

ESCUELA POLITÉCNICA NACIONAL

ESCUELA DE INGENIERÍA

ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL ÁREA DE SISMOLOGÍA DEL DEPARTAMENTO DE GEOFÍSICA DE LA ESCUELA POLITÉCNICA NACIONAL

**PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN
SISTEMAS INFORMATICOS Y DE COMPUTACION**

MICHAEL WLADIMIR VIZUETE NARANJO

CARLOS PATRICIO YELA SHININ

DIRECTOR: ING. MARIA HALLO

Quito, Marzo 2006

DECLARACIÓN

Nosotros, Michael Wladimir Vizuite Naranjo y Carlos Patricio Yela Shinín, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que hemos consultado las referencias bibliográficas que se incluyen en este documento.

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

Michael Wladimir Vizuite Naranjo

Carlos Patricio Yela Shinín

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Michael Wladimir Vizúete Naranjo y Carlos Patricio Yela Shinín, bajo mi supervisión.

Ing. María Hallo
DIRECTOR DE PROYECTO

CONTENIDO

1 CASO DE ESTUDIO	19
1.1 GENERALIDADES SOBRE SISMOLOGÍA.....	19
1.1.1 INTRODUCCIÓN.....	19
1.1.2 SISMOLOGÍA.....	19
1.1.2.1 Ondas Sísmicas	20
1.1.2.2 Medios de estudio	23
1.1.2.2.1 Sismógrafo	23
1.1.2.3 Estructura Interna de la tierra	26
1.1.2.3.1 Núcleo.	26
1.1.2.3.2 Manto.....	27
1.1.2.3.3 Corteza o Litosfera.	27
1.1.2.4 Placas Tectónicas	27
1.1.2.4.1 Interacción entre las Placas Tectónicas	30
1.1.2.5 Movimientos Sísmicos.....	31
1.1.2.6 Fallas Tectónicas	33
1.1.2.6.1 Tipos de fallas	33
1.1.2.7 Terremotos y Zonas Sísmicas.....	34
1.1.2.7.1 Tipos de Terremotos	35
1.1.2.8 Intensidad Sísmica	36
1.1.2.8.1 Escala de Mercalli	36
1.1.2.8.2 Escala de Richter	38
1.2 ESTUDIO DE LA SITUACION ACTUAL DEL AREA DE SISMOLOGÍA DEL INSTITUTO GEOFÍSICO DE LA ESCUELA POLITÉCNICA NACIONAL.....	40
1.2.1 ASPECTOS GENERALES	40
1.2.1.1 Descripción.....	40
1.2.1.2 Objetivos	40
1.2.1.3 Misión.....	41
1.2.1.4 Visión	41
1.2.2 ESTRUCTURA COMPUTACIONAL.....	42
1.2.2.1 Recursos de Hardware.....	42
1.2.2.1.1 Recursos de Red.....	42

1.2.2.1.2	Recursos de Hardware	43
1.2.2.2	Recursos de Software	43
1.3	DATA WAREHOUSE Y DATA MART	45
1.3.1	INTRODUCCION.....	45
1.3.2	DATA WAREHOUSE	46
1.3.2.1	Definiciones Generales	46
1.3.2.2	Características	47
1.3.2.3	Estructura del Data Warehouse	50
1.3.2.4	Usos de los Data Warehouse.....	53
1.3.2.5	Arquitectura Data Warehouse	53
1.3.2.6	Ventajas y desventajas	54
1.3.2.6.1	Ventajas.....	55
1.3.2.6.2	Desventajas.....	55
1.3.3	DATA MART.....	55
1.3.3.1	Definiciones generales	55
1.3.3.2	Características	56
1.3.3.3	Tipos de Data Mart.....	56
1.3.3.4	Carga de datos en Data Mart	57
1.3.3.5	Procesos de depuración.....	58
1.3.3.5.1	Modos de Procesamiento de depuración	58
1.3.3.6	Procesos de refrescamiento.....	58
1.3.3.7	Ventajas y desventajas	58
1.3.3.7.1	Ventajas.....	58
1.3.3.7.2	Desventajas.....	59
1.4	METODOLOGÍA PARA EL DESARROLLO DE UN DATA MART	59
1.4.1	PLANIFICACIÓN.....	59
1.4.1.1	Seleccionar la estrategia de implementación	59
1.4.1.1.1	Enfoque de arriba hacia abajo.....	60
1.4.1.1.2	Enfoque de abajo hacia arriba.....	60
1.4.1.1.3	Enfoque combinado.....	60
1.4.1.2	Seleccionar la metodología de desarrollo	60
1.4.1.2.1	Método de análisis y Diseño estructurado	60
1.4.1.2.2	Método Espiral.....	61

1.4.1.3	Selección del ámbito de implementación	63
1.4.1.4	Selección de la Arquitectura de implementación.....	63
1.4.1.4.1	El almacenamiento operacional con uso de datos operacionales.....	64
1.4.1.4.2	Solo Data Warehouse.....	64
1.4.1.4.3	Solo Data Mart.....	64
1.4.1.4.4	Data Warehouse y Data Mart	64
1.4.1.5	Cronograma y Presupuesto del proyecto	64
1.4.1.6	Recopilación de Metadatos	65
1.4.2	REQUERIMIENTOS.....	65
1.4.2.1	Definir los requerimientos generales o propios	66
1.4.2.2	Definir los requerimientos arquitectónicos.....	69
1.4.2.2.1	Arquitectura de datos	70
1.4.2.2.2	Arquitectura de aplicación	70
1.4.2.2.3	Arquitectura de tecnología.....	70
1.4.2.3	Definir los requerimientos del desarrollador	71
1.4.2.3.1	Requerimientos de tecnología	71
1.4.2.3.2	Requerimientos de despliegue	72
1.4.2.3.3	Requerimientos de disposición para la producción del Data Warehouse.....	73
1.4.2.3.4	Requerimientos para el desarrollo y despliegue del personal y sus habilidades	74
1.4.2.4	Definir los requerimientos de usuario final.....	74
1.4.2.4.1	Flujo de trabajo.....	74
1.4.2.4.2	Requerimientos de consulta	76
1.4.2.4.3	Requerimientos de reportes	77
1.4.3	ANÁLISIS	77
1.4.3.1	Analizar los factores que conducen el negocio (Business Drivers)	78
1.4.3.2	Analizar los objetivos del negocio	78
1.4.3.3	Analizar las necesidades de información de alto nivel	79
1.4.3.4	Recolectar información de procesos	79
1.4.3.5	Identificar áreas críticas	79

1.4.3.6	Identificar roles de personas	79
1.4.3.7	Revisar infraestructura informática.....	79
1.4.3.8	Revisar Planes Tácticos y Estratégicos	79
1.4.3.9	Identificar limitaciones y restricciones	80
1.4.3.10	Recolectar y revisar documentación de la empresa.....	80
1.4.3.11	Identificar procesos de alto nivel e indicadores claves de rendimiento.....	80
1.4.3.12	Analizar las fuentes de datos	80
1.4.4	DISEÑO.....	81
1.4.4.1	Diseño detallado de la arquitectura de datos	81
1.4.4.2	Diseño detallado de la arquitectura de la aplicación	81
1.4.5	IMPLEMENTACIÓN	81
1.4.5.1	Construcción	81
1.4.5.2	Implantación.....	83
1.4.6	PRUEBAS	84
1.4.6.1	Pruebas de Unidad.....	84
1.4.6.2	Pruebas de Integración	84
1.4.6.3	Pruebas de Validación	84
1.4.6.4	Pruebas del Sistema	84
2	DESARROLLO DE UN DATA MART PARA EL ÁREA DE SISMOLOGÍA DEL DEPARTAMENTO DE GEOFÍSICA DE LA ESCUELA POLITÉCNICA NACIONAL	86
2.1	PLANIFICACIÓN	86
2.1.1	SELECCIONAR LA ESTRATEGIA DE IMPLEMENTACIÓN.....	86
2.1.2	SELECCIONAR LA METODOLOGÍA DE DESARROLLO	86
2.1.3	SELECCIÓN DEL ÁMBITO DE IMPLEMENTACIÓN.....	87
2.1.4	SELECCIÓN DE LA ARQUITECTURA DE IMPLEMENTACIÓN	89
2.1.5	CRONOGRAMA Y PRESUPUESTO DEL PROYECTO.....	89
2.1.6	RECOPIACIÓN DE METADATOS	89
2.2	ETAPA DE DEFINICIÓN DE REQUERIMIENTOS	90
2.2.1	DEFINIR LOS REQUERIMIENTOS GENERALES O PROPIOS.....	90
2.2.2	DEFINIR LOS REQUERIMIENTOS ARQUITECTÓNICOS.....	90
2.2.3	DEFINIR LOS REQUERIMIENTOS DEL DESARROLLADOR.....	91

2.2.4	DEFINIR LOS REQUERIMIENTOS DE USUARIO FINAL	92
2.3	ETAPA DE ANÁLISIS.....	92
2.3.1	ANALIZAR DE LOS FACTORES QUE CONDUCEN EL NEGOCIO...92	92
2.3.2	ANALIZAR DE LOS OBJETIVOS DEL NEGOCIO	92
2.3.3	ANALIZAR LAS NECESIDADES DE INFORMACIÓN DE ALTO NIVEL	93
2.3.4	DEFINIR PROCESOS Y ROLES	93
2.3.4.1	Procesos del Instituto Geofísico para el Área de Sismología.....93	93
2.3.4.1.1	Monitoreo de los eventos Sísmicos	93
2.3.4.1.2	Análisis y estudios de Eventos Sísmicos.....94	94
2.3.4.1.3	Evaluación de peligros Sísmicos - Volcánicos.....94	94
2.3.4.1.4	Difusión de información	95
2.3.4.2	Roles dentro del Instituto Geofísico para el Área de Sismología..95	95
2.3.5	DEFINIR LOS INDICADORES CLAVES DE RENDIMIENTO	96
2.3.6	DIMENSIONES Y HECHOS.....96	96
2.3.7	ANALIZAR LAS FUENTES DE DATOS	97
2.3.7.1	Base de datos de sismos localizados e históricos.....97	97
2.3.7.2	Mecanismos focales	97
2.4	DISEÑO.....	97
2.4.1	DISEÑO DETALLADO DE LA ARQUITECTURA DE DATOS.....97	97
2.4.2	DISEÑO DETALLADO DE LA ARQUITECTURA DE LA APLICACIÓN.	98
2.5	IMPLEMENTACIÓN.....	99
2.5.1	CONSTRUCCIÓN	99
2.5.1.1	Selección de Plataformas de Software.....99	99
2.5.1.1.1	Sistema Operativo	100
2.5.1.1.2	Base de Datos	100
2.5.1.1.3	Herramientas de Carga de Datos	100
2.5.1.1.4	Transformación Multidimensional	100
2.5.1.1.5	Herramienta OLAP	101
2.5.1.1.6	Herramientas de Metadatos	101
2.5.1.2	Construcción de la Base de Datos	101
2.5.1.2.1	Requerimientos Físicos de la Base de Datos	101

2.5.1.2.2	Construcción de los Objetos del Data Mart de Sismología..	101
2.5.1.3	Extracción, Transformación y Carga de Datos	101
2.5.1.3.1	Acceso a las fuentes y Transformación de Datos.....	102
2.5.1.4	Transformación Multidimensional.....	105
2.5.1.5	Visualización de Datos	112
2.5.2	IMPLANTACIÓN.....	117
2.5.2.1	Instalación Inicial de las Herramientas	117
2.6	PRUEBAS.....	133
2.6.1	PRUEBAS DE UNIDAD.....	133
2.6.2	PRUEBAS DE INTEGRACIÓN.....	134
2.6.3	PRUEBAS DE VALIDACIÓN.....	135
2.6.4	PRUEBAS DEL SISTEMA.....	136
3	CONCLUSIONES Y RECOMENDACIONES	140
3.1	CONCLUSIONES	140
3.2	RECOMENDACIONES	141
	BIBLIOGRAFÍA	142
	LIBROS	142
	PAGINAS WEB	142

INDICE DE FIGURAS

Figura 1	Ondas Sísmicas	20
Figura 2	Ondas Primarias.....	21
Figura 3	Ondas Secundarias	21
Figura 4	Ondas Superficiales	22
Figura 5	Tipos de Ondas Superficiales.....	22
Figura 6	Sismógrafo	23
Figura 7	Registro de un sismo (Sismograma)	25
Figura 8	Principales capas que componen la Tierra.....	26
Figura 9	Corteza Terrestre	27
Figura 10	Placas Tectónicas.....	29
Figura 11	Interacción entre las placas tectónicas	30
Figura 12	Esquema de los desplazamientos de las Placas Tectónicas.....	31
Figura 13	Movimientos Sísmicos	32
Figura 14	Tipos de Fallas Tectónicas	33
Figura 15	Características de un Terremoto.....	34
Figura 16	Diagrama de la topología de la red del Instituto Geofísico.....	42
Figura 17	De tiempo variante.....	49
Figura 18	No volátil	50
Figura 19	Estructura de los datos de un Data Warehouse	51
Figura 20	Data Marts Dependientes	56
Figura 21	Data Marts Independientes.....	57
Figura 22	Método de Desarrollo Estructurado	61
Figura 23	Método de Desarrollo en Espiral.....	62
Figura 24	Requerimientos para un Proyecto de Data Warehouse.....	66
Figura 25	Pirámide de datos que muestra la relación entre la granularidad	68
Figura 26	Intereses comunes para la mayoría de organizaciones.....	69
Figura 27	Flujo de trabajo de una Investigación de Mercado previo al Data Warehouse	75
Figura 28	Flujo de trabajo de una Investigación de Mercado luego del Data Warehouse	76
Figura 29	Diseño de la Arquitectura de Datos	98
Figura 30	Diseño de la Arquitectura de la Aplicación.....	99

Figura 31	Pantalla de creación del paquete de transformación de datos	102
Figura 32	Pantalla de definición del origen de datos	103
Figura 33	Pantalla de definición del origen de los datos	103
Figura 34	Pantalla de definición de las transformaciones de los datos.....	104
Figura 35	Vista del Paquete DTS.....	104
Figura 36	Pantalla del Asistente para la generación de cubos	105
Figura 37	Pantalla de selección de la tabla de hechos	106
Figura 38	Pantalla de selección de la columnas numéricas de hechos	106
Figura 39	Pantalla del asistente para generar las dimensiones.....	107
Figura 40	Pantalla de selección de tipo de dimensiones	107
Figura 41	Pantalla de selección de las tablas de dimensiones	108
Figura 42	Pantalla de selección de la dimensión sobre la que trabajaremos.	108
Figura 43	Pantalla de selección de grado de granularidad de la dimensión tiempo	109
Figura 44	Pantalla de presentación de la dimensión Tiempo.....	109
Figura 45	Pantalla de selección de la dimensión sobre la que trabajaremos.	110
Figura 46	Pantalla de selección de grado de granularidad de la dimensión Región	110
Figura 47	Pantalla de presentación de la dimensión Región	111
Figura 48	Pantalla de selección de las dimensiones que contendrá el cubo .	111
Figura 49	Pantalla de presentación del Cubo de Sismología.....	112
Figura 50	Pantalla de inicio de la herramienta OPEN I.....	113
Figura 51	Pantalla de presentación de los proyectos del servidor OLAP	113
Figura 52	Pantalla de ingreso para administrar un proyecto.....	114
Figura 53	Pantalla de configuración de la fuente de datos	114
Figura 54	Pantalla de creación de filtros para generación de datos	115
Figura 55	Pantalla inicial de visualización de la información del Data Mart de Sismología.....	116
Figura 56	Pantalla de visualización de la información del Data Mart de Sismología con barras de herramientas para modificarlas consultas.....	116
Figura 57	Pantalla de visualización de la información del Data Mart de Sismología con visualización del texto de la consulta que se visualiza.....	117
Figura 58	Menú de Instalación de SQL Server 2000	118

Figura 59	Menú de componentes de SQL Server.....	118
Figura 60	Pantalla de inicio de la Instalación de SQL Server	119
Figura 61	Pantalla de selección del equipo a instalar	119
Figura 62	Pantalla de selección de la instalación	120
Figura 63	Pantalla de ingreso de la información de usuario	120
Figura 64	Pantalla del Contrato de Licencia de SQL Server 2000.....	121
Figura 65	Pantalla de selección del tipo de instalación.....	121
Figura 66	Pantalla para configurar el nombre de la instancia de SQL Server	122
Figura 67	Pantalla para seleccionar el tipo de instalación	122
Figura 68	Pantalla para especificar las cuentas de los servicios de SQL Server	123
Figura 69	Pantalla para especificar el modo de autenticación	123
Figura 70	Pantalla de información del inicio de la copia de archivos	124
Figura 71	Indicador de la progresión de la copia de archivos.....	124
Figura 72	Pantalla de información de un exitoso proceso de instalación del servidor de bases de datos de SQL Server.....	125
Figura 73	Pantalla de Inicio de instalación del Análisis Services de SQL Server 2000	125
Figura 74	Pantalla del Contrato de Licencia de SQL Server 2000.....	126
Figura 75	Pantalla de selección de componentes a instalar	126
Figura 76	Pantalla de selección de la carpeta donde se desplegará los accesos al Análisis Services	127
Figura 77	Pantalla de información de un exitoso proceso de instalación del Análisis Services de SQL Server.....	127
Figura 78	Menú Inicio con los nuevos accesos a SQL Server	128
Figura 79	Pantalla del Administrador de Servicios de SQL Server	128
Figura 80	Pantalla ya con los Servicios iniciados	129
Figura 81	Asistente de instalación de Java 2 SDK	129
Figura 82	Pantalla de selección de aplicaciones a instalar	130
Figura 83	Pantalla de copia de archivos de instalación	130
Figura 84	Asistente de instalación de Apache Tomcat	131
Figura 85	Pantalla de Configuración de usuario de Apache	131
Figura 86	Asistente de instalación de Microsoft XML	132

Figura 87 Muestra de Archivos copiados132

INDICE DE TABLAS

Tabla 1.1	Escalla de Mercalli.....	38
Tabla 1.2	Escala de Richter	39
Tabla 1.3	Equipos de red del Instituto Geofísico	42
Tabla 1.4	Recursos de Hardware del Instituto Geofísico	43
Tabla 1.5	Software Base del Instituto Geofísico.....	44
Tabla 1.6	Software Aplicativo del Instituto Geofísico.....	45
Tabla 1.7	Software de Utilitarios del Instituto Geofísico	45
Tabla 2.1	Volúmenes de Información del Instituto Geofísico.....	88
Tabla 2.2	Volúmenes de Información Documental.....	88
Tabla 2.3	Volúmenes de Información que se añadirá próximamente.....	89
Tabla 2.4	Roles	96

RESUMEN

En este trabajo hemos querido dar un breve recuento del desarrollo de una solución Data Mart para el Departamento de Sismología de la Escuela Politécnica Nacional con el fin de mejorar el manejo de la información de dicha área para prevenir sismos que han producido un gran impacto en muchas comunidades causando grandes pérdidas tanto humanas como económicas.

En el primer capítulo de este trabajo encontraremos generalidades sobre sismología lo cual nos ayudará a comprender de una mejor manera el objetivo de realizar este trabajo, además encontraremos definiciones sobre Data Warehouse y Data Mart las cuales nos guiarán en el proceso de desarrollo de la solución Data Mart que se implementa en el Instituto Geofísico y una Guía con la definición de los diferentes pasos que se deben seguir para el desarrollo de un Data Warehouse.

En el segundo capítulo encontramos el detalle de las distintas fases que se siguieron para el desarrollo de la solución Data Mart del Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional.

En el tercer capítulo encontraremos las conclusiones y recomendaciones que hemos obtenido del trabajo realizado en los dos primeros capítulos.

