta

Esta jerarquía agrupa ordenadamente para su correcta exploración todas las posibles respuestas que puede realizar un docente o un estudiante encuestado, esta jerarquía está formada por la relación recursiva que existe en esta dimensión, estableciendo las relaciones entre los diferentes niveles de la jerarquía: Grupo Respuesta, Respuesta, Respuesta 1 (Ejemplo de datos en los niveles de esta jerarquía, ver **Figura 39**). También se aprecia el atributo *Peso Respuesta Grupo*, el cual es un valor importante para calcular el valor del peso ponderado de la respuesta sobre el grupo de respuestas como se explicó anteriormente (Ver **Figura 45**).

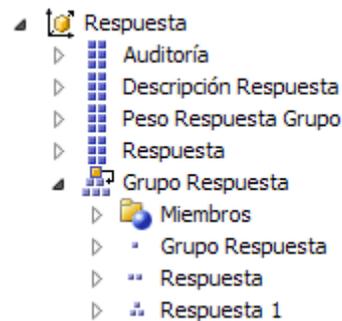


Figura 45. Jerarquía – Dimensión Respuesta



Figura 46. Ejemplo niveles Jerarquía – Dimensión Respuesta

INCONVENIENTES PRESENTADOS EN EL DISEÑO

En la implementación se presentaron inconvenientes para realizar la medición de las competencias en TIC, puesto que la herramienta se limita a realizar agregaciones sencillas que no permitían obtener los resultados necesarios para dicha medición, por lo anterior, fue necesario programar mediante el lenguaje MDX los comportamientos y operaciones requeridas utilizando un miembro calculado para el valor del peso ponderado en cada nivel de la jerarquía (Competencia, Desempeño, Indicador).

Dado a la complejidad de procesamiento para obtener el valor cuantitativo de respuesta se decidió hacer persistencia de este cálculo en la medida *Valor Respuesta* de la tabla de hechos Encuesta Beneficiario. A continuación se presenta el código MDX creado para calcular un valor porcentual de la cantidad de respuestas de una pregunta en particular:

```
CREATE MEMBER CURRENTCUBE.[MEASURES].[% Respuesta Encuesta Beneficiario]
AS IIF
(
    IsEmpty
    (
        [Measures].[Recuento Respuesta Cuestionario]
    ),
    NULL,
    CoalesceEmpty
    (
        (
            [Respuesta].[Grupo Respuesta].CurrentMember,
            [Measures].[Recuento Respuesta Encuesta Beneficiario]
        )
        /
        (
            [Respuesta].[Grupo Respuesta].[All].[All],
            [Measures].[Recuento Respuesta Encuesta Beneficiario]
        )
    )
)
```

```
    )  
    ;  
    0  
  )  
) ,  
FORMAT_STRING = "Percent";
```

El siguiente código MDX es el encargado de los controlar los comportamientos para los valores de los beneficiarios obtenidos por cada beneficiario encuestado:

```
CREATE MEMBER CURRENTCUBE.[MEASURES].[Valor Beneficiario]  
AS Case  
When [Encuesta].[Encuesta].CurrentMember.Level Is  
    [Encuesta].[Encuesta].[All]  
Then  
    Avg  
    (  
        Descendants  
        (  
            [Encuesta].[Encuesta].CurrentMember  
        ),  
        [Measures].[Valor Beneficiario]  
    )  
Else  
    Case  
    When [Indicador].[Grupo Indicador].CurrentMember.Level Is  
        [Indicador].[Grupo Indicador].[All]  
    Then  
        Avg  
        (  
            Descendants  
            (  
                [Indicador].[Grupo Indicador].CurrentMember  
            ),  
            [Measures].[Valor Beneficiario]  
        )  
    Else  
        (  
            [Indicador].[Grupo Indicador].CurrentMember.DataMember,  
            [Measures].[Valor Indicador]  
        )  
    End  
End,  
FORMAT_STRING = "#,0.000";
```

5.4. DIMENSION AUDITORIA

Un aspecto que vale resaltar dentro del proceso ETL es la construcción del sistema de auditoría, teniendo en cuenta el sistema propuesto por Ralph Kimball [8], el cual está compuesto de tres tablas, para poder realizar el seguimiento de los paquetes y los procesos que cargan los datos en las dimensiones y tablas de hechos.

La dimensión Auditoria en el DMSS registra la cantidad de filas extraídas, insertadas y actualizadas de todas las dimensiones y tablas de hechos de la bodega de datos. En la **Figura 47** se muestra el sistema desarrollado para llevar a cabo la auditoria para el DSS.

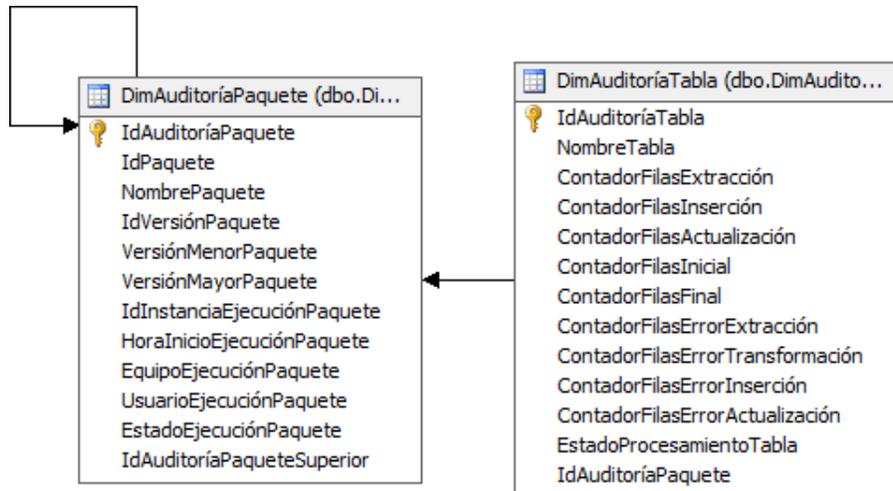


Figura 47. Sistema del Auditoría del DW

CAPITULO VI. DESCRIPCIÓN DEL PROCESO DE DESARROLLO DEL MODULO DE MINERIA DE DATOS

Para el desarrollo del módulo de DM se adoptó como base la metodología CRISP-DM [9]. Esta metodología presenta el ciclo de vida del proyecto de DM dividida en seis fases. La **Figura 48** se ilustra las diferentes fases por las que debe pasar todo proceso de DM.