INTRODUCCION

En los últimos años el Instituto Geofísico de la Escuela Politécnica Nacional ha cobrado una gran trascendencia a nivel nacional e internacional al haberse reactivado varios volcanes activos en nuestro país y debido al aumento de la actividad sísmica, con el consiguiente peligro para la ciudadanía asentada en las zonas de afectación tanto de los procesos eruptivos de dichos volcanes como en las zonas con mayor riesgo de sufrir desastres por causa de los sismos. Gracias a la información que se obtiene de la Red Nacional de Sismógrafos, el Instituto ha podido emitir alertas tempranas para que las autoridades y la población puedan tomar a tiempo las medidas preventivas y correctivas correspondientes, en base a los mapas de peligro previamente establecidos por sus científicos procesando la información almacenada en el Instituto Geofísico.

Por esto se decidió la construcción de una solución Data Mart para el Área de Sismología con el fin de mejorar el manejo y el manejo de la gran cantidad de información de los sismos desde 1541 que se almacena en diferentes archivos tanto digitales como físicos dentro del Departamento de Geofísica.

Para empezar con la construcción del Data Mart de Sismología de Instituto Geofísico de la Escuela Politécnica Nacional debemos primero conocer algunos aspectos generales sobre sismología, medios de estudio de sismos, la estructura interna de la Tierra, las placas y fallas tectónicas, los movimientos sísmicos, las zonas sísmicas y las escalas de medición de la intensidad de un sismo.

También evaluaremos la situación actual del Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional, para lo cual veremos una breve descripción del Área de Sismología, los objetivos que cumple en beneficio de nuestro país, su misión, su visión y su estructura computacional, tanto hardware como software.

Para continuar con el marco teórico veremos breves conceptos sobre Data Warehouse, sus características, su estructura, los usos que se le da, su arquitectura, sus ventajas y sus desventajas.

También conoceremos algunas definiciones de Data Mart, sus características, los tipos de Data Mart que existen, los procesos de depuración y refrescamiento de la información que se cargará en un Data Mart, sus ventajas y sus desventajas.

Para finalizar con el marco teórico vamos a notar la metodología que usaremos para el desarrollo de la solución Data Mart de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional, la misma que esta compuesta por varias fases como son: la Planificación, recolección de requerimientos de todos los involucrados en este desarrollo, el análisis de los requerimientos recolectados, el diseño de la solución Data Mart tanto en su parte de almacenamiento como en la aplicación de despliegue de datos, la implementación de la solución Data Mart iniciando por su construcción en base a los diseños previos y luego la implementación para el uso de los usuarios finales y finalmente las pruebas de la solución en distintos aspectos para asegurar su correcto funcionamiento.

A continuación tenemos el proceso de desarrollo de la solución Data Mart del Área de Sismología del Instituto Geofísico en el cual se implementara la Metodología que notamos en el Capítulo 1.

Para iniciar con el desarrollo se planifica en distintos aspectos el desarrollo del Data Mart de Sismología ya que debemos seleccionar una estrategia de implementación, una metodología de desarrollo, el ámbito y la arquitectura en el que se implementara y además crearemos un cronograma general para el desarrollo del Data Mart de Sismología.

Seguimos con la definición de los requerimientos de los distintos actores que participan en el desarrollo del Data Mart de Sismología para esto se recolectara los requerimientos de los propietarios, de los arquitectos que diseñaran el Data Mart, de los desarrolladores y de los usuarios finales que usaran el Data Mart de Sismología.

Luego de haber recolectado los requerimientos se los debe analizar y crear especificaciones para el diseño del Data Mart. Dentro de la información que se analizará tenemos los factores que guían al Instituto Geofísico, los objetivos que tiene el Área de Sismología, las necesidades de información que tienen los niveles ejecutivos, los procesos internos del Área de Sismología, además se

definirán los indicadores claves de rendimiento y se analizarán las distintas fuentes de datos.

Ya hecho un análisis de la información recopilada en las etapas anteriores se procede a diseñar la arquitectura del almacén de datos y la arquitectura de la Aplicación del Data Mart de Sismología.

Seguimos con la construcción del Data Mart, primero seleccionando el software que nos brindara las mejores condiciones para desarrollar el Data Mart en todos sus aspectos, luego seguimos con la construcción del almacén de datos y con la extracción, carga y transformación de los datos a sus diferentes tablas, a continuación se procede la transformación multidimensional de los datos del Data Mart de Sismología y finalizamos la con la construcción de la aplicación que permitirá una sencilla visualización de los datos.

Ya construido el Data Mart de Sismología se procede a instalar el software que se seleccionó anteriormente que permitirá el funcionamiento del Data Mart de Sismología y el uso por parte de los usuarios finales.

Finalmente procederemos a realizar distintas pruebas que aseguren correcto funcionamiento del Data Mart de Sismología en cuanto al manejo de toda la información sísmica que almacena el Instituto Geofísico de la Escuela Politécnica Nacional.

CAPITULO 1

1 CASO DE ESTUDIO

1.1 GENERALIDADES SOBRE SISMOLOGÍA

1.1.1 INTRODUCCIÓN

La ubicación del Ecuador cerca de la zona donde convergen las placas Nazca y Sudamericana determina una importante actividad sísmica en su territorio y esto se refleja en los grandes terremotos han ocurrido a lo largo de la historia, los cuales constan en forma destacada en los catálogos sísmicos mundiales por sus grandes magnitudes debido a la cercanía de sus epicentros.

Los terremotos en el Ecuador han impactado también en su economía ya que a lo largo de los años los terremotos han producido un gran impacto ya que han causado daños a muchas comunidades que fueron destruidas por sismos quedando su población en la extrema pobreza.

En el Ecuador, la Institución que se encarga de monitorear, identificar y analizar las fuentes sísmicas y sus efectos es el Instituto Geofísico de la Escuela Politécnica Nacional, es por esto que el manejo de la información sobre Sismología es muy importante no solo para esta institución sino para el país.

1.1.2 SISMOLOGÍA

La sismología es una rama de la Geofísica y es la ciencia que estudia las causas que producen los terremotos, el mecanismo por el cual se producen y propagan las ondas sísmicas y la predicción del fenómeno sísmico.

La sismología implica la observación de las vibraciones naturales del terreno y de las señales sísmicas generadas de forma artificial, con muchas ramificaciones teóricas y prácticas. Como rama de la geofísica, la sismología ha aportado contribuciones esenciales a la comprensión de la tectónica de placas, la estructura del interior de la Tierra, la predicción de terremotos y es una técnica valiosa en la búsqueda de minerales.

1.1.2.1 Ondas Sísmicas

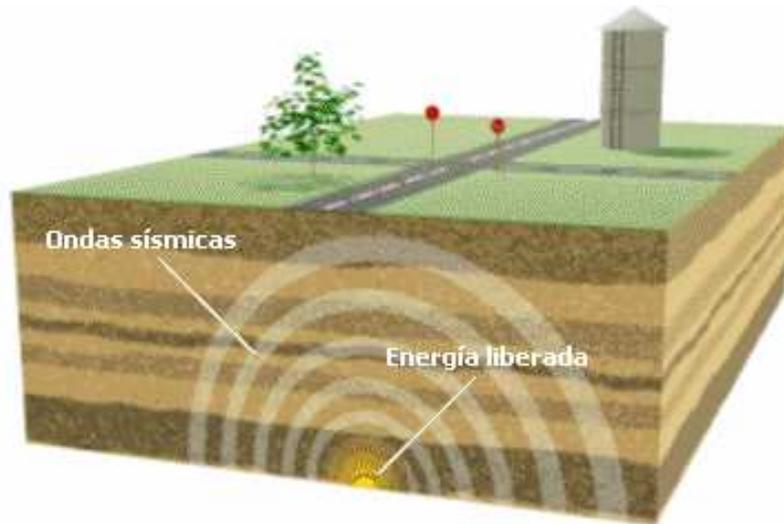


Figura 1 Ondas Sísmicas ¹

La deformación de los materiales rocosos produce distintos tipos de ondas sísmicas. Un deslizamiento súbito a lo largo de una falla, por ejemplo, produce ondas primarias, longitudinales o de compresión (ondas P) y secundarias, denominadas transversales o de cizalla (ondas S).

Los trenes de ondas P, de compresión, establecidos por un empuje (o tiro) en la dirección de propagación de la onda, causan sacudidas de atrás hacia adelante en las formaciones de superficie. La velocidad de propagación de las ondas P depende de la densidad de las rocas. En la propagación de las ondas de cizalla, las partículas se mueven en dirección perpendicular a la dirección de propagación. Las ondas P y las ondas S se transmiten por el interior de la Tierra; las ondas P viajan a velocidades mayores que las ondas S.

¹ Fuente: <http://www.obsis.unb.br/br/terremo.htm>

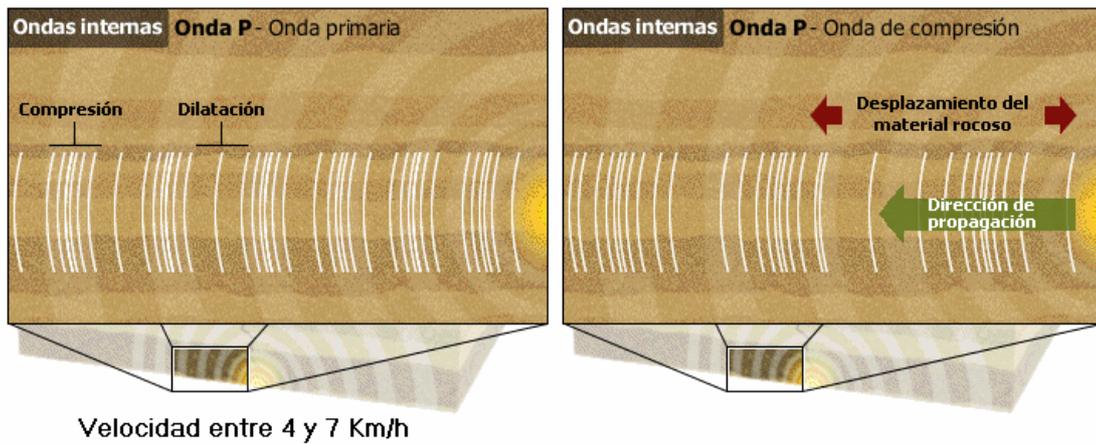


Figura 2 Ondas Primarias²

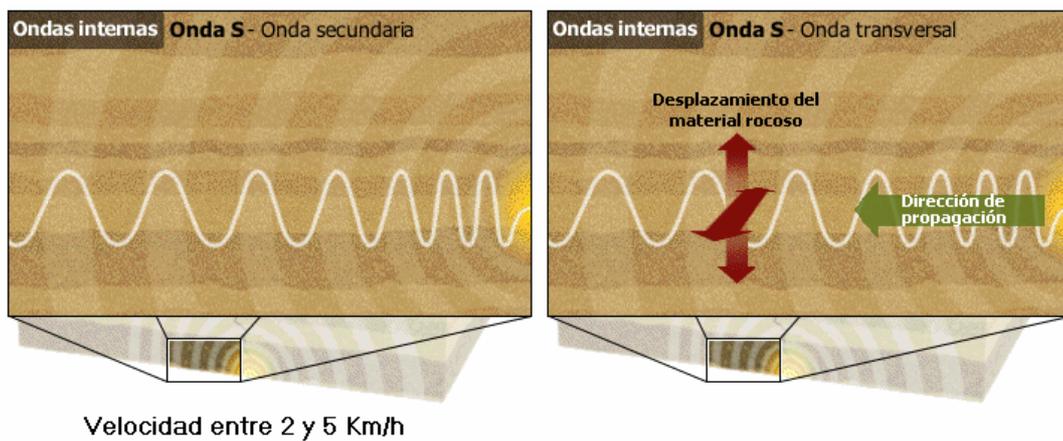


Figura 3 Ondas Secundarias

Cuando las ondas P y S encuentran un límite, como la discontinuidad de Mohorodovicic (Moho), que yace entre la corteza y el manto de la Tierra, se reflejan, refractan y transmiten en parte y se dividen en algunos otros tipos de ondas que atraviesan la Tierra. Las rocas graníticas corticales muestran velocidades típicas de onda P de 6 km/s, mientras que las rocas subyacentes máficas y ultramáficas (rocas oscuras con contenidos crecientes de magnesio y hierro) presentan velocidades de 7 y 8 km/s respectivamente.

²Fuente: <http://www.obsis.unb.br/br/terremo.htm>

Además de las ondas internas P y S, hay ondas superficiales que se propagan por la superficie terrestre y son las causantes de los mayores destrozos. Las ondas superficiales son más lentas que las ondas internas.

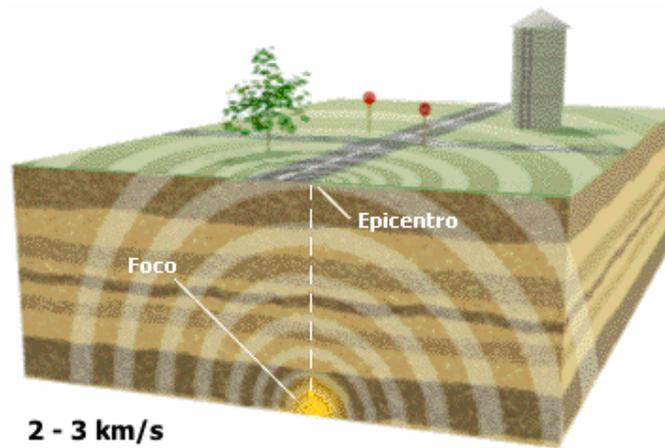


Figura 4 Ondas Superficiales³

Existen dos tipos de ondas superficiales; la más lenta es la llamada de Rayleigh en honor al físico británico, que al pasar produce en el piso movimientos verticales y movimientos horizontales paralelos a la dirección en que viaja. La otra es la onda de Love, llamadas así por el geofísico británico Augustus E. H. Love, cuya velocidad es intermedia entre las Secundarias y las de Rayleigh, y que produce solamente movimientos horizontales perpendiculares a la dirección de propagación.

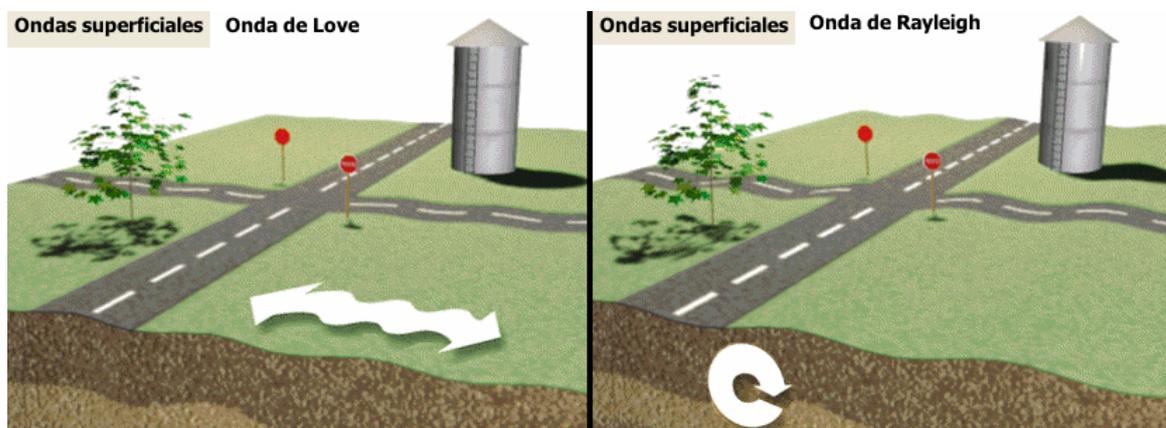


Figura 5 Tipos de Ondas Superficiales

³ Fuente: <http://www.obsis.unb.br/br/terremo.htm>

1.1.2.2 Medios de estudio

Las ondas sísmicas longitudinales, transversales y superficiales provocan vibraciones allí donde alcanzan la superficie terrestre. Los instrumentos sísmicos están diseñados para detectar estos movimientos con métodos electromagnéticos u ópticos. Los instrumentos principales, llamados sismógrafos, se han perfeccionado tras el desarrollo por el alemán Emil Wiechert de un sismógrafo horizontal, a finales del siglo XIX.

1.1.2.2.1 Sismógrafo

El sismógrafo es un instrumento que registra el movimiento del suelo causado por el paso de las ondas sísmicas producidas por un sismo. Los sismógrafos fueron ideados a fines del siglo pasado y perfeccionados durante este siglo hasta alcanzar un alto grado de perfeccionamiento electrónico; sin embargo, su principio básico no ha cambiado.

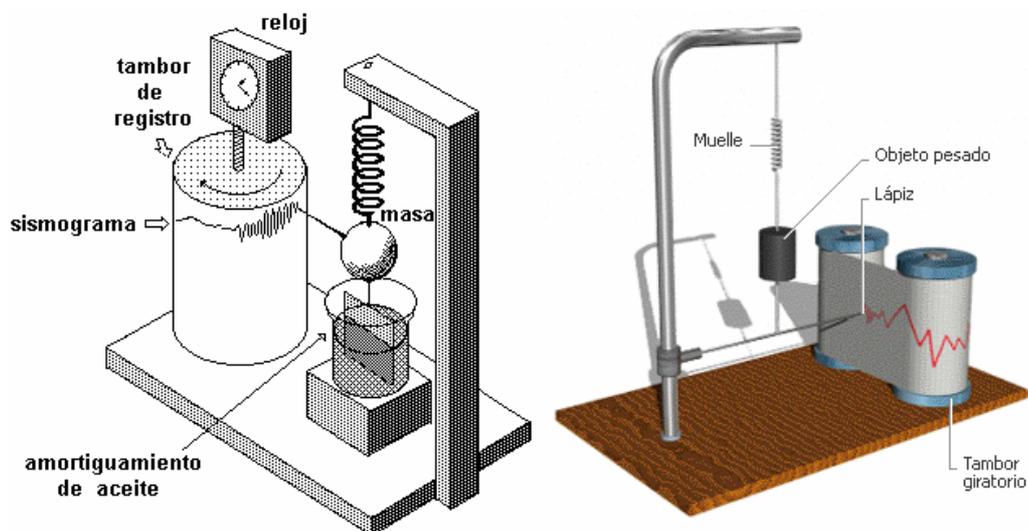


Figura 6 Sismógrafo ⁴

Para obtener el registro del movimiento del suelo es necesario tener como referencia un punto fijo al cual tendríamos que referirlo, lo cual sería imposible por cuanto el punto fijo se movería junto con el suelo. A fin de salvar esta dificultad habría que recurrir al principio de inercia de cualquier cuerpo y por el cual todos

⁴Fuente: <http://kaotica.no.sapo.pt/base1/sismografo1.gif>

los cuerpos tienden a resistirse al movimiento o a variar su velocidad. Por lo tanto, el movimiento del suelo puede ser medido en función de una masa suspendida por algún elemento que le permita mantenerse en reposo con respecto al suelo. El mecanismo más simple considera una masa suspendida de un resorte que está atado a un soporte anclado al suelo. Cuando el resorte se agita al paso de las ondas sísmicas, la inercia de la masa hace que ésta permanezca un instante en el mismo punto de reposo y cuando sale del mismo, tiende a oscilar. En los sismógrafos antiguos el amortiguamiento era realizado utilizando aceite y hoy en día se logra con el uso de bobinas o imanes que amortiguan la oscilación libre de la masa.

Si se adhiere a la masa suspendida un pincel o lápiz a fin de que inscriba en un papel sobre un cilindro que gira a tiempo constante, se registraría una componente del movimiento del suelo. El papel o lámina sobre el cual se registra el movimiento del suelo se llama sismograma. Debido a que el movimiento del suelo se realiza en tres dimensiones, es necesario registrarlo además de la componente vertical, en componentes horizontales con péndulos que oscilan en dirección similar al giro de una puerta con el eje ligeramente inclinado a fin de lograr mayor estabilidad en su movimiento.

Tal como se indicó, el movimiento del suelo con respecto a la masa suspendida se registraba inicialmente por medio de una pluma que inscribía en un tambor giratorio. Luego se implementó el registro en papel fotográfico o sobre película. Actualmente, el registro de los datos se realiza en medios magnéticos (cintas) o en sistemas digitales mejorando la calidad de los mismos. Asimismo, el control de tiempo es realizado mediante relojes satélites o los GPS.

En el mundo hay más de 1.000 observatorios sísmicos grabando continuamente, pero no todos son estaciones sísmicas independientes. En la década de 1960 se consiguió un importante avance con la construcción de estaciones sísmicas dispuestas en patrones geométricos (normalmente circulares o con forma de L). Ese tipo de formación de estaciones sísmicas operando a la vez, no sólo proporciona determinaciones más exactas de las características de los terremotos

(como su localización), sino que facilita información extra del camino que sigue la onda sísmica a lo largo de la formación.

Los gráficos producidos por los sismógrafos se conocen como sismogramas, y a partir de ellos es posible determinar el lugar y la intensidad de un terremoto. Muchos sismogramas son muy complicados y se requiere una técnica y experiencia considerables para interpretarlos, pero los más simples no son difíciles de leer. Éstos muestran las ondas primarias o de compresión (ondas P) del terremoto viajando en primer lugar (porque son las más rápidas), seguidas por las ondas secundarias o de cizalla (ondas S), más lentas, y, por último, por las ondas de superficie.



Figura 7 Registro de un sismo (Sismograma) ⁵

No todos los sismógrafos son de tipo pendular. En el sismógrafo de tensión, por ejemplo, un tubo rígido horizontal, normalmente de cuarzo fundido y de 20 a 50 m de largo, se fija a un estribo firmemente sujeto al suelo. El tubo se extiende casi hasta un segundo estribo y cuando pasa una onda sísmica los dos estribos se mueven el uno con respecto al otro y, por tanto, el hueco entre el tubo y el segundo estribo varía en anchura. Esta variación puede ser detectada y grabada por un aparato eléctrico que se conoce como transductor, situado en el hueco.

⁵ Fuente: <http://www.practiciencia.com.arsuperficie/endogenos/actisismica/terremoto/sismografo/>

1.1.2.3 Estructura Interna de la tierra

La Tierra está formada por tres capas concéntricas: corteza, manto y núcleo, con propiedades físicas distintas. Estas capas han podido ser detectadas y definidas, a partir del estudio de los registros del movimiento de su superficie, y mas concretamente por los estudios de los terremotos.

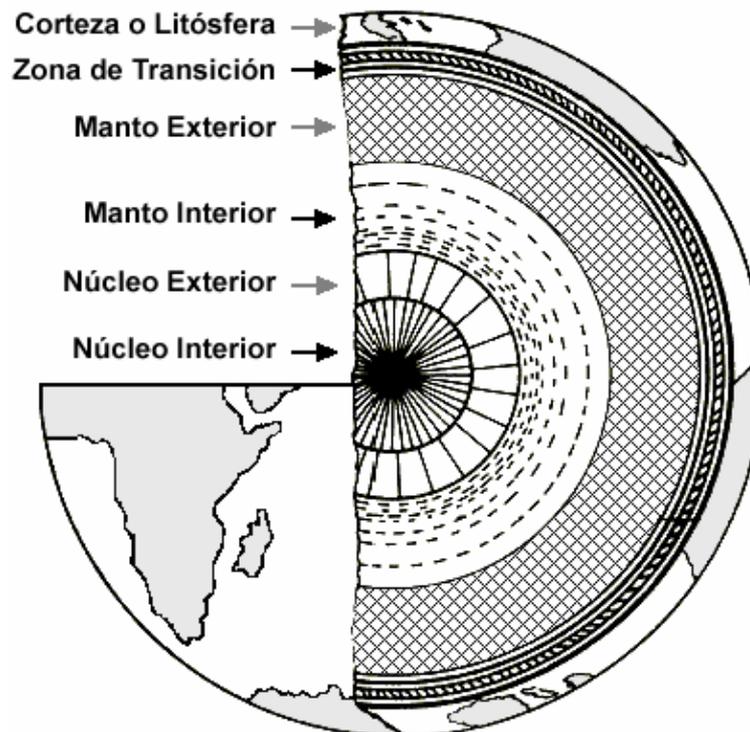


Figura 8 Principales capas que componen la Tierra⁶

Entre las principales capas que componen la Tierra tenemos a las siguientes:

1.1.2.3.1 Núcleo.

Con un radio de 3470 Km., esta constituido por núcleo interior formado por hierro, su consistencia es sólida y su temperatura es de 4300°C y núcleo exterior formado por hierro fundido, mezclado con pequeñas cantidades de níquel, sulfuros y silicio, su consistencia es líquido y su temperatura es de 3700°C.

⁶ Fuente: Autores

1.1.2.3.2 Manto.

Con un espesor de 2900 Km. y está dividido en manto interior, manto exterior, y zona de transición y esta formado por Magnesio, Hierro, Aluminio, Silicio y Oxígeno y su temperatura es de 100° C.

1.1.2.3.3 Corteza o Litosfera.

Es la capa exterior de la Tierra, es de elevada rigidez formada por roca, sabemos que es de espesor variable, que en algunos casos puede ser de 60 Km., en los continentes las formaciones son graníticas y basálticas en los fondos oceánicos.

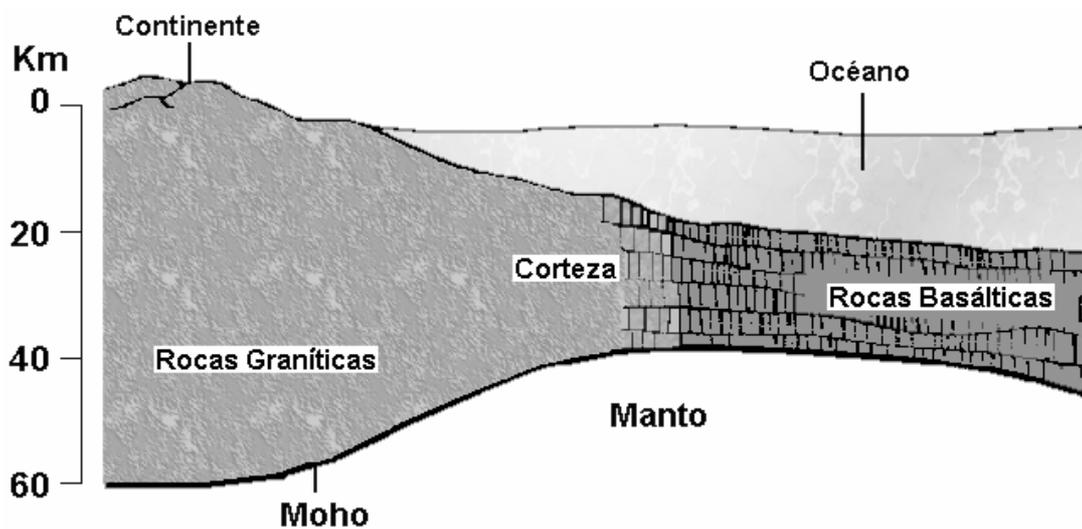


Figura 9 Corteza Terrestre⁷

1.1.2.4 Placas Tectónicas

Según la teoría de la tectónica de placas, la corteza terrestre está compuesta al menos por una docena de placas rígidas que se mueven independientemente. Estos bloques descansan sobre una capa de roca caliente y flexible, llamada astenosfera, que fluye lentamente a modo de alquitrán caliente. Los geólogos todavía no han determinado con exactitud como interactúan estas dos capas, pero las teorías más vanguardistas afirman que el movimiento del material espeso

⁷ Fuente: Autores

y fundido de la astenosfera fuerza a las placas superiores a moverse, hundirse o levantarse.

El concepto básico de la teoría de la tectónica de placas es simple: el calor asciende. El aire caliente asciende por encima del aire frío y las corrientes de agua caliente flotan por encima de las de agua fría; el mismo principio se aplica a las rocas calientes que están bajo la superficie terrestre: el material fundido de la astenosfera, o magma, sube hacia arriba, mientras que la materia fría y endurecida se hunde cada vez más hacia al fondo, dentro del manto; la roca que se hunde finalmente alcanza las elevadas temperaturas de la astenosfera inferior, se calienta y comienza a ascender otra vez. Este movimiento continuo y, en cierta forma circular, se denomina convección. En los bordes de la placa divergente y en las zonas calientes de la litosfera sólida, el material fundido fluye hacia la superficie, formando una nueva corteza.

Las placas tectónicas son gigantescos cascarones de la corteza terrestre, del tamaño de continentes, que se mueven unos hacia otros bajo la presión que ejercen sobre ellos los flujos de lava provenientes del núcleo del planeta Tierra.

En términos geológicos, una placa es una plancha rígida de roca sólida que conforma la superficie de la Tierra, flotando sobre la roca ígnea y fundida que conforma el centro del planeta.

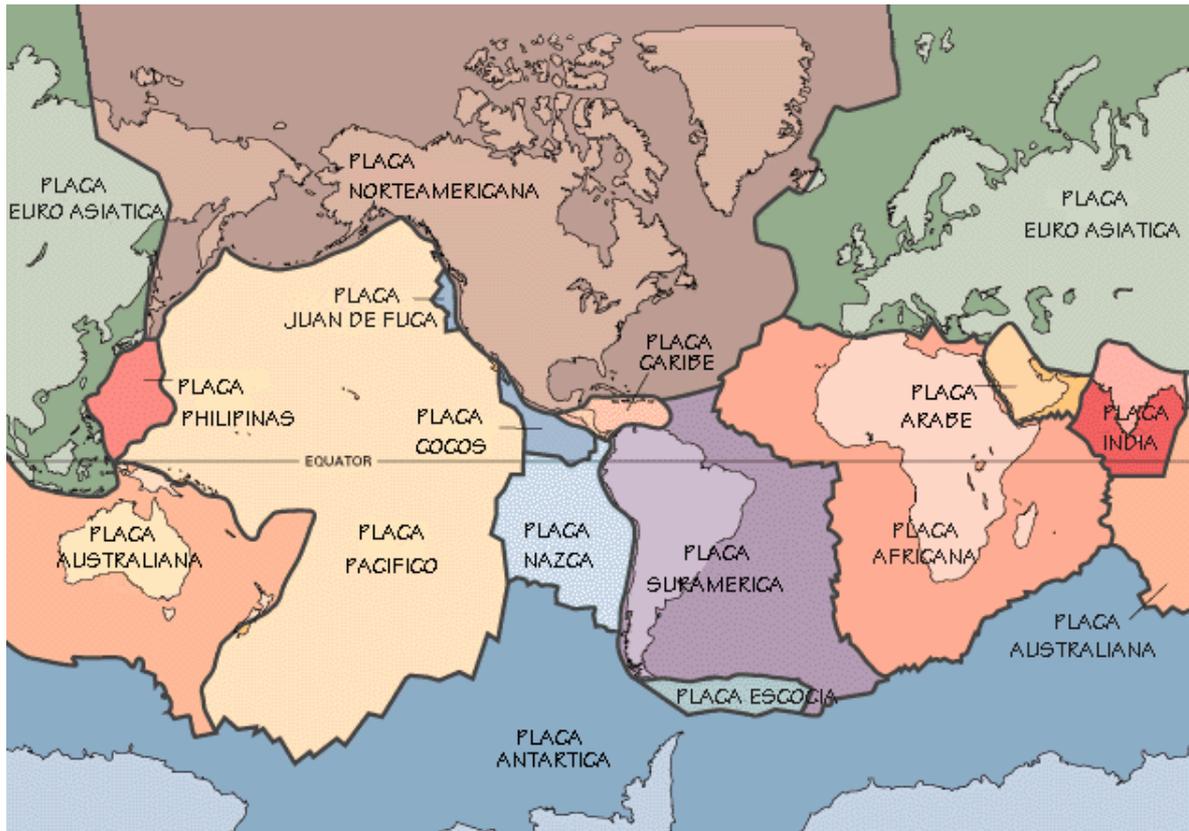


Figura 10 Placas Tectónicas⁸

Alfred Wegner en el año 1912 planteó que las doce grandes zonas de la corteza terrestre denominadas placas tectónicas, están en continua modificación y que los continentes se han formado a partir de un solo continente llamado Pangaea.

Los movimientos de deriva son los que han dado lugar a la formación de los actuales Continentes a partir del Pangaea.

⁸ Fuente: www.aldeaeduacativa.com

1.1.2.4.1 Interacción entre las Placas Tectónicas

Los modelos de Interacción entre las placas tectónicas son cuatro:

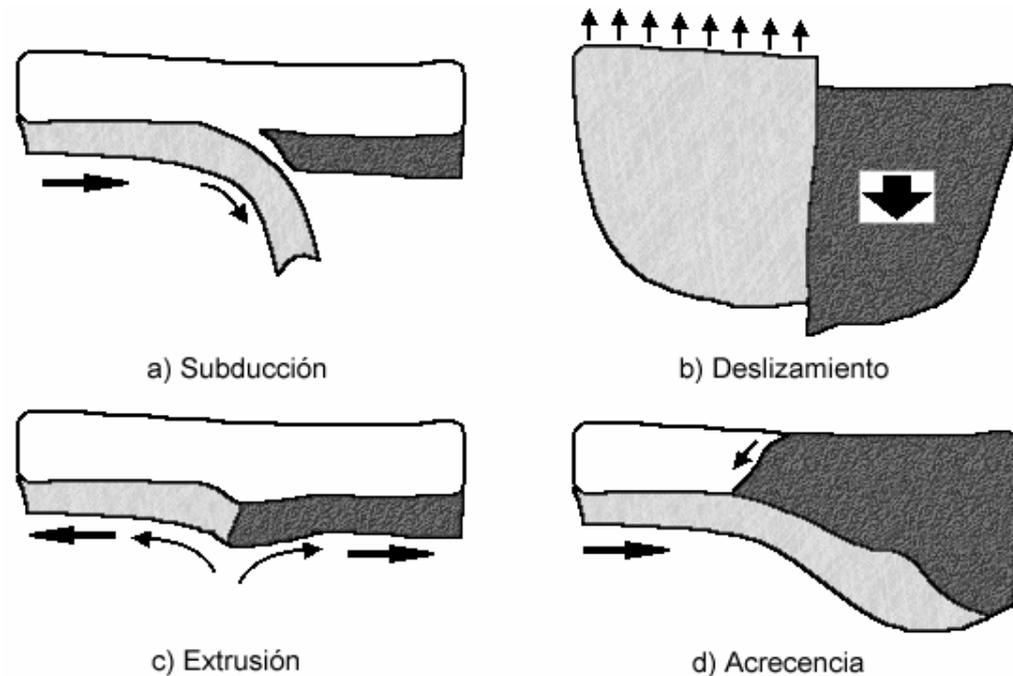


Figura 11 Interacción entre las placas tectónicas⁹

1.1.2.4.1.1 Subducción.

Ocurre cerca de las islas, donde dos placas de similar espesor entran en contacto entre sí.

1.1.2.4.1.2 Deslizamiento.

Se produce cuando entran en contacto dos placas oceánicas, o bien una continental y una oceánica.

1.1.2.4.1.3 Extrusión

Este fenómeno ocurre cuando se juntan dos placas tectónicas delgadas que se desplazan en direcciones opuestas, es el caso del contacto de dos placas del fondo del océano.

⁹ Fuente: Autores

1.1.2.4.1.4 Acrecencia

Tiene lugar cuando hay un impacto leve entre una placa oceánica y una continental.

Los movimientos de las placas están asociados con la energía calorífica que se concentra bajo la litosfera. En las zonas de extrusión aparece "nueva corteza", mientras en las zonas de subducción las placas que penetran por debajo se funden, por efecto del calor desarrollado en la interacción entre placas bajo condiciones de presión elevada, dando lugar al magma. Por ello los volcanes activos se sitúan frecuentemente en estas zonas de subducción.

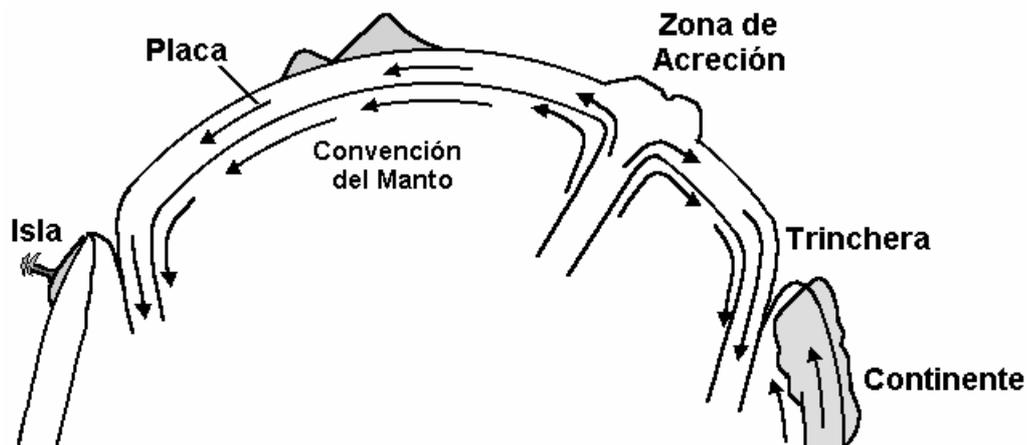


Figura 12 Esquema de los desplazamientos de las Placas Tectónicas¹⁰

1.1.2.5 Movimientos Sísmicos

Cuando se aplican esfuerzos sobre una roca ésta se deforma acumulando en su interior energía elástica de deformación, si dejamos de aplicar los esfuerzos, la roca recobrará su forma original; esto se conoce como comportamiento elástico de las rocas. Si mantenemos los esfuerzos aplicados sobre la roca durante muchos miles de años, la roca se deforma permanentemente, lo que se conoce como comportamiento plástico.

Dependiendo del tipo de roca y de las condiciones ambientales de temperatura y presión, ésta se comportará en forma más o menos elástica o plástica. Ante

¹⁰ Fuente: Autores

fuerzas que se aplican durante tiempos muy cortos todas las rocas se comportan de manera elástica; mientras que en el caso de fuerzas que actúan durante cientos o miles de años, todas actúan plásticamente.

Al aparecer un plano que está relativamente libre de esfuerzos el material localizado a ambos lados de él puede desplazarse con libertad y la roca vuelve a tomar aproximadamente su forma original en forma súbita, y este movimiento repentino de grandes masas de roca produce ondas elásticas, conocidas como ondas sísmicas, que viajan unas a través de y otras por la superficie de la Tierra, dando lugar a un sismo. Sismo es el término técnico para referirse a todos los temblores de tierra; a menudo se emplea la palabra terremoto para aludir a los grandes sismos.

La ruptura comienza en un punto y de allí se propaga o se extiende a puntos cercanos y de allí a otros hasta romper todo el plano de falla; este proceso se lleva a cabo en cuestión de fracciones de segundo en el caso de sismos pequeños y puede durar minutos enteros cuando se trata de grandes terremotos. La posición del punto inicial de una ruptura sísmica se llama hipocentro y el punto de la superficie terrestre situado arriba de él se llama epicentro; al volumen de roca cuyo desplazamiento causó el sismo y dentro del cual se encuentra la falla, se le llama fuente o foco sísmico.