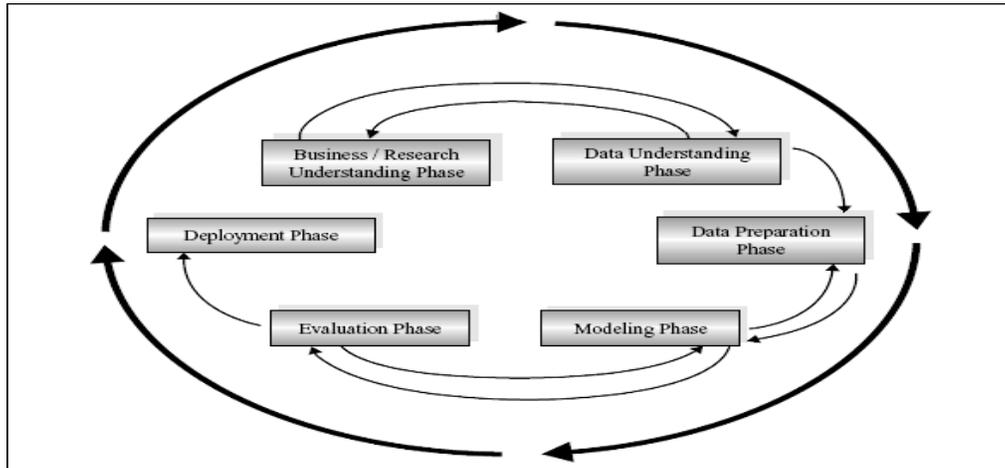


Figura 48. Fases del modelo de referencia de CRISP-DM, tomado de [9].

A continuación se describe la secuencia (no rígida) de las fases de esta metodología [9] que permitieron el desarrollo del presente módulo de minería de datos:

6.1. Entendimiento del negocio

Esta fase inicial se enfoca en el entendimiento de los objetivos y requerimientos del proyecto desde una perspectiva del negocio, para luego convertir este conocimiento en una definición del problema de DM y diseñar un plan preliminar para conseguir los objetivos [9].

La idea de incorporar minería de datos en el DSS se presentó con el planeamiento inicial del proyecto como mecanismo para poder descubrir nuevas correlaciones, patrones y/o tendencias ocultas en la información almacenada en el DW, que permitan a los directivos de los convenios de las universidades y CPE, realizar análisis más profundos sobre el proceso que están realizando, en particular la Fase de Profundización, para poder de esta manera mejorar los resultados obtenidos con la ejecución de las diferentes actividades relacionadas con la fase mencionada.

Con los resultados obtenidos por medio del proceso de minería de datos, se podrán tomar decisiones que impliquen grandes beneficios sobre el proceso ejecutado por las universidades, consiguiendo mejores beneficios para las entidades educativas, docentes y estudiantes, y un mejor rendimiento en la cobertura, capacitación, acompañamiento, instalación de equipos y demás actividades, según las correlaciones encontradas entre la información disponible en el DW.

De acuerdo a lo anterior en esta fase se determinó el objetivo del negocio; se evaluaron las situaciones; se determinaron los objetivos de la minería de datos y se elaboró el plan del proyecto.

Objetivo del Negocio

Modelar y construir un módulo de DM, que permita realizar un proceso de descubrimiento de nuevas correlaciones, patrones y/o tendencias ocultas en la información almacenada en el DW.

Objetivo de Minería de Datos

1. Realizar una clasificación adecuada de los indicadores del modelo de competencias de los docentes, tomando como base la información relacionada con la formación del docente brindada por CPE (Docente, Entidad Educativa, Formador, Universidad)
2. Realizar una clasificación adecuada de los indicadores del modelo de competencias de los estudiantes, tomando como base la información relacionada con la formación de sus docentes brindada por CPE (Estudiante, Entidad Educativa, Formador, Universidad)
3. Obtener las reglas que permitan descubrir la relación entre los indicadores del modelo de competencias en TIC de los docentes y la información relacionada con la formación del docente brindada por CPE (Estudiante, Entidad Educativa, Formador, Universidad)
4. Obtener las reglas que permitan descubrir la relación entre los indicadores del modelo de competencias en TIC de los estudiantes y la información relacionada con la formación de sus docentes brindada por CPE (Docente, Entidad Educativa, Formador, Universidad)

6.2. Entendimiento de datos

La fase de entendimiento de datos empieza con una colección de datos inicial y prosigue con las actividades para familiarizarse con los datos, identificar los problemas de calidad de datos, descubrir los primeros discernimientos en los datos o detectar subconjuntos interesantes y para formar hipótesis sobre la información oculta [9].

Sobre el DW se almacena información relacionada con la fase de profundización (formadores, convenios, universidades, etc.), los beneficiarios (docentes y estudiantes), entidades educativas, el diligenciamiento de las encuestas por parte de los beneficiarios, los modelos de competencias en TIC definidas por CPE; todos estos datos se encuentran representados en las dimensiones y hechos del DW, información indispensable para realizar el proceso de descubrir co-relaciones, patrones y/o tendencias ocultas.