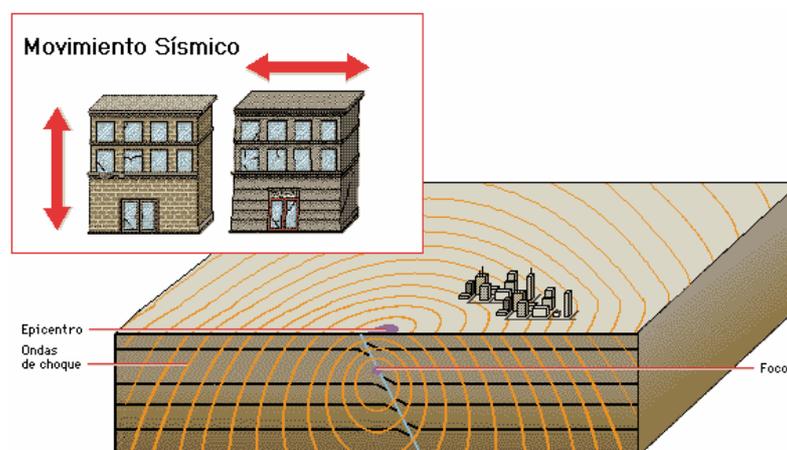


Figura 13 Movimientos Sísmicos¹¹

¹¹ Fuente: <http://www.obsis.unb.br/br/terremo.htm>

1.1.2.6 Fallas Tectónicas

Una roca se comporta elásticamente mientras las deformaciones producidas por el esfuerzo aplicado sean relativamente pequeñas. Si el esfuerzo aplicado es tan grande que produce deformaciones demasiado grandes, la roca se rompe y se dice que falla, esta ruptura es súbita y ocurre a lo largo de planos llamados planos de falla o simplemente fallas tectónicas.

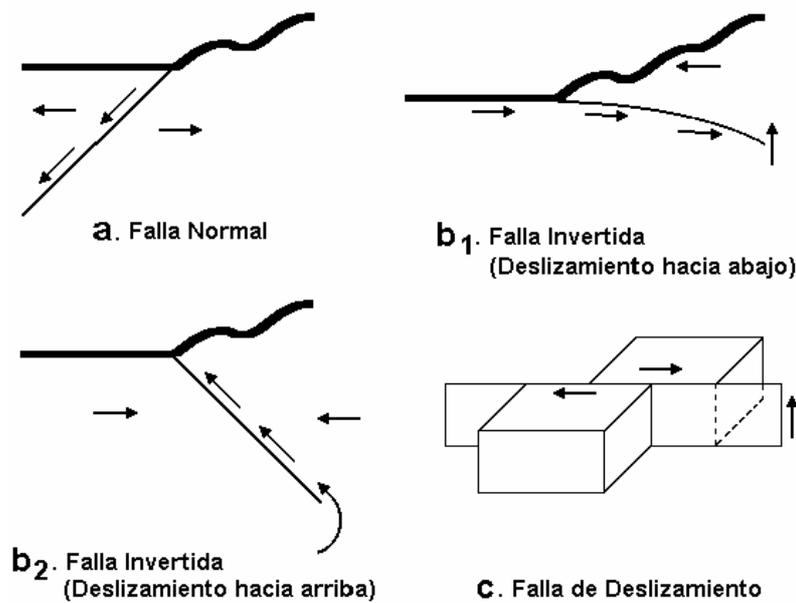


Figura 14 Tipos de Fallas Tectónicas¹²

1.1.2.6.1 Tipos de fallas

Los tipos de fallas tectónicas más importantes son las siguientes:

1.1.2.6.1.1 Falla normal.

Que corresponde a las zonas donde la corteza terrestre está en extensión, uno de los dos bloques de la falla se desliza hacia abajo.

1.1.2.6.1.2 Falla invertida.

Que corresponden a las zonas en compresión, existen dos casos:

¹² Fuente: Autores

1.1.2.6.1.3 Deslizamiento hacia abajo

Una de las dos porciones de corteza que están en contacto penetra bajo la otra que en general es una placa continental.

1.1.2.6.1.4 Deslizamiento hacia arriba

Una de las placas se desliza hacia arriba de la otra.

1.1.2.6.1.5 Falla de deslizamiento.

Que implica deslizamientos horizontales entre los dos bordes de la falla.

1.1.2.7 Terremotos y Zonas Sísmicas

Los terremotos pueden definirse como movimientos caóticos de la corteza terrestre, caracterizados por una dependencia en el tiempo de amplitudes y frecuencias. Un terremoto se produce debido a un choque producido a una cierta profundidad bajo la superficie terrestre en un determinado punto llamado foco o hipocentro. A la proyección del foco sobre la superficie terrestre se le denomina epicentro.

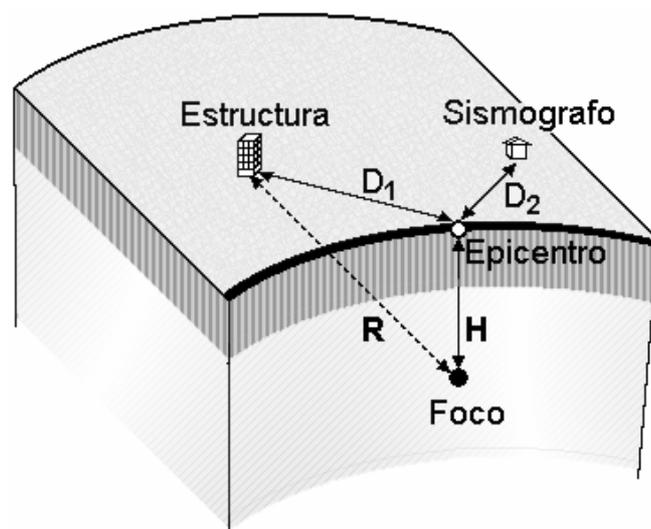


Figura 15 Características de un Terremoto¹³

¹³ Fuente: Autores

Las principales zonas sísmicas del mundo coinciden con los contornos de las placas tectónicas y con la posición de los volcanes activos de la Tierra. Esto se debe al hecho de que la causa de los terremotos y de las erupciones volcánicas están fuertemente relacionadas con el proceso tectónico del Planeta.

Los tres principales cinturones sísmicos del Mundo son: el Cinturón de Fuego del Pacífico, el Cinturón Trans asiático (Himalaya, Irán, Turquía, Mar Mediterráneo, Sur de España) y el Cinturón situado en el Centro del Océano Atlántico.

Al hablar de regiones sísmicas, hay que clarificar dos conceptos importantes. La intensidad sísmica es una medida de los efectos de los terremotos en el entorno y en particular sobre las estructuras. La sismicidad se define como la frecuencia de ocurrencia de fenómenos sísmicos por unidad de área incluyendo, al mismo tiempo cierta información de la energía sísmica liberada.

1.1.2.7.1 Tipos de Terremotos

La sismología distingue las siguientes clases de terremotos:

1.1.2.7.1.1 Terremotos de colapso.

Son terremotos de baja intensidad originados en cavidades subterráneas y debido al colapso de las mismas.

1.1.2.7.1.2 Terremotos de origen volcánico.

Las erupciones volcánicas y los terremotos tienen el mismo origen, pero además la explosión de gases en las erupciones volcánicas pueden originar terremotos que en general son de baja intensidad y que afectan a pequeñas superficies.

1.1.2.7.1.3 Terremotos tectónicos.

Son los de mayor intensidad y frecuencia, están originados por la rotura violenta de las masas rocosas a lo largo de las fallas o superficies de fractura.

1.1.2.7.1.4 Terremotos causados por explosiones.

El hombre produce explosiones que a veces se pueden detectar a distancias considerables originando sacudidas sísmicas que pueden afectar a las estructuras de algunos edificios.

1.1.2.8 Intensidad Sísmica

La intensidad sísmica está íntimamente relacionada con los efectos producidos por un terremoto en las reacciones de las personas, el grado de destrozos producidos en las construcciones y las perturbaciones provocadas en el terreno tales como grietas, deslizamientos o desprendimientos.

Describiendo de manera subjetiva el potencial destructivo de los sismos se han propuesto varias escalas para medir la intensidad, entre las principales escalas tenemos las siguientes:

1.1.2.8.1 Escala de Mercalli

La Escala de Mercalli es una escala de 12 puntos desarrollada para evaluar la intensidad de los terremotos a través de los daños causados a distintas estructuras. El nombre de dicha escala se debe al físico y geólogo italiano Giuseppe Mercalli. Unos años después de desarrollada fue modificada en 1931 por Harry O. Wood y Frank Neuman.

Se trata de una escala subjetiva de intensidad, porque evalúa la percepción humana del sismo, y se expresa en números romanos para diferenciarla de la Escala de Richter.

Los niveles bajos de la escala están asociados por la forma en que las personas sienten el temblor, mientras que los grados más altos se relacionan con el daño estructural observado.

Grado	Descripción
I. Muy débil	No se advierte sino por unas pocas personas y en condiciones de perceptibilidad especialmente favorables.
II. Débil	Se percibe sólo por algunas personas en reposo, particularmente aquellas que se encuentran ubicadas en los pisos superiores de los edificios.
III. Leve	Se percibe en los interiores de los edificios y casas.
IV. Moderado	Los objetos colgantes oscilan visiblemente. La sensación percibida es semejante a la que produciría el paso de un vehículo pesado. Los automóviles detenidos se mecen.
V. Bastante fuerte	La mayoría de las personas lo percibe aún en el exterior. Los líquidos oscilan dentro de sus recipientes y pueden llegar a derramarse. Los péndulos de los relojes alteran su ritmo o se detienen. Es posible estimar la dirección principal del movimiento sísmico.
VI. Fuerte	Lo perciben todas las personas. Se siente inseguridad para caminar. Se quiebran los vidrios de las ventanas y los objetos frágiles. Los muebles se desplazan o se vuelcan. Se hace visible el movimiento de los árboles o se les oye crujir.
VII. Muy fuerte	Los objetos colgantes se estremecen. Se experimenta dificultad para mantenerse en pie. Se producen daños de consideración en estructuras de albañilería mal construidas o mal proyectadas. Se dañan los muebles. Caen trozos de mampostería, ladrillos, parapetos, cornisas y diversos elementos arquitectónicos. Se producen ondas en los lagos.
VIII. Destructivo	Se hace difícil e inseguro el manejo de vehículos. Se producen daños de consideración y aún el derrumbe parcial

	en estructuras de albañilería bien construidas. Se quiebran las ramas de los árboles. Se producen cambios en las corrientes de agua y en la temperatura de vertientes y pozos.
IX. Ruinoso	Pánico generalizado. Todos los edificios sufren grandes daños. Las casas sin cimentación se desplazan. Se quiebran algunas canalizaciones subterráneas, la tierra se fisura.
X. Desastroso	Se destruye gran parte de las estructuras de albañilería de toda especie. El agua de canales, ríos y lagos sale proyectada a las riberas.
XI. Muy desastroso	Muy pocas estructuras de albañilería quedan en pie. Los rieles de las vías férreas quedan fuertemente deformados. Las cañerías subterráneas quedan totalmente fuera de servicio.
XII. Catastrófico	El daño es casi total. Se desplazan grandes masas de roca. Los objetos saltan al aire. Los niveles y perspectivas quedan distorsionados.

Tabla 1.1 Escala de Mercalli¹⁴

1.1.2.8.2 Escala de Richter

La escala sismológica de Richter es una escala logarítmica nombrada así en honor al sismólogo estadounidense Charles Richter.

Richter desarrolló su escala en 1930, en la que calculó que la magnitud de un terremoto puede ser medido mediante la siguiente fórmula:

$$M = \log_{10}A(\text{mm}) + 3\log_{10}(8\Delta t(\text{s})) - 2.92$$

¹⁴ Fuente: IGEPN

Donde: A es la amplitud en milímetros y t el tiempo en segundos, en la cual se asigna una magnitud arbitraria pero constante a terremotos que liberan la misma cantidad de energía. El uso del logaritmo en la escala es para reflejar la gran cantidad de energía que se desprende en un terremoto. El logaritmo incorporado a la escala hace que los valores asignados a cada nivel aumenten de forma exponencial, es decir con un multiplicador, y no de forma lineal.

A continuación se muestra una tabla de las magnitudes de la escala:

Magnitud en Escala Richter	Efectos del terremoto
Menos de 3.5	Generalmente no se siente, pero es registrado
3.5 - 5.4	A menudo se siente, pero sólo causa daños menores
5.5 - 6.0	Ocasiona daños ligeros a edificios
6.1 - 6.9	Puede ocasionar daños severos en áreas muy pobladas.
7.0 - 7.9	Terremoto mayor. Causa graves daños
8 o mayor	Gran terremoto. Destrucción total a comunidades cercanas.

Tabla 1.2 Escala de Richter¹⁵

¹⁵ Fuente: IGEPN

1.2 ESTUDIO DE LA SITUACION ACTUAL DEL AREA DE SISMOLOGÍA DEL INSTITUTO GEOFÍSICO DE LA ESCUELA POLITÉCNICA NACIONAL

1.2.1 ASPECTOS GENERALES

1.2.1.1 Descripción

El Instituto Geofísico de la Escuela Politécnica Nacional se constituye como el principal Centro de Investigación existente en el país para en el diagnóstico y la vigilancia de los peligros sísmicos y volcánicos, los cuales pueden causar gran impacto en la población, en los proyectos de inversión y en el entorno natural.

Conjuntamente con el diagnóstico de la amenaza, el Instituto Geofísico mantiene un activo programa de monitoreo instrumental en tiempo real, que asegura la vigilancia científica permanente sobre volcanes activos y fallas tectónicas en el territorio nacional.

Una serie de publicaciones tanto científicas como de divulgación general a nivel nacional e internacional dan fe de la capacidad y mística de trabajo de los científicos y técnicos que conforman el Instituto.

En los últimos dos años el Instituto ha cobrado una gran trascendencia a nivel nacional e internacional al haberse reactivado los volcanes Guagua Pichincha y Tungurahua, con el consiguiente peligro para la ciudadanía asentada en las zonas de afectación de los productos eruptivos de dichos volcanes. Gracias la Red Nacional de Sismógrafos y a la Red de Observatorios Volcánicos, el Instituto ha podido emitir alertas tempranas para que las autoridades y la población puedan tomar a tiempo las medidas preventivas correspondientes, en base a los mapas de peligro previamente establecidos por sus científicos.

El Instituto ha realizado además una serie de proyectos de evaluación del peligro sísmico y del peligro volcánico; así como también la evaluación de la vulnerabilidad y de la mitigación del riesgo sísmico y volcánico en los asentamientos humanos.

1.2.1.2 Objetivos

- Reducción del impacto de desastres sísmicos y volcánicos

- Difusión de los resultados de la vigilancia e investigación de la actividad sísmica y volcánica y comunicación de recomendaciones a entidades locales y público en general
- Servicio Nacional de Vulcanología
- Servicio Nacional de Sismología
- Emisión de pronósticos y alertas a las autoridades y población

El Instituto Geofísico - Departamento de Geofísica - de la Escuela Politécnica Nacional es el encargado según el artículo primero del decreto ejecutivo **No. 3593** del 20 de enero de 2003, de las siguientes actividades:

- a) Vigilancia, detección y comunicación mediante la red nacional de sismógrafos de los movimientos sísmicos ocurridos en el territorio nacional, así como el estudio e investigación de la sismicidad con fines de reducción del riesgo sísmico;
- b) Vigilancia, detección y comunicación mediante la red de observatorios volcánicos de las erupciones ocurridas en el territorio nacional, así como el estudio de investigación del volcanismo activo con fines de reducción del riesgo volcánico; y;
- c) Identificación de las amenazas volcánicas y sísmicas y preparación de los mapas de peligro respectivos.

1.2.1.3 Misión

A través del estudio científico y de la aplicación tecnológica, comprender el origen, los procesos y los efectos de los fenómenos naturales fundamentalmente sísmicos y volcánicos y, propiciando la concientización y preparación ciudadana y estatal, lograr reducir su impacto en la sociedad mediante la incorporación del concepto y las acciones de prevención en todos los ámbitos del quehacer individual, local y nacional, de esta manera propendiendo al desarrollo sostenible y sustentable del país.

1.2.1.4 Visión

Instituto autogestionario, planificado, dinámico y optimista, líder en el país y en la región en el estudio científico de los fenómenos sísmicos y volcánicos para la reducción de su impacto, promotor de políticas de Estado para la creación de una

cultura de prevención, asesor de diversas instancias gubernamentales en los temas contemplados en su misión.

1.2.2 ESTRUCTURA COMPUTACIONAL

1.2.2.1 Recursos de Hardware

1.2.2.1.1 Recursos de Red

Internamente el Instituto Geofísico posee una topología tipo estrella, esta topología se la muestra a continuación, junto con un breve resumen de los concentradores y dispositivos de interconexión.

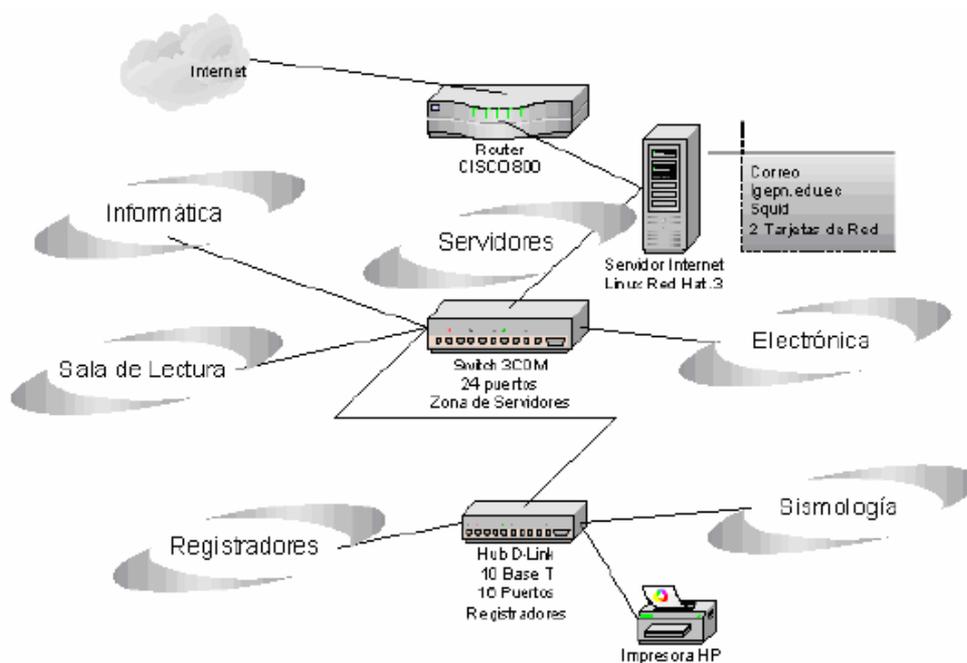


Figura 16 Diagrama de la topología de la red del Instituto Geofísico.¹⁶

Equipos	Cantidad
Hubs	5
Switches	2
Routers	1

Tabla 1.3 Equipos de red del Instituto Geofísico¹⁷

¹⁶ Fuente: IGEPN

1.2.2.1.2 Recursos de Hardware

Dentro de los recursos de hardware que posee el Instituto se ha podido extraer el siguiente cuadro explicativo.

Equipos	Cantidad
Servidores	1
PC's	36
Impresoras	7
Scanners	2
Plotters	1
Unidades ZIP	3
CD Writers	5
UPS	2
PC Portatiles	7

Tabla 1.4 Recursos de Hardware del Instituto Geofísico¹⁸

1.2.2.2 Recursos de Software

Dentro de los recursos de software que posee el Instituto se ha podido extraer el siguiente cuadro:

Sistemas Operativos
DOS 6.2
Windows 95
Windows 98
Windows 2000 Professional
Windows 2000 Server
Windows XP
Linux Fedora Core 1
Linux Red Hat 6.0
Linux Suse 9.0

¹⁷ Fuente: IGEPN

¹⁸ Fuente: IGEPN

Mac OS

Tabla 1.5 Software Base del Instituto Geofísico¹⁹

Software Aplicativo	
Sísmica	Earthworm
	ACQ 3
	Sismalp
	Scream
	Software de Marslite
	Piqev
	Triev
	Tensor
Lahares y fluviómetros	OPC Systems
	Sammy
Inclinometría	Tilt
	RSam, Bat
Térmica	Termal CAM
GIS	ArcGis
	ArcView
	AutoCad
	Goleen Surfer
Gráficos y Video	Adobe Illustrator
	Adobe Photoshop
	Pixela
Matemáticas	MatLab
	Octave
Programación y	Visual Studio

¹⁹ Fuente: IGEPN

Desarrollo	Eclipse
	Java

Tabla 1.6 Software Aplicativo del Instituto Geofísico²⁰

Utilitarios
Microsoft Office
Adobe Acrobat
Wsftp
Winfax
Winzip

Tabla 1.7 Software de Utilitarios del Instituto Geofísico²¹

1.3 DATA WAREHOUSE Y DATA MART

1.3.1 INTRODUCCION

En la actualidad toda empresa pública o privada necesita herramientas que le permitan depositar mucha confianza en la toma de decisiones sobre los negocios que efectúa, para la toma de estas decisiones se necesita de hechos y cifras, ya que además la competitividad en los negocios crece aceleradamente entonces las decisiones que debemos tomar en nuestra empresa deben ser mas rápidas y deben estar basadas en buenos cimientos y para esto debemos manejar y analizar adecuadamente la información que posee la empresa en el menor tiempo posible ya que las demoras pueden causar incluso el cierre del negocio.

Por ese motivo se requieren herramientas que nos ayuden a minimizar el tiempo para analizar mucha información con mayor velocidad y precisión ya que utilizando dichas herramientas logramos mantenernos competitivos debido a que nuestros negocios deben reaccionar de acuerdo al cambio del mercado. De otro modo el mercado globalizado, la presión inmensa de la competencia, los arranques tecnológicos entre otros cambios drásticos a que se han dado, irán debilitando nuestra empresa. Esto nos muestra que las empresas en la actualidad

²⁰ Fuente: IGEPN

²¹ Fuente: IGEPN

están invirtiendo en tecnología y soluciones con las cuales se mantienen competitivos en este mundo cambiante, ahora las empresas no dependen tan solo de factores como ubicación, productos o marketing. Sino también del conocimiento basado en información comprensible, detallada y relevante que es crucial para lograr y sostener ventajas competitivas. El poseer conocimientos correctos significa tener respuestas correctas y realizar decisiones estratégicas para su ejecución en beneficio de la empresa. Pero las tareas de recolectar, procesar, limpiar y transformar la información necesaria para la toma de decisiones no es una tarea sencilla mas si consideramos que una empresa tiene distintas áreas que a veces se encuentran alejadas de los ejecutivos de negocios. El Componente de Inteligencia de Negocios (Business Intelligence) que resuelve este caos en el manejo y análisis la información o datos es el Data Warehouse. El Data Warehouse es un conjunto de procesos y acciones, es una colección de datos orientados a un tema, integrados y no volátiles en el soporte al proceso de toma de decisiones de la gerencia.

1.3.2 DATA WAREHOUSE

1.3.2.1 Definiciones Generales

A continuación veremos algunas definiciones que diferentes especialistas en el tema han dado sobre Data Warehouse:

Data Warehouse es el centro de la arquitectura para los sistemas de información en la década de los '90. Soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis. Facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se necesitan para el procesamiento analítico, informático sobre una amplia perspectiva de tiempo.

Un Data Warehouse es una colección de datos orientado a temas, integrado, no volátil, de tiempo variante que es considerada la solución integral y oportuna para desarrollar negocio y se usa para el soporte del proceso de toma de decisiones gerenciales.

Data Warehouse es una herramienta con procesos para consolidar y administrar datos de variadas fuentes con el propósito de responder preguntas de negocios y tomar decisiones.

Data Warehouse es un proceso que agrupa datos desde múltiples fuentes heterogéneas, incluyendo datos históricos para soportar la continua necesidad de consultas, reportes analíticos y soporte de decisiones.

John Edwards en un artículo reciente de CIO Magazine, define DATA WAREHOUSE como: "Un Data Warehouse toma información de múltiples sistemas y bases de datos y la almacena de una manera que está diseñada para dar a los usuarios acceso más rápido, más fácil y más flexible a los aspectos claves"

Susan Osterfeldt publicó en 1993 una definición que sin duda nos describe más fielmente lo que es el Data Warehouse: "Yo considero al Data Warehouse como algo que provee dos beneficios empresariales reales: Integración y Acceso de datos. Data Warehouse elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operacional clásico"

Existen muchas definiciones para el Data Warehouse pero creemos que la más conocida fue propuesta por W. H. Inmon que es considerado como el padre de las Bases de Datos en 1992: " Un Data Warehouse es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales".

1.3.2.2 Características

Un Data Warehouse tiene muchas características entre las que podemos destacar las siguientes:

- **Orientado al tema**

Una primera característica del Data Warehouse es que la información se clasifica en base a los aspectos que son de interés para la empresa. Siendo así, los datos tomados están en contraste con los clásicos procesos orientados a las aplicaciones.

El ambiente operacional se diseña alrededor de las aplicaciones y la base de datos combina estos elementos en una estructura que acomoda las necesidades de la aplicación.

La alineación alrededor de las áreas de los temas afecta el diseño y la implementación de los datos encontrados en el Data Warehouse. Las

principales áreas de los temas influyen en la parte más importante de la estructura clave.

Las aplicaciones están relacionadas con el diseño de la base de datos y del proceso. En Data Warehouse se enfoca el modelamiento de datos y el diseño de la base de datos. El diseño del proceso no es separado de este ambiente.

Las diferencias entre la orientación de procesos y funciones de las aplicaciones y la orientación a temas, radican en el contenido de la Data a nivel detallado. En el Data Warehouse se excluye la información que no será usada por el proceso de sistemas de soporte de decisiones, mientras que la información de las orientadas a las aplicaciones, contiene datos para satisfacer de inmediato los requerimientos funcionales y de proceso, que pueden ser usados o no por el analista de soporte de decisiones.

- **Integrado**

El aspecto más importante del ambiente Data Warehouse es que la información encontrada al interior está siempre integrada.

La integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples y otros.

A través de los años, los diseñadores de las diferentes aplicaciones han tomado sus propias decisiones sobre cómo se debería construir una aplicación. Se diferencian en la codificación, en las estructuras claves, en sus características físicas, en las convenciones de nombramiento y otros. La capacidad colectiva de muchos de los diseñadores de aplicaciones, para crear aplicaciones inconsistentes, es fabulosa.

- **De tiempo variante**

Toda la información del Data Warehouse es requerida en algún momento. Esta característica básica de los datos en un depósito, es muy diferente de la información encontrada en el ambiente operacional. En éstos, la información se requiere al momento de acceder. En otras palabras, en el ambiente operacional, cuando usted accede a una unidad de información, usted espera que los valores requeridos se obtengan a partir del momento de acceso.

Como la información en el Data Warehouse es solicitada en cualquier momento los datos encontrados en el depósito se llaman de "tiempo variante". Los datos históricos son de poco uso en el procesamiento operacional. La información del depósito por el contraste, debe incluir los datos históricos para usarse en la identificación y evaluación de tendencias como vemos en la figura a continuación.

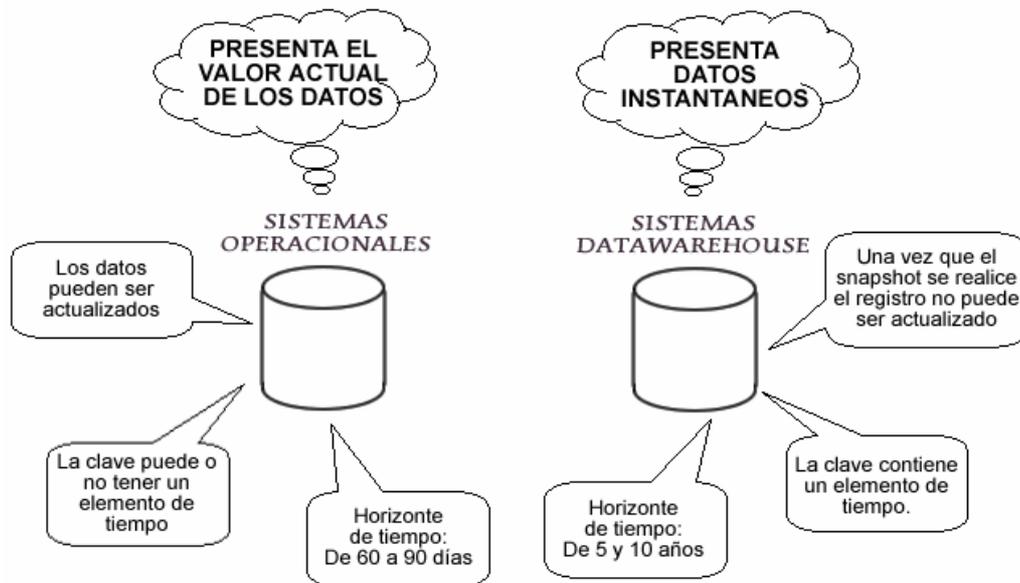


Figura 17 De tiempo variante²²

- **No volátil**

La información es útil sólo cuando es estable. Los datos operacionales cambian sobre una base momento a momento. La perspectiva más grande, esencial para el análisis y la toma de decisiones, requiere una base de datos estable.

En la siguiente Figura se muestra que la actualización es decir el proceso de insertar, borrar y modificar, se hace regularmente en el ambiente operacional sobre una base de registro por registro. Pero la manipulación básica de los datos que ocurre en el Data Warehouse es mucho más simple. Hay dos únicos tipos de operaciones: la carga inicial de datos y el acceso a los mismos. No hay actualización de datos en el depósito, como una parte normal de procesamiento.

²² Fuente: Autores

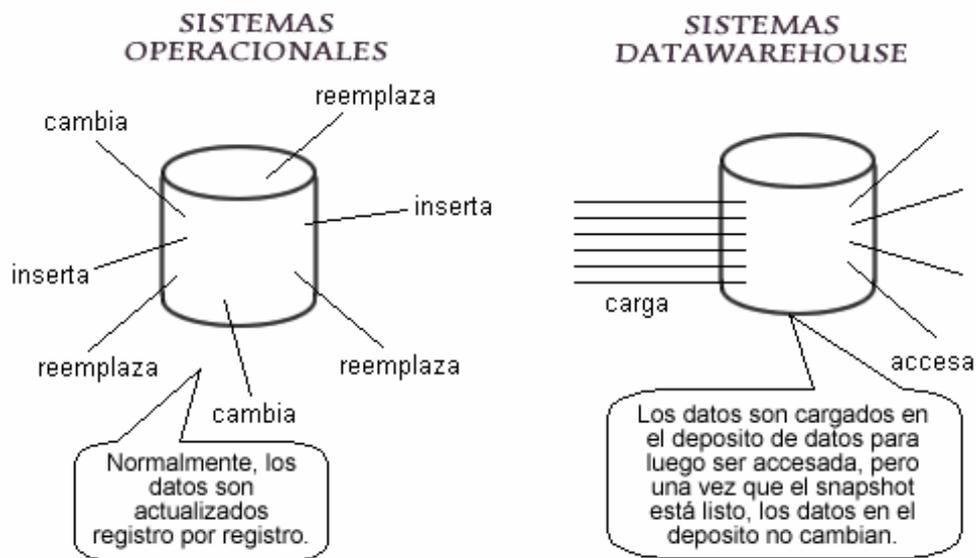


Figura 18 No volátil²³

Hay algunas consecuencias muy importantes de esta diferencia básica, entre el procesamiento operacional y del Data Warehouse. En el nivel de diseño, la necesidad de ser precavido para actualizar las anomalías no es un factor en el Data Warehouse, ya que no se hace la actualización de datos. Esto significa que en el nivel físico de diseño, se pueden tomar libertades para optimizar el acceso a los datos, particularmente al usar la normalización y denormalización física.

- **Escalable**

Cuando la organización está lista para implementar una solución de Data Warehouse, la solución necesita acomodarse al incremento dramático de la demanda de los datos. Como las instituciones crecen en otras áreas, la solución de Data Warehouse necesita localizar los nuevos orígenes de datos y debe variar en su tamaño de acuerdo a las necesidades.

1.3.2.3 Estructura del Data Warehouse

Los Data Warehouse tienen una estructura distinta. Hay niveles diferentes de esquematización y detalle que delimitan el Data Warehouse. La estructura de un Data Warehouse se muestra a continuación:

²³ Fuente: Autores

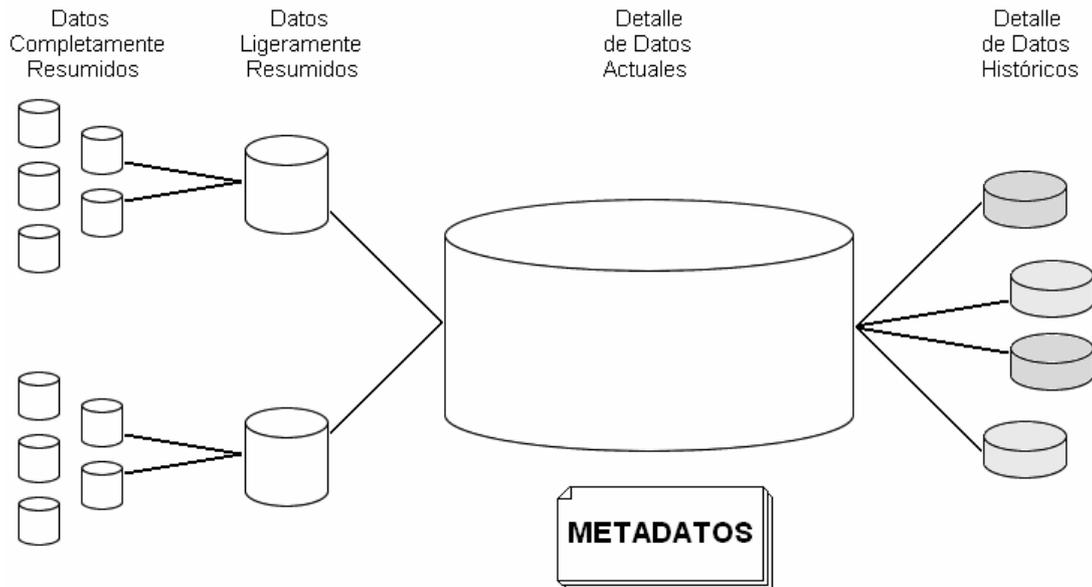


Figura 19 Estructura de los datos de un Data Warehouse²⁴

En la figura, se muestran los diferentes componentes del Data Warehouse y son:

- **Datos completamente resumidos**

El siguiente nivel de datos encontrado en el Data Warehouse es el de los datos completamente resumidos. Estos datos son compactos y fácilmente accesibles.

A veces se encuentra en el ambiente de Data Warehouse y en otros, fuera del límite de la tecnología que ampara al Data Warehouse.

- **Datos ligeramente resumidos**

La Data ligeramente resumida es aquella que proviene desde un bajo nivel de detalle encontrado al nivel de detalle actual. Este nivel del Data Warehouse casi siempre se almacena en disco. Los puntos en los que se basa el diseñador para construirlo son:

Que la unidad de tiempo se encuentre sobre la esquematización hecha.

Qué contenidos (atributos) tendrá la Data ligeramente resumida.

- **Detalle de datos actuales.**

En gran parte, el interés más importante radica en el detalle de los datos actuales, debido a que:

Refleja las ocurrencias más recientes, las cuales son de gran interés

²⁴ Fuente: Autores

Es voluminoso, ya que se almacena al más bajo nivel de granularidad.

Casi siempre se almacena en disco, el cual es de fácil acceso, aunque su administración sea costosa y compleja.

- **Detalle de datos históricos**

La Data antigua es aquella que se almacena sobre alguna forma de almacenamiento masivo. No es frecuentemente accesada y se almacena a un nivel de detalle, consistente con los datos detallados actuales. Mientras no sea prioritario el almacenamiento en un medio de almacenaje alterno, a causa del gran volumen de datos unido al acceso no frecuente de los mismos, es poco usual utilizar el disco como medio de almacenamiento.

- **Metadatos**

El componente final del Data Warehouse es el de los Metadatos. De muchas maneras los Metadatos se sitúan en una dimensión diferente al de otros datos del Data Warehouse, debido a que su contenido no es tomado directamente desde el ambiente operacional.

Los Metadatos juega un rol especial y muy importante en el Data Warehouse y son usados como:

Un directorio para ayudar al analista a ubicar los contenidos del Data Warehouse.

Una guía para el mapping de datos de cómo se transforma, del ambiente operacional al de Data Warehouse.

Una guía de los algoritmos usados para la esquematización entre el detalle de datos actual, con los datos ligeramente resumidos y éstos, con los datos completamente resumidos, etc.

Los Metadatos juegan un papel mucho más importante en un ambiente Data Warehouse que en un operacional clásico.

Los Metadatos contienen al menos:

- La estructura de los datos
- Los algoritmos usados para la esquematización
- El mapping desde el ambiente operacional al Data Warehouse

La información adicional que no se esquematiza es almacenada en el Data Warehouse. En muchas ocasiones, allí se hará el análisis y se producirá un tipo u otro de resumen. El único tipo de esquematización que se almacena

permanentemente en el Data Warehouse, es el de los datos que son usados frecuentemente. En otras palabras, si un analista produce un resumen que tiene una probabilidad muy baja de ser usado nuevamente, entonces la esquematización no es almacenada en el Data Warehouse.

1.3.2.4 Usos de los Data Warehouse

El Data Warehouse puede ser utilizado en diferentes casos, entre los principales mencionaremos los siguientes:

Se tienen consultas muy complejas.

- Existen grandes cantidades de datos a los cuales deben ingresar los usuarios.
- Pocos usuarios acceden a los datos de manera concurrente.
- Las consultas que realizan los usuarios no tienen tiempos de respuesta críticos.

1.3.2.5 Arquitectura Data Warehouse

Dentro de arquitectura de un Data Warehouse, se define la estructura de los datos, así como todos los procesos que intervienen en el manejo de los mismos, así tenemos las comunicaciones, procesamiento y presentación de los datos.

- Base de datos Intermedia.
- Nivel de acceso a la información.
- Nivel de acceso a los datos.
- Metadatos.
- Gestión de proceso.
- Mensajes de la aplicación.
- Data Warehouse
- Organización de los datos.
- *Base de datos Intermedia.* Debido al gran volumen de información que se maneja en las organizaciones, es importante tener una base de datos intermedia a partir de la cual se va a cargar datos al Data Warehouse.
- *Nivel de acceso a la información.* Son los diferentes mecanismos que tienen los usuarios para ingresar a los datos, es decir para acceder a la información que se encuentra almacenada en el Data Warehouse. Existen

herramientas que permiten hacer esto, por ejemplo Microsoft Excel, Acces, etc.

- *Nivel de acceso a los datos.* Las diferentes maneras por cuales se va a acceder a los datos fuente para desplegarlos hacia los usuarios finales. Se tienen diferentes mecanismos para realizar este trabajo, así: DBMS.
- *Metadatos.* Información relevante respecto a los datos empresariales.
- *Gestión de proceso.* Tareas a realizar para construir el Data Warehouse.
- *Mensajes de la aplicación.* Información de cómo se transportan los datos a través de la red de datos empresarial.
- *Data Warehouse:* En el data warehouse (núcleo) es donde ocurre la data actual, usada principalmente para usos estratégicos. En algunos casos, uno puede pensar del data warehouse simplemente como una vista lógica o virtual de datos. En muchos ejemplos, el data warehouse puede no involucrar almacenamiento de datos.