Dado que el modelo de minería de datos se basará de los datos almacenados en el DW no será necesario detallar el proceso de recolección, ya que éste se explicó previamente.

Dado que algunos datos de las entidades no fueron suministrados en su totalidad por CPE se determinó que la información relacionada con las jornadas ofrecidas, sub-modalidades ofrecidas y población atendida no se tomará en cuenta para la generación de los modelos de minería de datos en presente documento, además no se tomarán en cuenta los datos de localización de la universidad y localización de residencia del

formador porque por el momento el DW no almacena información de otros convenios, sin embargo estos datos quedarán definidos en la estructura de los modelos de minería.

A continuación en la

Tabla 9 se define la estructura genérica de minería de datos, en el Anexo 32 se detallan cada una de las variables relacionadas y sus posibles valores.

Tabla 9. Definición Estructura Genérica de Minería de Datos

Dimensión	Clasificación de Variables
Beneficiario	Demografía
	Localización Residencia
	Formación en Informática
	Disponibilidad Recursos Tecnológicos
	Educación
Docente	Educación
	Información Laboral
Estudiante	Educación
Entidad Educativa	Disponibilidad Recursos Tecnológicos
	Localización
	Sub-modalidades Ofrecidas
	Población Atendida
Formador	Jornadas Ofrecidas
	Demografía
	Localización Residencia
	Educación
	Ruta CPE
Universidad	Localización
Indicador	Indicador
	Valor Indicador

Nota: La información del Formador, la Universidad (Entidad Formadora) y Ruta CPE corresponde a la información de la Estrategia de Acompañamiento Educativo.

6.3. Preparación de datos

La fase de preparación de datos cubre todas las actividades para construir el conjunto de datos final sin procesamiento inicial. Las tareas de preparación de datos son probablemente para ser realizadas muchas veces y no en cualquier orden prescrito [9].

Dado que los datos base para el modelo de minería de datos se encuentran en el DW, la limpieza y formateada de los datos no será necesario ya éste se realizó previamente en el proceso de cargue (ETL) del DW.

En esta fase se diseñaron cuatro estructuras de datos que integran la información relacionada con la formación brindada por CPE a los docentes. En cada una de éstas estructuras se analizará cada uno de los objetivos de minería definidos previamente. A continuación en la Tabla 10 se presentan las estructuras definidas para cada uno de los objetivos.

Tabla 10. Estructuras de minería de datos para cada objetivo de minería de datos

No	Estructura DM	Objetivos de DM
1	Indicador Docente	Realizar una clasificación adecuada de los indicador del modelo de competencias de docentes, tomando como base la información relacionada con la formación brindada por CPE a los docentes (Docente, Entidad Educativa, Formador, Universidad)
2	Indicador Estudiante	Realizar una clasificación adecuada de los indicador del modelo de competencias de estudiantes, tomando como base información relacionada con la formación brindada por CPE a sus docentes (Estudiante, Entidad Educativa, Formador, Universidad)
3	Indicadores Docente	Obtener las reglas que permitan descubrir la relación entre los indicadores del modelo de competencias en TIC de los docentes y la información relacionada con la formación brindada por CPE a los docentes (Estudiante, Entidad Educativa, Formador, Universidad)
4	Indicadores Estudiante	Obtener las reglas que permitan descubrir la relación entre los indicadores del modelo de competencias en TIC de los estudiantes y la información relacionada con la formación brindada por CPE a sus docentes (Docente, Entidad Educativa, Formador, Universidad)

Se construyeron los modelos de datos que integran la información detallada de cada una de las dimensiones (Beneficiario, Entidad Educativa, Formador, Universidad e Indicador) que intervienen en cada una de las estructuras de minería de datos, en el **Anexo 33** se detalla el modelo de datos de cada una de las estructuras de minería de datos. A continuación en la Tabla 11 se presentan la clasificación de las variables de entrada o de predicción que tuvieron en cuenta en cada una de las estructuras de minería de datos.

Tabla 11. Estructuras de minería de datos, (E = Variable de Entrada, P = Variable de Predicción, X = No aplica)

Clasificación de Variables		Estructuras DM			
		1	2	3	4
Beneficiario	Demografía	E	E	E	E
	Localización Residencia	E	E	E	E
	Formación en Informática	E	E	E	E
	Disponibilidad Recursos Tecnológicos	E	E	E	E
	Educación	E	E	E	E
Docente	Educación	E	E	E	E
	Información Laboral	E	X	E	X
Estudiante	Educación	X	E	X	E
Entidad Educativa	Disponibilidad Recursos Tecnológicos	E	E	E	E
	Localización	E	E	E	E
	Sub-modalidad Ofrecidas	E	E	E	E
	Población Atendida	E	E	E	E
	Jornadas Ofrecidas	E	E	E	E

SISTEMA DE SOPORTE A LA TOMA DE DECISIONES EN EL MARCO DE COMPUTADORES PARA EDUCAR
(CPE) EN LA FASE DE PROFUNDIZACIÓN (FP)

Formador	Demografía	E	E	E	E
	Localización Residencia	E	E	E	E
	Educación	E	E	E	E
	Ruta CPE	E	E	E	E
Universidad	Localización	E	E	E	E
Indicador	Indicador	X	X	P*	P*
	Valor Indicador	P	P	P*	P*

La información de los indicadores de las estructuras 3 y 4 se filtran a los indicadores que tenía el nivel más bajo de la jerarquía Competencia->Desempeño->Indicador1->Indicador2->..., ya que existe una relación directa entre el indicador y su indicador padre, porque su valor es calculado en base al valor de su indicador hijo.