En un data warehouse físico, copias, en algunos casos, muchas copias de datos operacionales y/o externos, son almacenados realmente en una forma que es fácil de acceder y es altamente flexible. Cada vez más, los data warehouses son almacenados sobre plataformas cliente/servidor, pero por lo general se almacenan sobre mainframes.

- *Organización de Datos:* El componente final de la arquitectura data warehouse es la organización de los datos. Se llama también gestión de copia o réplica, pero de hecho, incluye todos los procesos necesarios como seleccionar, editar, resumir, combinar y cargar datos en el depósito y acceder a la información desde bases de datos operacionales y/o externas. La organización de datos involucra con frecuencia una programación compleja, pero cada vez más, están creándose las herramientas data warehousing para ayudar en este proceso. Involucra también programas de análisis de calidad de datos y filtros que identifican modelos y estructura de datos dentro de la data operacional existente.

1.3.2.6 Ventajas y desventajas

El Data Warehouse, permite administrar la información en forma accesible, correcta, uniforme y actualizada, lo que ayuda a que los costos para la toma de

decisiones sean mucho menores, aumentando la flexibilidad en el entorno de trabajo, lo que permite el rediseño de los procesos. Esta es la principal ventaja para la utilización o implementación de un negocio con el Data Warehouse.

1.3.2.6.1 Ventajas

A continuación se desglosarán las ventajas más significativas para el uso de los Data Warehouse:

- Puedo ejecutar o procesar una gran cantidad de información.
- Es más rápida y flexible al momento de acceder a la información.
- Permite mejorar la toma de decisiones y el aumento de la productividad del negocio, puesto que permite conocer los resultados de la empresa, sean estos resultados positivos o negativos, evaluando de esta forma los tiempos de respuesta y los costos de operaciones.
- Es fiable la comunicación entre los diferentes departamentos de la empresa.
- El Data Warehouse permite realizar los planes de la empresa en forma más efectiva, lo que ayuda para mejorar las relaciones entre el cliente y el proveedor.

1.3.2.6.2 Desventajas

Los Data Warehouse presentan también algunas desventajas, las cuales se describen:

- El Data Warehouse realiza la reestructuración de los sistemas operacionales, lo que implica altos costos.
- Presenta revisiones continuas de los modelos de datos, objetos, transacciones, lo que provoca que el diseño sea complejo.
- Es necesario de aplicaciones, sistemas, y almacenamientos específicos.

1.3.3 DATA MART

1.3.3.1 Definiciones generales

Para entender el significado del Data Mart se consideran las siguientes definiciones:

- Los Data Mart son Data Warehouse departamentales construidos velozmente para brindar soluciones a un determinado negocio dentro de la institución.

- Son subconjuntos de Data Warehouse diseñados para satisfacer necesidades específicas de un área de la institución.
- Es la implementación de un Data Warehouse con funciones restringidas, que sirve a un Departamento específico de la organización.
- Es un almacén de datos para dar soporte a un departamento o unidad de negocio en particular. Puede ser independiente, parte de una red distribuida de Data Mart o dependiente de los datos de un Data Warehouse.²⁵
- El concepto de Data Mart es el de una tienda local, que da servicio a un vecindario, en vez de una gran tienda departamental que atiende a un suburbio o ciudad completos.²⁶

1.3.3.2 Características

Al igual que el Data Warehouse, los Data Mart tienen las mismas características de integración, no volatilidad y orientación temática, referirse al ítem 1.3.1.2 Características.

1.3.3.3 Tipos de Data Mart

Se definen dos tipos de Data Mart, los dependientes y los independientes:

Dependientes: Son los que se construyen a partir de un Data Warehouse central, es decir reciben sus datos de un repositorio empresarial central.

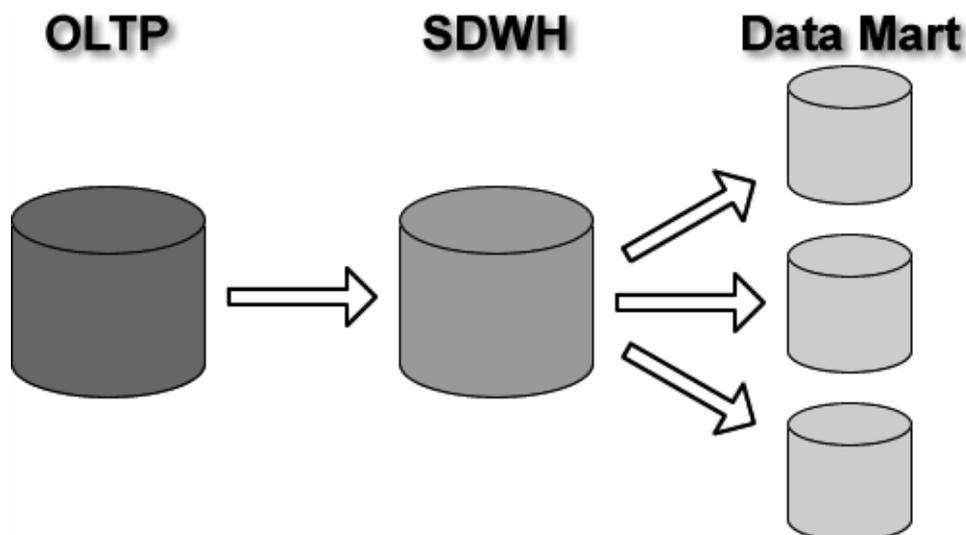


Figura 20 Data Marts Dependientes²⁷

²⁵ Global Communications, Enero 1999

²⁶ Data Warehousing – Cliente/Servidor, Harjinder S. Gill y Prakash C. Rao.

Independientes: Son aquellos Data Mart que no dependen de un Data Warehouse central, ya que pueden recibir los datos directamente del ambiente operacional, ya sea mediante procesos internos de las fuentes de datos o de almacenes de datos operacionales (ODS).

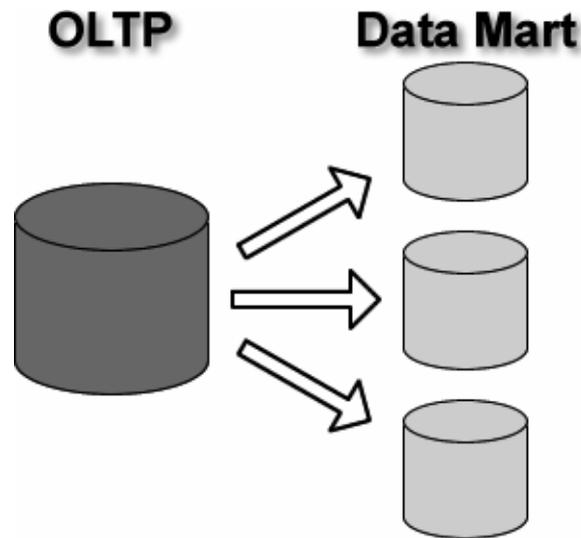


Figura 21 Data Marts Independientes²⁸

1.3.3.4 Carga de datos en Data Mart

Para la carga de datos hacia el Data Mart se pueden utilizar técnicas de carga para las herramientas OLAP, pero se debe tener en cuenta la capacidad de soportar extracción de grandes volúmenes de datos desde las fuentes, para así no sobrecargar las mismas. Además los tiempos de carga y la calidad de los datos a ser transportados hacia el Data Mart.

Una de las herramientas para la carga de datos en Microsoft SQL Server son los DTS, con los cuales podemos realizar la carga con solo escoger las fuentes y los destinos

Se debe tener cuidado que los datos sean coherentes es decir que los datos que extraemos sean los que se cargan.

Esta Fase comprende: preparación, integración, alto nivel de agregación y customización de los datos

²⁷ Fuente: Autores

²⁸ Fuente: Autores

1.3.3.5 Procesos de depuración

Este proceso se encarga de las siguientes tareas:

- Identificar anomalías; eliminar inconsistencias y detectar datos incorrectos
- Desarrollar una estrategia y metodología para mantener el sistema de detección y de limpieza de datos

1.3.3.5.1 Modos de Procesamiento de depuración

Se puede usar software para acceder y manipular los datos de orígenes diversos para su transferencia a múltiples plataformas

El software debe poder ser personalizado para la detección automática de anomalías, inconsistencias, limpiar y transformar los datos de acuerdo a reglas preestablecidas

1.3.3.6 Procesos de refrescamiento

Similar a la fase de carga, con la diferencia que los cambios son a bajo nivel.

Depende de varios parámetros: aplicaciones, restricciones de las fuentes de datos y limitantes en el sistema.

El refrescamiento puede ser:

- A los pocos segundos de haberse actualizado los datos base
- Primero se transforman los datos y luego se los carga en base a los requerimientos de almacenamiento.
- Los datos son extraídos diariamente o con alguna sincronización.

1.3.3.7 Ventajas y desventajas

1.3.3.7.1 Ventajas

A continuación se desglosarán las ventajas más significativas para el uso de los Data Marts:

- Dado que un DATA MART soporta menos usuarios que un DATA WAREHOUSE se puede optimizar para recuperar más rápidamente los datos que necesitan los usuarios.
- Menores cantidades de datos implican que se procesan antes, tanto las cargas de datos como las consultas.

- Las peticiones pueden acotarse al área o red que sirve esos datos, sin afectar al resto de los usuarios.
- La aplicación cliente, que pide la consulta es independiente del servidor que la procesa y del servidor de bases de datos que almacenan la información.
- Los costos que implica la construcción de un Data Mart son mucho menores a los de la implementación de un Data Warehouse.

1.3.3.7.2 Desventajas

- No permite el manejo de grandes volúmenes de información por lo que muchas veces se debe recurrir a un conjunto de Data Marts para cubrir todas las necesidades de información de la empresa.

1.4 METODOLOGÍA PARA EL DESARROLLO DE UN DATA MART

La metodología para la construcción de un Data Mart que a continuación exponemos se tomó de la combinación de distintas metodologías para desarrollo Data Warehouse que se encuentran explicadas en textos que enumeramos en nuestra bibliografía.

1.4.1 PLANIFICACIÓN

La planificación es la fase más importante para la construcción del Data Mart ya que aquí se determinarán todos y cada uno de los pasos a seguir para su desarrollo e implementación.

1.4.1.1 Seleccionar la estrategia de implementación

Antes de desarrollar el Data Mart, una de las tareas críticas es el desarrollo de una estrategia apropiada que se adapte a las necesidades de la empresa y de los usuarios de la misma. La decisión de que estrategia será la más conveniente, dependerá mucho de las características de la organización y de cómo se desarrollan las actividades dentro de la misma. Existen varios enfoques que pueden ser utilizados para la selección de la estrategia.

1.4.1.1.1 Enfoque de arriba hacia abajo

De necesidades generales a tareas específicas. Se basan en las metas que tiene la organización. Se comienza haciendo un diagrama generalizado de las tareas propias, obteniendo un nivel cada vez más bajo, en pocas palabras se subdivide a cada tarea en sub tareas hasta llegar a tareas mucho más específicas.

Materializar este tipo de enfoque en la práctica es muy difícil ya que en muchas ocasiones la organización no se encuentra muy al tanto de la tecnología y no tienen la suficiente experiencia en el desarrollo de aplicaciones. Una de las características de este enfoque conlleva a una poca consulta a los usuarios finales del Data Mart ya que su implementación se planifica principalmente en la cabeza de la organización.

1.4.1.1.2 Enfoque de abajo hacia arriba

Este enfoque parte de lo específico a lo general. Se comienza con elementos simples, encaminando la solución hacia arriba para buscar una mayor complejidad, es decir se selecciona un subconjunto del problema general, para luego ir hacia la solución organizacional global.

1.4.1.1.3 Enfoque combinado

El enfoque combinado como su nombre lo indica lleva una mezcla de los enfoques arriba – abajo y abajo – arriba.

Este enfoque tiene como principal ventaja la combinación de los anteriores, siendo complejo de manejar.

1.4.1.2 Seleccionar la metodología de desarrollo

Existen varias metodologías que pueden ser utilizadas para el desarrollo de un Data Mart. Para obtener los resultados esperados es necesario escoger una metodología que se adapte a los requerimientos de la organización.

1.4.1.2.1 Método de análisis y Diseño estructurado

Este método se basa en recolectar los requerimientos, a su vez analizarlos y satisfacerlos en base a módulos y prototipos.

El principal objetivo de éste método es obtener un producto que satisfaga los requerimientos partiendo únicamente de la información obtenida al principio de su implementación.

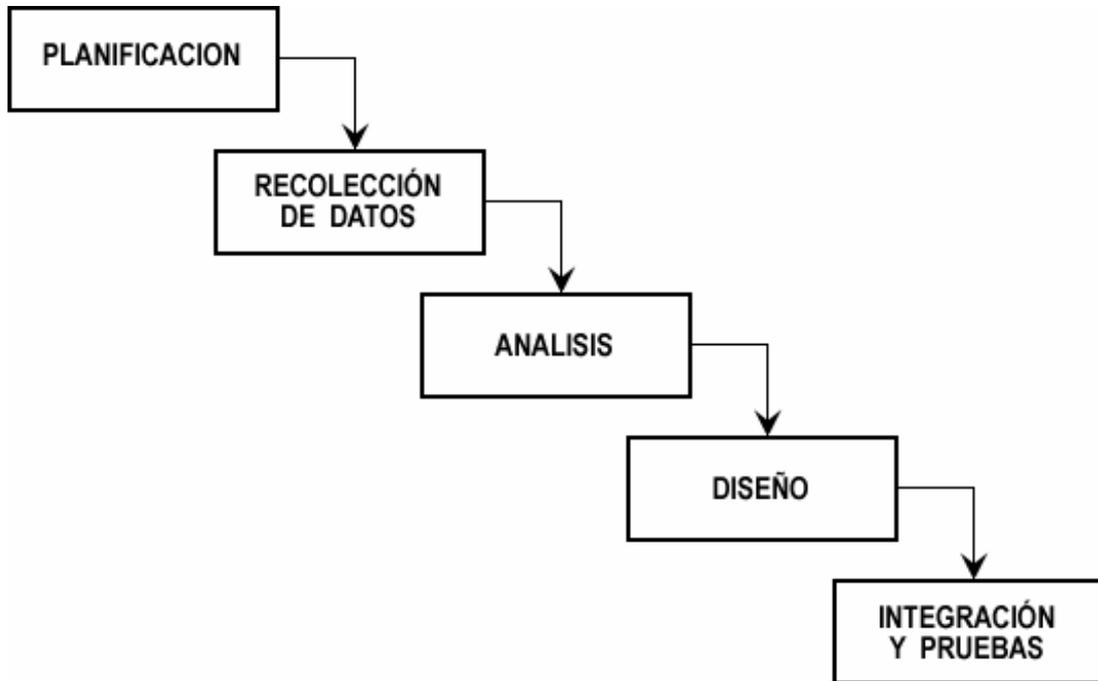


Figura 22 Método de Desarrollo Estructurado²⁹

1.4.1.2.2 Método Espiral

Este método se basa en el tiempo y en la velocidad de implementación, obteniendo una idea clara de los requerimientos que no puedan ser especificados al inicio del desarrollo del Data Mart.

Al momento de implementar un Data Mart debemos tener en cuenta que la flexibilidad a cambios, es decir la capacidad de soportar nuevas requerimientos, es muy importante.

Con éste método, lograremos combinar dos aspectos muy importantes en el desarrollo del Data Mart: Velocidad y Flexibilidad a cambios.

A continuación se presenta un gráfico descriptivo de las diferentes fases del método de Espiral:

²⁹ Fuente: Autores

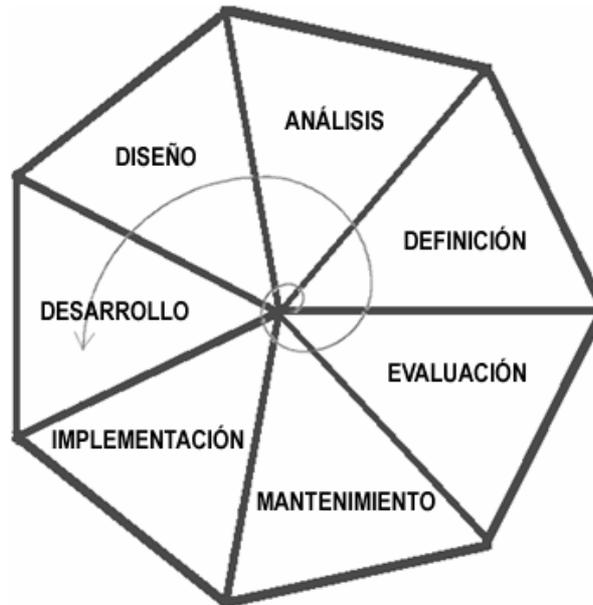


Figura 23 Método de Desarrollo en Espiral³⁰

El método de desarrollo en espiral está conformado por las siguientes fases:

1.4.1.2.2.1 Definición

Como su nombre lo indica en esta fase se definirá el porqué del Data Mart, es decir, es en ésta donde se obtiene una idea clara del objetivo y el alcance del Data Mart.

1.4.1.2.2.2 Análisis

Aquí se estudia de una manera más detallada los aspectos obtenidos en la etapa anterior, así como también los documentos adquiridos en la misma.

1.4.1.2.2.3 Diseño

Éste etapa es muy importante debido a que es aquí de donde se obtendrá el modelo de implementación, a partir del cual se empezará el desarrollo del Data Mart.

1.4.1.2.2.4 Desarrollo

Depende mucho de la fase de Diseño, es en ésta donde se Desarrolla el Diseño, se construyen las estructuras de datos

³⁰ Fuente: Autores

1.4.1.2.2.5 Implementación

En esta etapa del desarrollo en Espiral se obtiene el Data Mart, trabajando con el usuario final.

1.4.1.2.2.6 Mantenimiento

Esta etapa se trabaja fuera del proceso de desarrollo del Data Mart, ya que es una fase en la cual se implementan nuevas necesidades y requerimientos propios de cambios en la estructura de la organización, así como también variantes que vienen puedan ser necesarias luego de que se haya obtenido el Data Mart.

1.4.1.2.2.7 Evaluación

Aquí se evaluará el resultado obtenido luego de terminar todas y cada una de las etapas de la construcción del Data Mart, iniciando otra iteración, si ésta fuera necesaria. Es en esta fase crítica, en la cual el desarrollador optimizará trabajo, esfuerzo y recursos para la obtención de un nuevo producto.

1.4.1.3 Selección del ámbito de implementación

Para seleccionar el ámbito de implementación se ha dividido en dos categorías las mismas que se detallan a continuación:

- Ámbito considerando la perspectiva del usuario empresarial.
- Ámbito considerando la tecnología empresarial.

Se deben considerar:

- Hacia donde se quiere llegar con la implementación del Data Mart
- Que personas lo van a utilizar.
- Las consultas que se van a realizar.
- El tamaño de los Metadatos.
- El tamaño de los datos dentro del Data Mart.
- Fuentes de entrada.

1.4.1.4 Selección de la Arquitectura de implementación

En el análisis de un Data Warehouse se tomará en cuenta la flexibilidad arquitectónica la cual permite implementarla, considerando las siguientes opciones:

1.4.1.4.1 El almacenamiento operacional con uso de datos operacionales

Este tipo de arquitectura implementa un esquema virtual de Data Warehouse, direccionando sus consultas a los metadatos de las base de datos operacionales bajo la guía del meta modelo del Data Warehouse.

1.4.1.4.2 Solo Data Warehouse

Permite establecer una arquitectura independiente de la base de datos operacional, obteniendo un Data Warehouse centralizado con lo cual la depuración, integración, adición y demás tareas relacionadas con el manejo de las fuentes de la base de datos se optimizaran, debido a que la carga de estos datos se lo hará una sola vez, permitiendo trabajar directamente con ellos.