Dado que las estructuras de datos deben ser analizadas en contextos diferentes (por modelo de competencias⁸ o por indicadores de un modelo de competencias) se diseñó un sistema para poder configurar y parametrizar cada una de las estructuras de datos, permitiendo así realizar análisis por cada uno de estos contextos. A continuación en la **Figura 49** se muestra la tabla que permite realizar esta configuración y parametrización.

EstructuraMinería (dmo.EstructuraMinería)	
🔑	ID
	Nombre
	IdSIFPTipoBeneficiario
	IdCuestionario
	IdIndicador
	RequiereCuestionario
	RequiereIndicador

Figura 49. Configuración de Estructura de Minería de Datos

6.4. Modelado

En esta fase varias técnicas de modelado son seleccionadas, aplicadas y sus parámetros son calibrados para valores óptimos. Básicamente, hay algunas técnicas para el mismo tipo de problema de DM, otras tienen requerimientos específicos sobre la forma de los datos, por lo tanto, devolverse a la fase de preparación de datos es a menudo necesario [9].

En esta fase se seleccionó la herramienta para la minería de datos, se seleccionaron las técnicas de modelado que se aplicarán a cada uno de los objetivos de minería de datos y se implementó y validaron cada uno de los modelos propuestos.

Herramienta para la minería de Datos

Se seleccionó la herramienta Microsoft SQL Server 2005 que incorpora la herramienta SQL Analysis Server, la cual establece facilidades para desarrollar el módulo de minería de datos del proyecto, entre las facilidades que se cuentan [21]:

- El procesamiento de los modelos de una misma estructura de minería ocurre en paralelo, en una sola lectura de los datos.

⁸ Los modelos de competencias son articulados por las preguntas definidos para cada cuestionario
 SISTEMA DE SOPORTE A LA TOMA DE DECISIONES EN EL MARCO DE COMPUTADORES PARA EDUCAR
 (CPE) EN LA FASE DE PROFUNDIZACIÓN (FP)

- Proporciona más de 12 visores de resultados para los algoritmos que ayudarán a comprender mejor los patrones encontrados en el proceso de minería.
- Proporciona gráficos de elevación, de beneficios y una matriz de clasificación que permite establecer una comparación de lo real con lo previsto; para contrastar y comparar la calidad de los modelos.
- Posee un lenguaje para la creación de consultas de minería (DMX) similar al SQL que facilita la tarea de creación de aplicaciones de minería de datos.
- Posee una interfaz gráfica para generar las consultas DMX.
- Cuenta con los algoritmos de minería más avanzados: Naive Bayes, Clustering, Clústeres de Secuencia, Árboles de Decisión, Redes Neuronales, Series Temporales, Reglas de Asociación, Regresión Logística, y Regresión Lineal y minería de textos.
- Marco de desarrollo para agregar nuevos algoritmos y también para construir visores propios para los modelos generados [22].
- Existe documentación y ejemplos de guía para creación de aplicaciones de minería de datos [23].

Los algoritmos disponibles en la herramienta Analysis Services de Microsoft para minería de datos (híbridos) que se seleccionaron para aplicar a cada una de las estructuras son:

- Clasificación:
 - Algoritmo árboles de decisión: algoritmo híbrido que incorpora distintos métodos para crear un árbol, y que admite varias tareas de análisis (regresión, clasificación y asociación), permite además el modelado de atributos discretos y continuos; nos ofrece elementos como: Visor de árboles de decisión de Microsoft, Red de dependencia, Gráfico de elevación y Matriz de Clasificación.
 - Algoritmo de clusteres: este algoritmo proporciona dos métodos para crear clústeres y asignar puntos de datos a dichos clústeres, el algoritmo mediana-K (agrupación de clusteres en duro) y el método Expectation Maximization (agrupación de métodos en blando), nos ofrece elementos de Visor de clústeres de Microsoft, Diagrama del clúster, Perfiles del clúster, Características del clúster y Distinción del clúster.
- Asociación:
 - Algoritmo Reglas de Asociación: implementación del algoritmo Apriori, este genera y luego cuenta conjuntos de elementos candidatos (un elemento puede representar un evento, un producto o el valor de un atributo, dependiendo del tipo de datos que se analice), nos ofrece elementos como: Visor de árboles de decisión de Microsoft, Conjuntos de Elementos, Reglas y Red de dependencia.

A continuación en la Tabla 12 se detallan las técnicas de minería de datos que se aplicaron a cada uno de los modelos de minería de datos para poder lograr cada uno de los objetivos de minería de datos definidos previamente.

Tabla 12. Técnicas de minería de datos utilizadas en los modelos de minería de datos del proyecto

Estructura DM	Modelo DM	Técnica DM
Indicador Docente	Clasificación de Indicador Docente	Algoritmo de Clustering de Microsoft
	Árboles de Decisión de Indicador Docente	Algoritmo de Árboles de Decisión de Microsoft
Indicador Estudiante	Clasificación de Indicador Estudiante	Algoritmo de Clustering de Microsoft
	Árboles de Decisión de Indicador Estudiante	Algoritmo de Árboles de Decisión de Microsoft
Indicadores Docente	Redes de Asociación de Indicadores Docente	Algoritmo de Reglas de Asociación de Microsoft
Indicadores Estudiante	Redes de Asociación de Indicadores Estudiantes	Algoritmo de Reglas de Asociación de Microsoft

En la siguiente figura se aprecian los valores para la probabilidad mínima (por defecto es 0,40) y para la Importancia mínima (-0,32), los cuales como para este algoritmo de reglas de asociación, para este algoritmo y los demás se tienen parametrizados sus resultados por medio de la selección de la estructura de minería de datos, el cuestionario y por un indicador en particular, información sobre la cual queremos analizar.

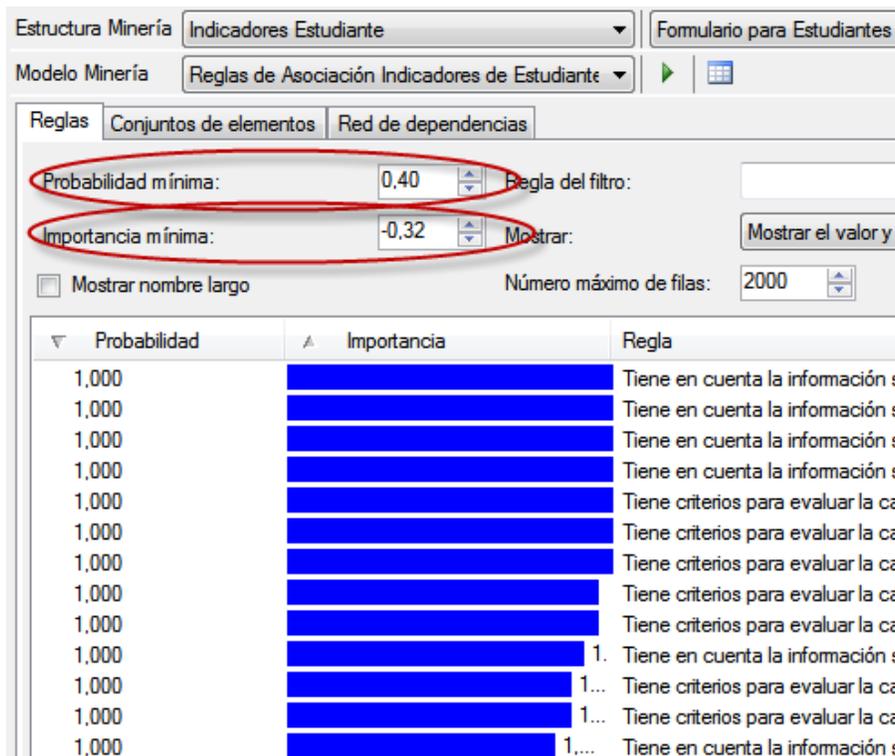


Figura 50. Configuración algoritmo de reglas de asociación

Implementación de los modelos de minería de datos

A continuación en la **Figura 51** se muestra el entorno de desarrollo de SQL Server Analysis Services para implementar el modulo de minería de datos del proyecto. En el **Anexo 33** se detalla la implementación de las estructuras de minería de datos y sus respectivos modelos.

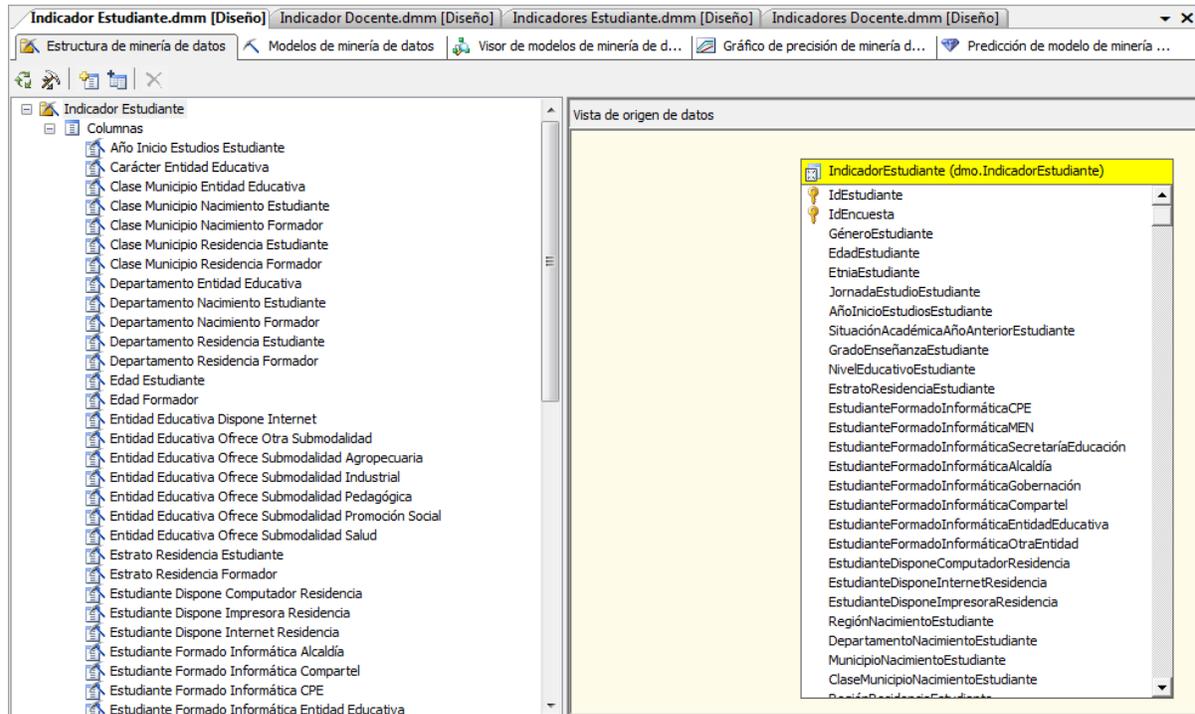


Figura 51. Entorno de Desarrollo de SQL Server Analysis Services para Modulo de Minería de Datos

Validación de los modelos de minería de datos

Para realizar la validación de los modelos de minería de datos se partió de que las estructuras de minería de datos se analizan en contextos diferentes (por modelo de competencias o por indicadores de un modelo de competencias) por lo tanto se seleccionaron al azar algunos de éstos contextos como entrenamiento y otros de prueba.