1.4.1.4.3 Solo Data Mart

Al utilizar este tipo de arquitectura, el Data Mart permite analizar independientemente a un departamento específico del negocio, asumiendo que este tiene sus propias necesidades, estructuras y requerimientos.

1.4.1.4.4 Data Warehouse y Data Mart

Para que el negocio permita satisfacer las necesidades del usuario final es necesario implementar el Data Warehouse corporativo y uno o varios Data Mart departamentales, esto quiere decir que el Data Warehouse controlará el flujo de datos de los Data Mart que componen el negocio, ayudando en la toma de decisiones.

1.4.1.5 Cronograma y Presupuesto del proyecto

En esta fase se procede a desarrollar un cronograma con todas y cada una de las actividades que se van a desempeñar durante el proceso de desarrollo del Data Warehouse.

Cabe señalar que el cronograma debe ser muy puntual y demás indicar recursos y presupuestos para obtenerlos.

Dentro del presupuesto del proyecto se debe indicar costos y justificarlos.

1.4.1.6 Recopilación de Metadatos

Un tema importante dentro de la planificación de un Data Warehouse se refiere a la recopilación de la información de los datos, debido a que partimos de aquí para construir el Data Warehouse.

Los metadatos pueden ser conseguidos de diferentes fuentes, a continuación daremos una breve definición de cada uno de ellos:

- Modelos empresariales contruidos por cada organización, debido a que tiene información relevante de las organizaciones.
- Diccionarios de datos que posean los administradores de los datos operacionales de las instituciones.
- Fuentes externas que sean de relevancia para la organización. Esta información es brindada por organismos externos como por ejemplo aquellos que regulan las entidades públicas y privadas como Contraloría, etc.

1.4.2 REQUERIMIENTOS

La fase de requerimientos de la implementación del Data Warehouse es una especificación de las funciones que tendrá el Data Warehouse. Además de las características y funciones necesarias, los requerimientos describirán con claridad el ambiente operativo en el que se entregará el Data Warehouse. La fase de definición de requerimientos es una transición de la visión del propietario.

El volumen de recolección de requerimientos va de acuerdo al enfoque de implementación que se utilice. Por ejemplo si vamos a usar el enfoque de arriba hacia abajo, la actividad de definición de requerimientos es muy grande pero se pueden catalogar con facilidad, mientras que si usamos el enfoque de abajo hacia arriba la mayoría de los requerimientos se definen en base a las expectativas que se tiene del proyecto y estos requerimientos son significativamente menos que los anteriores.

Una forma directa de definir los requerimientos es a través de los diferentes puntos de vista de la gente que esta involucrada en el proyecto de Data Warehouse.

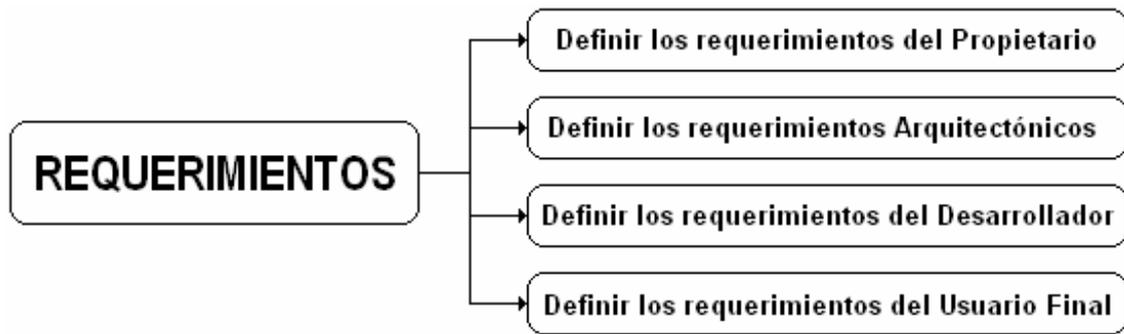


Figura 24 Requerimientos para un Proyecto de Data Warehouse³¹

A continuación detallaremos las diferentes perspectivas de la gente involucrada en un proyecto de Data Warehouse sobre la definición de requerimientos:

1.4.2.1 Definir los requerimientos generales o propios

Los propietarios del Data Warehouse proponen los siguientes cuestionamientos para su construcción:

- ¿Por qué construir un Data Warehouse o un Data Mart?
- ¿Qué problema empresarial abordará?
- ¿Cuáles son los objetivos empresariales?
- ¿Cuánto costará?
- ¿Cuándo estará listo?
- ¿Qué impacto tendrá sobre la gente?
- ¿Cuál es el impacto sobre la organización?
- ¿Tenemos la capacidad de hacerlo?
- ¿Cuáles son los riesgos?

La recolección de requerimientos debe dar respuestas a las preguntas anteriores. Estos requerimientos formarán entonces un criterio desde la perspectiva del propietario.

Las áreas potenciales de recolección de requerimientos son las siguientes:

- Objetivos empresariales
- Ámbito y objetivos del Data Warehouse, Data Mart y de los clientes
- Requerimientos del Cliente
- Fuentes de datos

³¹ Fuente: Data Warehousing Cliente/Servidor; Harjinder S. Gill

- En los distintos planes de la empresa, en los calendarios, presupuestos o recursos.
- Impacto sobre las inversiones actuales, como en personal, tecnología y capacitación.

Parte de los requerimientos empresariales son también especificaciones de ámbito para el Data Warehouse en los siguientes términos:

Áreas Tema

Estos son los temas de interés de las distintas funciones empresariales. Una cuidadosa y atinada selección de los Áreas Tema maximiza la utilidad del Data Warehouse

Granularidad

Se refiere al nivel de detalle de la información requerida. La Granularidad tiene una relación directa con las actividades de resumen y adición que deben realizarse sobre los datos fuente. A menor Granularidad mayor cantidad de detalle. Los datos operacionales son considerados como el nivel más bajo de granularidad. Para incrementar la granularidad los datos operacionales deben resumirse y acumularse más con el fin de incrementar su utilidad para la toma de decisiones. Por lo general, entre mayor sea la granularidad, mayor será el procesamiento para convertir y resumir los datos operacionales. Al mismo tiempo los datos con alta granularidad requieren menos volumen de almacenamiento y pueden ser más rápidamente consultados. El costo del procesamiento es continuo para cada actualización del Data Warehouse con nueva información.

La figura a continuación muestra la pirámide de datos que describe la relación entre:

- El incremento de granularidad
- Incremento de utilidad para quienes toman decisiones en altos niveles
- Disminución del volumen de datos
- Incremento del cómputo

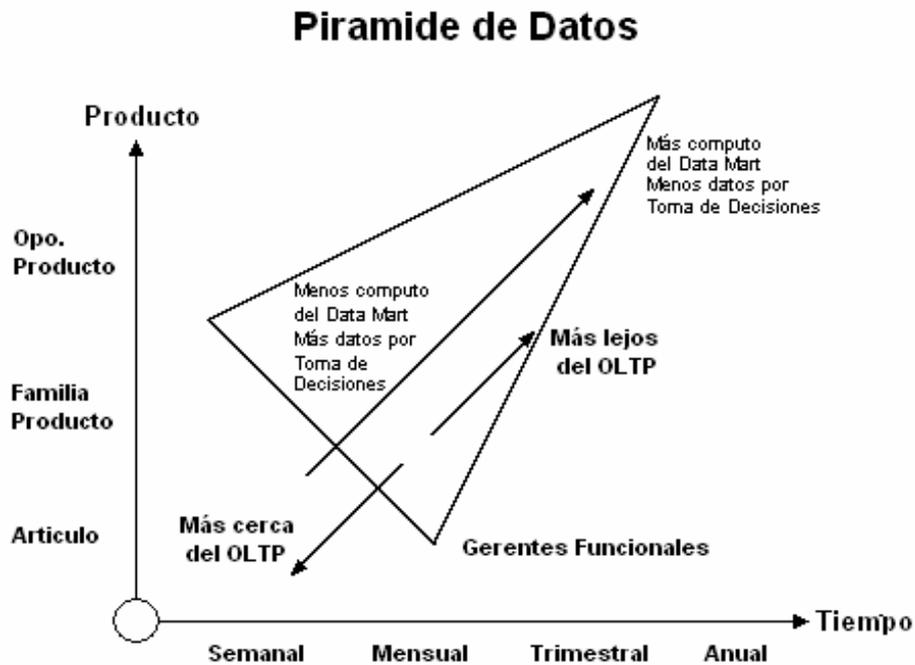


Figura 25 Pirámide de datos que muestra la relación entre la granularidad³²

Dimensiones

Un Data Warehouse organiza un gran conjunto de datos operacionales e históricos mediante múltiples dimensiones de categorización una de las dimensiones mas importantes es el tiempo. A los datos operacionales se les asigna un registro de tiempo en la fuente. Luego el Data Warehouse es capaz de agrupar todos los datos que ocurrieron dentro de un determinado rango de tiempo para responder a una determinada solicitud de consulta.

La figura siguiente muestra la ampliación de algunas de las dimensiones mencionadas anteriormente para indicar como se utilizan para la clasificación de datos dentro del Data Warehouse.

³² Fuente: Data Warehousing Cliente/Servidor; Harjinder S. Gill

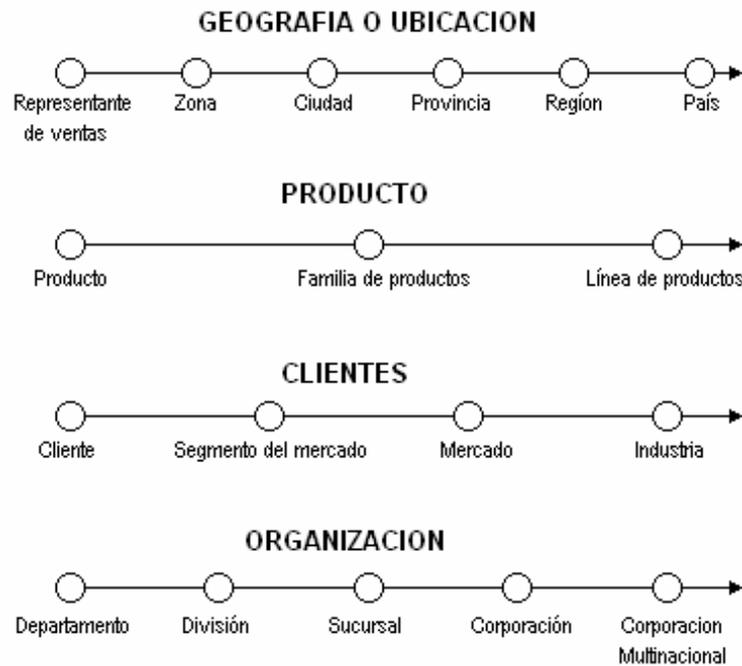


Figura 26 Intereses comunes para la mayoría de organizaciones³³

1.4.2.2 Definir los requerimientos arquitectónicos

La persona responsable de diseñar los distintos componentes del Data Warehouse el arquitecto ya que dispone de los conocimientos para construir un proyecto que sustente las necesidades actuales y futuras.

Los arquitectos determinarán los siguientes requerimientos arquitectónicos:

- El uso de estándares e interfaces
- El rango de plataformas necesarias para la implementación
- El rango de funciones y características a ofrecer por parte del Data Warehouse
- La flexibilidad para incorporar mejoras

Una de las funciones de los arquitectos es reunir una serie de requerimientos que concuerden con la visión del propietario, así como un conjunto de requerimientos que guíe para escoger la tecnología que se implementará. Uno de los enfoques estructurados más para implementar la arquitectura obteniendo un buen diseño es la metodología EAP (Enterprise Architecture Planning) que significa Planeación de Arquitectura Empresarial

³³ Fuente: Data Warehousing Cliente/Servidor; Harjinder S. Gill

Esta metodología nos describe tres tipos de arquitectura que describiremos a continuación:

1.4.2.2.1 Arquitectura de datos

En esta arquitectura describimos los diferentes datos y las relaciones entre ellos. La caracterización de datos se ve como una actividad fundamental. La razón es que las aplicaciones no pueden desarrollarse sin definir las características de los datos que deben ser creados o modificados. Generalmente las arquitecturas de datos son descritas como Modelos Entidad - Relación.

1.4.2.2.2 Arquitectura de aplicación

Si consideramos a un sistema como una combinación de diversas aplicaciones podemos decir entonces que juntas estas aplicaciones permiten la funcionalidad de un sistema. La arquitectura de aplicación se define después de la arquitectura de datos. La arquitectura de aplicación es un catálogo de aplicaciones junto con las funciones que ofrecen cada una y las interfaces entre cada aplicación. La arquitectura de aplicación también se confronta con la arquitectura de datos utilizando una matriz "CRUD" (Create, Read, Update o Delete), se establecen referencias cruzadas desde cada aplicación con uno o más datos que van a ser creados, leídos, actualizados o eliminados por otra aplicación.

1.4.2.2.3 Arquitectura de tecnología

Arquitectura de tecnología es una descripción de todos los componentes de tecnología utilizados en el proyecto Data Warehouse. Una arquitectura de tecnología se construye dividiendo un sistema en elementos de tecnología como una computadora servidor, una estación de trabajo, la interfaz gráfica de usuario, el RDBMS y el diccionario de datos, y luego seleccionando candidatos con base en criterios de evaluación, tales como los estándares y el costo.

Cada una de estas arquitecturas se desarrolla utilizando la arquitectura de referencia del Data Warehouse como un mecanismo de categorización. La arquitectura de datos se desarrolla mediante la identificación de los diversos elementos de datos y metadatos de los distintos bloques de la arquitectura de referencia. La arquitectura de aplicación se desarrolla mediante una lista las características y funciones de las diversas aplicaciones para los distintos bloques

de la arquitectura de referencia. También la arquitectura de tecnología se desarrolla por medio de una lista de las distintas opciones de tecnología que encontramos en las diferentes capas de la arquitectura de referencia.

1.4.2.3 Definir los requerimientos del desarrollador

El arquitecto ve al Data Warehouse en abstracto mientras que el desarrollador debe verlo ya en concreto. Los requerimientos del desarrollador están muy cercanos a la arquitectura de implementación. El desarrollador requiere que las arquitecturas de datos de aplicación y de tecnología formuladas por el arquitecto se subdividan en lo posible en computadoras, aplicaciones, interfaces, bases de datos, comunicaciones y pantallas de interfaz de usuario específicas. Por lo tanto, los requerimientos del desarrollador son un refinamiento de los requerimientos del arquitecto, con decisiones tomadas respecto a la selección de plataformas y la separación de la arquitectura de datos y la arquitectura de aplicaciones sobre las plataformas seleccionadas. También los requerimientos del desarrollador se relacionan con descripciones detalladas de la arquitectura de tecnología para la especificación de elementos tales como el lenguaje de programación, el acceso al RDBMS y los protocolos de comunicación. Además, los desarrolladores necesitan también de los siguientes requerimientos:

1.4.2.3.1 Requerimientos de tecnología

Los requerimientos de tecnología para los diversos bloques de la arquitectura de referencia son los siguientes:

- Para el bloque de **Fuentes de Datos**, se requieren los siguientes requerimientos:
 - Fuentes de datos y metadatos
 - Extracción de datos y metadatos
 - Almacenamiento de datos
 - Administración y manejo de datos
 - Redes y comunicaciones
 - Procesadores OLTP
 - Ambientes operativos
 - Administración del flujo de trabajo y estándares

- Para el bloque del Data Warehouse, necesitamos incorporar:
 - El refinamiento y reingeniería de datos
 - Redes y comunicaciones
 - Procesadores de Data Warehouse y ambiente operativo
 - Almacenamiento de datos y metadatos en el Data Warehouse
 - Catálogo de metadatos
 - Administración del flujo de trabajo y estándares
- Para el bloque del Data Mart, se necesita poner en escena:
 - Refinamiento y reingeniería de datos
 - Redes y comunicaciones
 - Procesadores de mercado de datos y ambiente operativo
 - Almacenamiento de datos y metadatos en el Data Mart
 - Catálogo de metadatos
 - Administración del flujo de trabajo y estándares
- Para el bloque de Acceso del Usuario Final y Herramientas, se requiere:
 - Middleware de acceso y recuperación
 - Almacenamiento local
 - Ambiente de datos multidimensional
 - Navegación de metadatos y herramientas de reporte
 - Estándares
 - Herramientas de usuario empresarial como análisis y reportes
 - Modelado empresarial
 - Minería de datos
 - Navegación de datos y procesamiento analítico en línea OLAP
 - Nuevas aplicaciones de producción.

1.4.2.3.2 Requerimientos de despliegue

Estos se relacionan con la capacidad del Data Warehouse para distribuir información de manera conveniente y oportuna y proporcionar acceso a ella. El Data Warehouse debe proporcionar un rango de métodos de acceso para herramientas de usuario final y aplicaciones especializadas del Data Warehouse. Debe además proporcionar un rango de caminos de conectividad desde la navegación mediante una PC portátil, hasta las estaciones de trabajo conectadas

a una red de área local. Por lo tanto, deben considerarse los siguientes temas de requerimientos de despliegue del Data Warehouse:

- Métodos de acceso
- Herramientas de acceso
- Métodos de recuperación
- Requerimientos de conectividad
- Requerimientos de plataforma cliente

1.4.2.3.3 Requerimientos de disposición para la producción del Data Warehouse

Además de los aspectos relacionados con el despliegue, varios requerimientos se relacionan con el mejoramiento en la producción de una solución de Data Warehouse. Un sistema de Data Warehouse se convierte en un sistema de producción cuando se vuelve de Misión Crítica y se utiliza con frecuencia para apoyar las decisiones operacionales y estratégicas. Los requerimientos de disposición para la producción se relacionan principalmente con el manejo de la solidez y la disponibilidad, la conservación de la consistencia y la precisión de la información, el manejo del desempeño al crecer el almacenamiento de datos, la definición de políticas y procedimientos para la actualización y el mantenimiento tanto del metamodelo del Data Warehouse como de los datos, proporcionando control de acceso y procedimientos de seguridad. Entre los requerimientos de disposición para la producción tenemos a los siguientes:

- Mantener la consistencia, confiabilidad y actualidad de la información
- Manejar los metadatos y el metamodelo del Data Warehouse
- Asegurar que los mecanismos de transporte, base de datos, computadoras y mecanismos de comunicación estén listos y disponibles todo el tiempo
- Proporcionar soporte técnico inmediato y capacidad de ayuda en línea para asistir a los usuarios cuando falle el sistema o cuando tengan preguntas operacionales
- Proporcionar políticas y procedimientos de seguridad de acceso y autenticación
- Administrar el tamaño de las bases de datos que utiliza el Data Warehouse, incluyendo la selección de Tecnología de Bases de Datos muy grandes, lo cual incluye la selección de múltiples niveles de almacenamiento en diversos

tipos de medios. Es común que en un proyecto de Data Warehouse los costos de almacenamiento sean de alrededor del 50 % del presupuesto de hardware.

- Mejorar el tiempo de respuesta de acceso. Se obtienen mejoras en el desempeño mediante una recuperación eficiente; así como por medio del manejo del tamaño de los datos que debe mantener el Data Warehouse. Se logran mejoras a través del uso de arreglos, índices múltiples, mecanismos de apagado o bloqueo selectivo, revisión de datos no utilizados y respaldo de datos de baja utilización.

1.4.2.3.4 Requerimientos para el desarrollo y despliegue del personal y sus habilidades

El desarrollo del Data Warehouse requiere de distintas habilidades para las diferentes fases del desarrollo. Algunas de estas fases son muy costosas y se requieren por periodos cortos. Una útil actividad de acopio de requerimientos consiste en identificar cuáles habilidades se requieren para qué fase del ciclo de vida del desarrollo del Data Warehouse. Otro método para identificar habilidades es utilizar la arquitectura de referencia del Data Warehouse para separar las habilidades en categorías, con base en los bloques y capas arquitectónicos.