Tanto para los contextos de entrenamiento y de prueba se seleccionaron:

- 2 indicadores de la dimensión técnica y tecnológica
- 2 indicadores de la dimensión pedagógica
- 2 indicadores de la dimensión comunicativa y colaborativa

Además se exploraron los resultados obtenidos de cada uno de los modelos de minería de datos en cada una de las estructuras de datos, los cuales se pueden visualizar a través de los diferentes visores de los modelos de minería de datos que proporciona Analysis Services, permitiéndonos comparar directamente los modelos de minería de datos y elegir el más preciso o útil.

En este proceso de validación de los modelos de minería de datos se fueron descartando algunas variables que consideraron no influir en los modelos evaluados.

En el **Anexo 34** se detallan los modelos de minería de datos resultado.

6.5. Evaluación

En esta etapa se ha desarrollado en el proyecto un(os) modelo(s) que parecen tener alta calidad de una perspectiva de análisis de datos. Antes de seguir al despliegue final del modelo, es importante evaluarlo minuciosamente y examinar los pasos ejecutados para formularlo [9].

En esta fase se procedió a evaluar los resultados obtenidos en cada uno de los modelos y se realizaron las siguientes conclusiones:

- En la información almacenada en el DW, se presentan casos en los cuales los beneficiarios obtuvieron resultados de los indicadores con valores homogéneos (en su gran mayoría se clasifican en niveles medios y bajos), por esta razón los modelos de minería de datos dan como resultado conclusiones homogéneas.
- Para que los modelos de minería nos permitan encontrar correlaciones, patrones y/o tendencias ocultas en la información, es necesario incluir información de más convenios del programa CPE, por el momento, tan solo se tiene la información del convenio CPE-Universidad del Cauca y de una muestra representativa de la totalidad de la población beneficiada en el año 2007.
- Se debe realizar un monitoreo continuo de los modelos de minería de datos con la nueva información que le sea suministrada al sistema.

Los modelos de minería de datos fueron modelados de tal manera que soporten información de varios convenios. Para obtener un mejor grado de exactitud a la realidad, es necesario que se le suministre al sistema, información representativa de mas convenios de CPE con las diferentes universidades del país, teniendo de esta manera datos representativos de las diferentes Estrategias de Acompañamiento Educativo utilizadas para la ejecución de las respectivas Fases de Profundización de cada convenio en las instituciones beneficiadas, obteniendo información suficiente para obtener conclusiones importantes y valiosas para el personal administrativo de los convenios de las universidades con CPE.

6.6. Despliegue

La creación del modelo no es generalmente al final del proyecto, incluso si el propósito del modelo es incrementar el conocimiento de los datos, el conocimiento adquirido necesitará ser organizado y presentado de una manera que el cliente puede usarlo [9].

En esta fase se procedió a integrar los visores de los diferentes modelos de minería resultado al Sistema de Soporte de Toma de Decisiones FP-CPE como se muestra en la **Figura 52**. Además se adicionó al aplicativo un sistema para que el usuario pudiera parametrizar el modelo de competencias y su respectivo indicador en cada uno de los modelos de minería de datos.

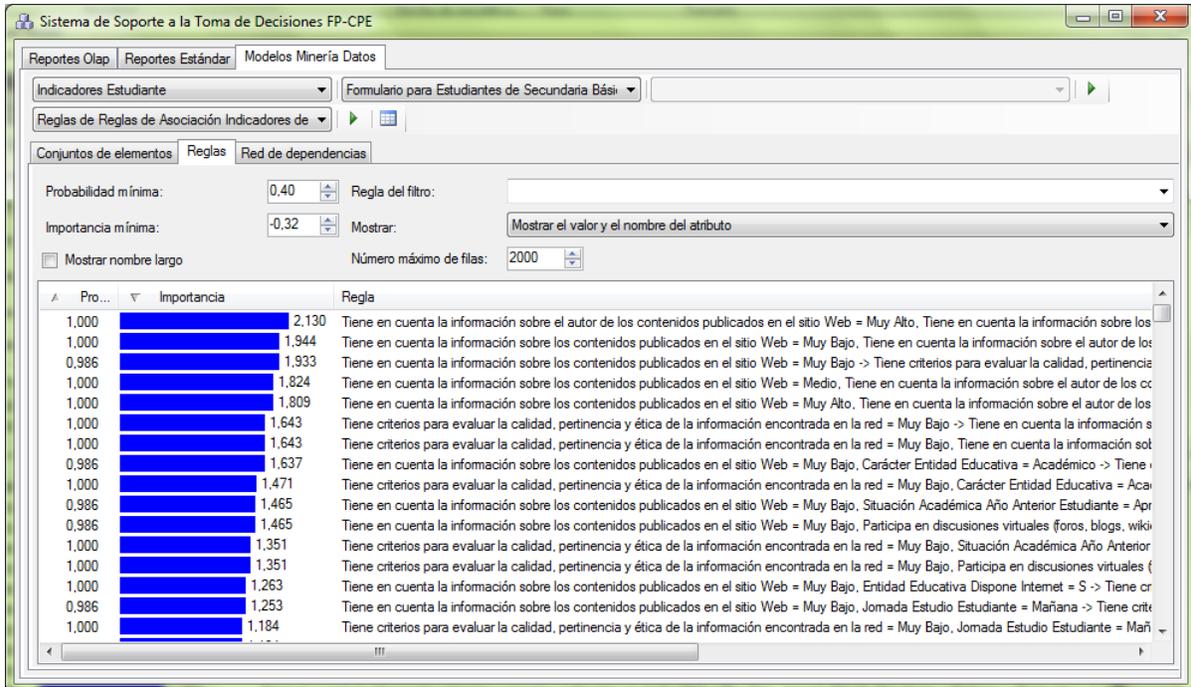


Figura 52. Modelo de minería de datos en el Sistema de Soporte a la Toma de Decisiones de FP-CPE.

CAPÍTULO VII. CONCLUSIONES Y TRABAJOS FUTUROS

Conclusiones

A continuación se definen las conclusiones que se derivaron del presente trabajo de grado. Se sintetizará sobre los resultados más importantes de una forma clara y precisa. Se presentarán conclusiones relacionadas con los casos de diseño del DW, los modelos de minería de datos generados y las herramientas de acceso al DSS.

- En el Modelo de Competencias en TIC definido por CPE las competencias, desempeños e indicadores presentan una forma jerárquica, por esta razón lo más conveniente fue diseñar la dimensión indicador (DimIndicador) en el DW con una jerarquía de organización [8] o de elementos primarios y secundarios [19], la cual articula físicamente los diferentes niveles de la jerarquía (Competencia -> Desempeño -> Indicador -> etc.).
- En el Sistema de Información de la Fase de Profundización (SIFP) las posibles respuestas de una pregunta son categorizadas o agrupadas para su presentación por esta razón lo más conveniente fue diseñar la dimensión respuesta, (DimRespuesta) en el DW con una jerarquía de organización [8] o de elementos primarios y secundarios [19], la cual articula físicamente los diferentes niveles de la jerarquía (Grupo de Respuestas->Sub Grupo de Respuesta->Respuesta).
- Con la dimensión indicador (DimIndicador) se presentó un caso de diseño particular donde se necesitaba modelar el peso de un indicador sobre su indicador padre. Este caso se solucionó diseñando un puente entre el indicador y su indicador padre en el cual se almacenaba el peso que relacionan estos indicadores, pero en la implementación se almacenó el peso (medida) en la dimensión indicador y utilizar la dimensión indicador como una tabla de hechos. Este caso es muy similar al tratado en [8] como una dimensión degenerada y en [19] como dimensión de hechos, en donde una tabla de hechos es utilizada como una dimensión.⁹
- La aplicación de sistemas de inteligencia de negocios están dirigidos a necesidades empresariales, en donde se maneja muchos datos de tipo estadístico; muy útil en un DW, a diferencia de los contextos educativos donde no se manejan muchos de éstos tipos de datos, por esta razón en el DW se generaron muchas tablas de hechos y puentes sin medidas, sin embargo a través del sistema DSS desarrollado en este trabajo de grado se logró dar un valor cualitativo a la información relacionada en éstos contextos a través de valores porcentuales (respuestas de encuestas) e indicadores (modelos de competencias).
- En el proceso de implementación se encontró, que las herramientas seleccionadas no ofrecen un soporte adecuado de agregaciones para el control del comportamiento de las medidas de pesos ponderados utilizadas en las jerarquías de organización creadas para el proceso de la medición de las competencias en

⁹ Este caso de diseño también se presenta en la dimensión respuesta (DimRespuesta)
SISTEMA DE SOPORTE A LA TOMA DE DECISIONES EN EL MARCO DE COMPUTADORES PARA EDUCAR
(CPE) EN LA FASE DE PROFUNDIZACIÓN (FP)

TIC de los beneficiarios, lo cual hace difícil elaborar soluciones para este caso de diseño dentro de este tipo de contextos, implicando la utilización de cálculos, los cual ocasiona problemas de rendimiento.

- A través del Sistema de Soporte a la Toma de Decisiones desarrollado, el personal del convenio CPE-Universidad del Cauca, podrá apoyarse para definir mejor las estrategias de acompañamiento educativo a través de los diferentes reportes estándar y consultas ad-hoc que pueden generarse en el sistema, brindando información histórica de las fases de profundización (formadores, jornadas de formación, beneficiarios, convenios, modelos de competencias, entre otros).
- Los modelos de minería de datos fueron desarrollados de tal manera que soporten información relacionada de varios convenios de CPE, de esta manera se estaría considerando diferentes Estrategias de Acompañamiento Educativo, al tener disponible esta información los modelos le permitirán al personal de CPE descubrir correlaciones, patrones y/o tendencias ocultas en esta información.
- Los modelos de minería seleccionados permitirán poder encontrar correlaciones, patrones y/o tendencias ocultas en la información a los directivos de CPE a nivel nacional, cuando se tenga información almacenada en el DW de varios o todos los convenios que tiene CPE con las diferentes universidades del país.

Recomendaciones y Trabajos Futuros

- Aplicar el formulario de Información general y Socio-económica, al final o durante el transcurso de las Jornadas de Formación, dado que esta información está disponible en aquel momento y no resulta dispendiosa su recolección.
- Aplicar algunas de las preguntas del formulario de las Competencias Técnicas y Tecnológicas, al finalizar las Jornadas de Formación, debido a que parte de los desempeños y habilidades técnicas y tecnológicas, se refieren al conocimiento teórico adquirido recientemente durante las jornadas.
- Aplicar los formularios de las Competencias Pedagógicas, y Comunicativas y Colaborativas al año siguiente de haber finalizado las Jornadas de Formación, ya que luego de este periodo de tiempo los docentes ya han desarrollado y aplicado el conocimiento práctico adquirido en las jornadas de formación.
- La implementación del Data Mart de la Fase Inicial de la estrategia de acompañamiento educativo, que tiene como objetivo crear otros tipos de indicadores que permitan por ejemplo identificar la disponibilidad real o más precisa de los recursos tecnológicos de las instituciones educativas en las fases de profundización.
- La implementación de un Sistema Integral de la Fase de Profundización que maneje la información relacionada de las jornadas de formación (programación, temáticas, formadores, etc.), encuestas, visitas, evaluaciones, muestreos, etc.

Sobre el cual el Sistema de Soporte a la Toma de Decisiones desarrollado se base para obtener información más precisas.

- Investigaciones sobre otras técnicas o casos diseños para realizar los cálculos en tiempo real de los indicadores de los modelos de competencia en TIC, permitiendo optimizar así las consultas sobre las herramientas OLAP o de presentación de información.
- Realizar el proceso de recolección de información para incluir en el DSS información de los docentes y estudiantes de las instituciones beneficiadas durante la ejecución de las Estrategias de Acompañamiento Educativo de los convenios que tiene CPE con las diferentes universidades del país, teniendo información representativa de todas las Estrategias de Acompañamiento Educativo utilizadas para poder realizar análisis y tomar decisiones a nivel nacional.

BIBLIOGRAFIA

- [1] MC. (28 de Junio de 2006). *Principios Generales para la Construcción de una Política Nacional de TIC*. Obtenido de Ministerio de Comunicaciones: <http://www.mincomunicaciones.gov.co>
- [2] CPE. (2007). *Términos de Referencia Fase de Profundización 2008*. Computadores para Educar, Bogotá.
- [3] CPE. (2004). *¿Qué es CPE?* Recuperado el 1 de Febrero de 2008, de Sitio Web de Computadores para Educar: http://www.computadoresparaeducar.gov.co/que_es.html
- [4] SAT (s.f.). *TT04 - Proceso de incorporación de las TIC en la escuela*. Coreduc SAT - Instituciones Operadoras del Sistema de Aprendizaje Tutorial SAT.
- [5] MEN. (2005). *Colombia Aprende - Programa Nacional de Nuevas Tecnologías*. Recuperado el 21 de Febrero de 2008, de Colombia Aprende: <http://www.colombiaprende.edu.co/html/home/1592/article-102549.html>
- [6] Cobo, Cristóbal (s.f.). *Propuestas para apoyar la incorporación de las TIC en la escuela*. FLACSO. México.
- [7] Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W. (1998). *The Data Warehouse Lifecycle Toolkit* (First Edition ed.). New York, New York, USA: John Wiley & Sons, Inc.
- [8] J. Mundy, W. Thomthwaite, R. Kimball. *The Microsoft Data Warehouse Toolkit: With SQL Server 2005 and the Microsoft business Intelligence Toolset*, wiley Publishing , Inc., Indianápolis, Indiana, 2006.
- [9] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., y otros. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*.
- [10] CPE. (2004). *¿Quiénes somos?* Recuperado el 1 de Febrero de 2008, de Sitio Web de Computadores para Educar: http://www.computadoresparaeducar.gov.co/quienes_somos.html
- [11] CPE. (2004). *Acompañamiento Educativo*. Recuperado el 1 de Febrero de 2008, de Sitio Web de Computadores para Educar: http://www.computadoresparaeducar.gov.co/apoyo_apost_escuelas.html
- [12] Imhoff, C., Gallemmo, N., & Geiger, J. G. (2003). *Mastering Data Mining Desing*. (R. M. Elliott, Ed.) Indianapolis, Indiana, USA: John Wiley & Sons, Inc.
- [13] Thomas, L. (s.f.) *Inteligencia de Negocios*. Recuperado el 2 de Marzo de 2008, del Sitio Web Idensa: <http://www.idensa.com/index.html>
- [14] Inmon, W. H. (2002). *Building the Data Warehouse* (Third Edition ed.). (R. M. Elliott, Ed.) John Wiley & Sons, Inc.
- [15] Mundy, J., Thornthwaite, W., & Kimball, R. (2006). *The Microsoft Data Warehouse Toolkit*. (R. M. Elliott, Ed.) Indianapolis, Indiana, USA: John Wiley & Sons, Inc.
- [16] Gartner, G. (s.f.). *Glosary*. Recuperado el 19 de Febrero de 2008, de Sitio Web de Gartner Group: <http://www.gartner.com>

- [17] Oncins de Frutos, M. (s.f.). Encuestas: metodología para su utilización. Centro Nacional de Condiciones de Trabajo.
- [18] Instituto Canario de Evaluación y Calidad Educativa (ICEC) - Consejería de Educación Cultura y Deportes del Gobierno de Canarias (Junio de 2004). Competencias básicas en las Tecnologías de la Información y la Comunicación. Canarias. ISBN 84-688-6927-9.
- [19] MSDN, Microsoft Developer Network, <http://msdn.microsoft.com>
- [20] Microsoft TechNet, Recursos para Profesionales de TI, <http://technet.microsoft.com>
- [21] Introduction to SQL Server 2005 Data Mining, Visitado el 15 de Noviembre de 2009, MSDN, Microsoft Developer Network, [http://msdn.microsoft.com/en-us/library/ms345131\(SQL.90\).aspx](http://msdn.microsoft.com/en-us/library/ms345131(SQL.90).aspx)
- [22] SQL Server Data Mining: Plug-In Algorithms, MSDN, Microsoft Developer Network, Visitado el 15 de Noviembre de 2009, [http://msdn.microsoft.com/en-us/library/ms345133\(SQL.90\).aspx](http://msdn.microsoft.com/en-us/library/ms345133(SQL.90).aspx)
- [23] Proyectos de minería de datos (Analysis Services - Minería de datos), MSDN, Microsoft Developer Network, Visitado el 15 de Noviembre de 2009, <http://msdn.microsoft.com/es-es/library/ms174861.aspx>