La implementación del Data Warehouse es un esfuerzo de equipo. Después de identificar a los diferentes miembros del equipo, es también importante definir sus funciones en el proyecto. Esto se hace utilizando la arquitectura de referencia mediante la asignación de diversas funciones de miembros del equipo dentro de los bloques y capas arquitectónicos.

1.4.2.4 Definir los requerimientos de usuario final

El usuario final ve al Data Warehouse como un proyecto muy bien guardado cuyo acceso principal es a través de aplicaciones y herramientas de consulta y reportes, junto con cierto tipo de ubicación de la información contenida dentro del Data Warehouse. Los requerimientos del usuario final pueden ubicarse en una o más de las siguientes categorías:

1.4.2.4.1 Flujo de trabajo

¿Cómo se ajusta la funcionalidad que ofrece el Data Warehouse con el flujo de trabajo diario del usuario final?

En la figura a continuación presentamos un proceso de muestra para Investigación de mercados en ausencia de las capacidades de apoyo a las decisiones que ofrece el Data Warehouse.

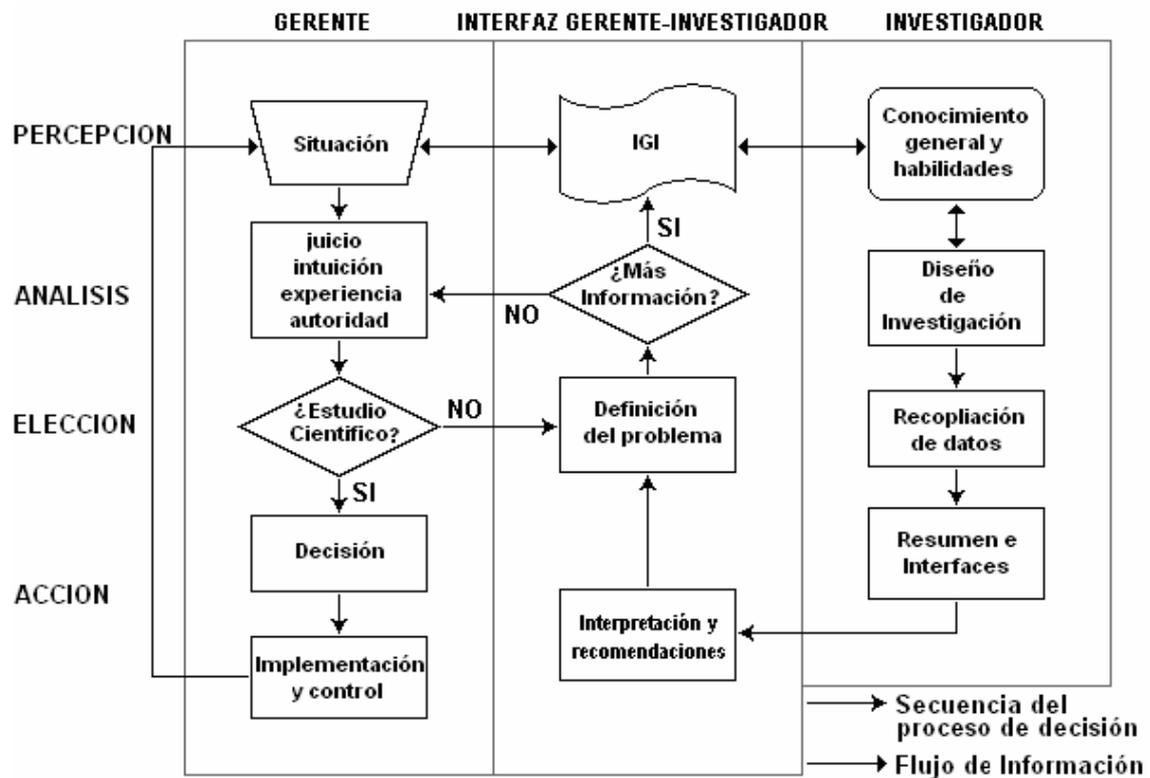


Figura 27 Flujo de trabajo de una Investigación de Mercado previo al Data Warehouse³⁴

A continuación en la figura vemos como las capacidades del Data Warehouse se incorporan en un flujo de trabajo modificado. En este escenario el usuario final es el especialista de investigación de mercado con acceso al Data Warehouse.

³⁴ Fuente: Data Warehousing Cliente/Servidor; Harjinder S. Gill

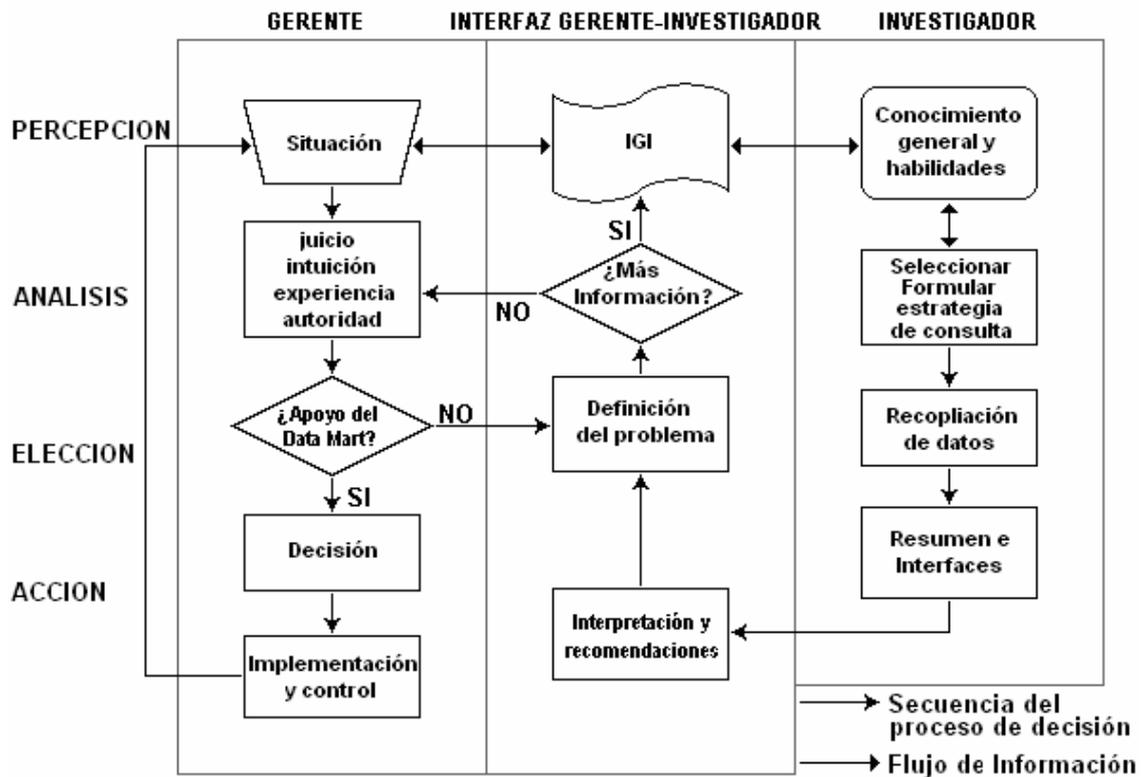


Figura 28 Flujo de trabajo de una Investigación de Mercado luego del Data Warehouse³⁵

1.4.2.4.2 Requerimientos de consulta

Como se mencionó en la fase de planificación el empleo de escenarios de uso empresarial son una valiosa herramienta para realizar prototipos de las capacidades del Data Warehouse y también para establecer las expectativas que tienen cada uno de los usuarios finales. Los requerimientos de consulta son simplemente la obtención de las consultas empresariales del usuario final en su propia terminología.

Además los usuarios finales pueden también especificar requerimientos de consultas de datos que les permita tener la información que cada uno de ellos necesite a tiempo y eficientemente. Los siguientes son algunos ejemplos de estos requerimientos:

- Satisfacer las necesidades de una la mayor variedad de usuarios posible

³⁵ Fuente: Data Warehousing Cliente/Servidor; Harjinder S. Gill

- Tener un acceso rápido, buena manipulación y excelente presentación de los datos
- Permitir a los usuarios la creación de sus propias consultas utilizando términos empresariales que ya conocen y ofrecer estructuras de datos consistentes
- Tener una capacitación y soporte mínimos para cada usuario de acuerdo a sus necesidades, tanto en el manejo del hardware como del software utilizados en el Data Warehouse

Los usuarios finales pueden además especificar los tipos de análisis que se efectuaran sobre los datos recuperados del Data Warehouse. Como ejemplos de requerimientos de análisis de datos tenemos a los siguientes:

- Tipos de actividades:
 - Separar elementos de datos en varias formas
 - Exponer más detalle de manera progresiva
 - Buscar patrones ocultos de datos (Data Mining)
 - Examinar de una manera indirecta (Navegación)
 - Copiar y realizar modificaciones locales
 - Construir modelos empresariales
- Visualización de datos:
 - En dos dimensiones: Hojas de cálculo y relacional
 - En múltiples dimensiones
 - Reportes y cuadros
 - Base de datos de muestra activa

1.4.2.4.3 Requerimientos de reportes

Los usuarios finales del Data Warehouse pueden obtener diversos requerimientos de reportes.

Los reportes generados por los requerimientos de reportes permitirán mejorar la producción de la empresa.

1.4.3 ANÁLISIS

La fase de análisis del ciclo de desarrollo del Data Warehouse significa convertir los requerimientos recolectados en la fase de requerimientos, en un conjunto de especificaciones que puedan apoyar al diseño.

El proceso de análisis consiste en derivar modelos físicos y lógicos de datos para el Data Warehouse y los Data Mart y definir los procesos necesarios para conectar las fuentes de datos, el Data Warehouse, los mercados de datos y las herramientas de acceso del usuario final

Para realizar la etapa de análisis debemos seguir varias etapas que describiremos a continuación:

1.4.3.1 Analizar los factores que conducen el negocio (Business Drivers)

En esta etapa se analizan los distintos factores que afectan a la empresa y obligan a tomar decisiones, como ejemplo de dichos factores tenemos a los siguientes:

- Ingreso de nuevos competidores
- Ingreso de nuevos clientes
- Cambios en las regulaciones legales
- Desarrollo de nuevos productos y servicios
- Incrementar la producción y competitividad

Todos estos factores producen que dentro de la empresa se deban tomar decisiones para aumentar principalmente la producción y la competitividad y optimizar la tecnología disponible mejorando los distintos procesos que se realizan dentro de la empresa que implementara en Data Warehouse.

1.4.3.2 Analizar los objetivos del negocio

Ahora analizaremos las metas o los objetivos que tiene la empresa con la finalidad de que la solución de Data Warehouse cumpla con estos objetivos, para esto debemos tener las metas definidas con claridad. Algunos ejemplos de las metas que tienen las empresas son los siguientes:

- Incrementar ventas y ganancias
- Aumentar el número de clientes y expandir el mercado
- Minimizar costos
- Incrementar la satisfacción del cliente
- Mejorar la productividad

1.4.3.3 Analizar las necesidades de información de alto nivel

Se analizan la información recopilada a través de datos obtenidos con los distintos usuarios que tendrá el Data Warehouse; para esto se necesita buscar expertos o encargados en cada una de las áreas a analizar que colaboraran en este proceso; además recopilar estudios de indicadores gerenciales.

1.4.3.4 Recolectar información de procesos

En esta parte se analizara cada proceso y la información requerida por el mismo; además de los responsables de dicho proceso; con la finalidad de tener un claro panorama de cada uno de los procesos de la empresa.

1.4.3.5 Identificar áreas críticas

Ahora analizaremos cada área de la organización en la que la solución de Data Warehouse puede mejorar los procesos y manejar de mejor forma la información.

1.4.3.6 Identificar roles de personas

Ahora realizaremos una breve identificación de los roles que cumplen cada una de las personas que usaran la solución Data Warehouse para identificar los diferentes perfiles de usuario que tendremos dentro de nuestro proyecto.

1.4.3.7 Revisar infraestructura informática

También debemos realizar una revisión de la infraestructura informática de la empresa sobre la que se implantará la solución del Data Warehouse para confirmar que se posee de todas las facilidades tanto en hardware como en software para implementar correctamente la solución Data Warehouse.

1.4.3.8 Revisar Planes Tácticos y Estratégicos

Dentro de los diferentes documentos que debemos revisar de una empresa tenemos a los planes tácticos y estratégicos que posee la empresa, ya que la información que obtengamos de estos planes ayudaran a un mejor diseño del Data Warehouse.

1.4.3.9 Identificar limitaciones y restricciones

También debemos identificar las diferentes limitaciones y restricciones que se nos presentan en la empresa donde se implementara la solución Data Warehouse tanto físicas, operacionales, tecnológicas, entre otras.

1.4.3.10 Recolectar y revisar documentación de la empresa

Dentro de la documentación que la empresa posee podemos encontrar importante información que a veces no tomamos en cuenta y que nos servirá para el mejor desarrollo del Data Warehouse.

Dentro de la documentación que deberíamos recolectar y revisar tenemos los siguientes documentos:

- Orgánico Funcional de la empresa
- Manuales de Procesos o Documentos de procesos
- Planes informáticos, planes de contingencia, etc.
- Manuales y documentación de sistemas que están en funcionamiento y que proveerán al Data Warehouse de datos
- Oficios recibidos por el área de sistemas o que originaron en dicha área

Toda la información que obtengamos a partir de estos distintos documentos será muy útil para convertirlos en especificaciones que mejoren el análisis para la construcción de la solución Data Warehouse.

1.4.3.11 Identificar procesos de alto nivel e indicadores claves de rendimiento

Los procesos de alto nivel serán identificados por los gerentes o por expertos en cada área en reuniones internas que tendrán durante el desarrollo de la solución de Data Warehouse.

Luego se hayan identificado los procesos de alto nivel se definirán las dimensiones de análisis.

1.4.3.12 Analizar las fuentes de datos

Especificaciones de requerimientos de fuentes de datos que delinear las fronteras de información disponible en las fuentes de datos actuales.

1.4.4 DISEÑO

Luego de tener correctamente desarrollado el análisis, viene la parte de diseño, la misma que es básica para luego continuar con la construcción del Data Warehouse.

Si se desarrolla un correcto diseño de del Data Warehouse, se tendrá velocidad en las consultas, con lo cual la demora el realizar una búsqueda puede ser mínima o llegar a tener tiempos muy altos.

1.4.4.1 Diseño detallado de la arquitectura de datos

Dentro del diseño de la arquitectura de datos, tenemos desarrollo de modelos físicos de datos para:

- Almacenamiento de Data Warehouse.
- Almacenamiento Local.
- Almacenamiento de Data Marts.

1.4.4.2 Diseño detallado de la arquitectura de la aplicación

Se deben tomar en cuenta los datos propios de la empresa, así como también los procesos que va a desarrollarse durante el ciclo de vida del data warehouse tanto para:

- Procesos con las fuentes de datos propias del Data Warehouse
- Manejo interno del Data Warehouse con respecto a los procesos.
- Procesos externos al Data Warehouse que se conectan al mismo para buscar cierta información.

Como se puede notar claramente, la información del Data Warehouse puede ser también acezada por otros procesos externos al mismo, lo cual implica que el diseño del Data Warehouse debe contemplar estas características de adaptabilidad.

1.4.5 IMPLEMENTACIÓN

1.4.5.1 Construcción

En la fase de construcción se implementa físicamente los diseños que se desarrollaron durante la fase de diseño. Para la construcción de una solución de Data Warehouse debemos decidir primero si es mejor construir o comprar esta

solución, ya teniendo una clara respuesta a este cuestionamiento consideramos en integrar a la empresa una solución de Data Warehouse.

La construcción del Data Warehouse requiere de muchas aplicaciones integradas que se construirán y este proyecto se puede comparar con el desarrollo de un sistema grande de base de datos relacional.

Algunas de las aplicaciones que deberemos desarrollar en un proyecto Data Warehouse son los siguientes:

- Aplicaciones que se encarguen de la extracción de datos de las distintas fuentes que posee un sistema de Data Warehouse tales como sistemas externos, hojas de cálculo, reportes, entre otros.
- Aplicaciones que creen, actualicen o modifiquen las bases de datos para del Data Warehouse y los mercados de datos.
- Aplicaciones que realicen transformaciones de datos, tales como integración, resumen y adición.
- Programas que efectúen búsquedas en bases de datos muy grandes. Para esto tenemos herramientas de usuario final han optimizado utilerías de búsqueda para acelerar las consultas que generan y también capacidades de recuperación en paralelo.

Además tenemos distintos fabricantes que ofrecen en el mercado este tipo de aplicaciones.

Dentro de la fase de construcción se generan varios retos con respecto a dar una verdadera solución a través del Data Warehouse a las necesidades de la empresa y entre los principales retos podemos mencionar los siguientes:

- La toma de decisiones inteligentes entre comprar o construir un Data Warehouse
- El entender cómo invertir de mejor manera las inversiones existentes en plataformas, tecnología y mejoramiento de las habilidades
- La selección y evaluación adecuada de componentes que se utilizarán en la construcción del Data Warehouse
- La capacidad de integración de sistemas para juntar las aplicaciones con las fuentes de datos existentes y las herramientas de acceso de datos

- La administración de metadatos

1.4.5.2 Implantación

La fase de implantación nos conlleva todo el proceso de la instalación de las diferentes aplicaciones sobre las que funcionara el Data Warehouse, la puesta en servicio de dichas aplicaciones y el funcionamiento de la solución de Data Warehouse.

Dentro de las diferentes actividades que se requieren en esta fase tenemos a las siguientes:

- Asegurar la integración dentro de la infraestructura existente
- Proporcionar la instalación inicial, incluyendo las conexiones básicas de datos con las fuentes y la actualización y sincronización de datos
- Proporcionar capacitación y orientación a los usuarios
- Proporcionar la administración de usuarios y sistemas
- Proporcionar la capacidad de generar respaldos
- Asegurar la completa disponibilidad de los procesos para manejar caídas de los sistemas y sus componentes de infraestructura
- Planear e implementar la actualización de plataformas y dar el mantenimiento necesario al Data Warehouse cuando se requiera

Además de los requerimientos normales de implantación para cualquier sistema de información empresarial, el sistema de Data Warehouse tiene un requerimiento adicional muy importante que es que la mayoría de los usuarios finales del Data Warehouse no son técnicos, por lo que necesitan ver definiciones de la información contenida en el Data Warehouse en términos simples y lenguaje que comprendan con facilidad.

La implantación de la solución Data Warehouse requiere capacidades adicionales para su correcto funcionamiento y uso de sus diferentes usuarios, como por ejemplo los catálogos de información, los browsers de información, entre otros. Algunas de estas capacidades ya están disponibles en los sistemas de gestión de base de datos que encontramos en el mercado.

1.4.6 PRUEBAS

Para evaluar el funcionamiento la solución Data Mart debemos realizar una pruebas en distintos aspectos, con el rendimiento que obtengamos de cada una de las pruebas podemos observar las fortalezas y debilidades que tengamos dentro del desarrollo para mejorar en los aspectos débiles. Dentro de las pruebas que realizaremos tenemos a las siguientes:

1.4.6.1 Pruebas de Unidad

El proceso de pruebas empieza con las pruebas de unidad, pero hay que entender que es una prueba de unidad y adicionalmente definir qué es una unidad.

Una unidad se define como una pieza de análisis a muy bajo nivel. Una unidad puede estar expresado por un path de direccionamiento, un dato de una base de datos, un reporte o un procedimiento almacenado. Es decir que en las pruebas de unidad examinaremos el correcto funcionamiento de cada una de las unidades que conforman una solución Data Warehouse.

1.4.6.2 Pruebas de Integración

Las pruebas de integración son aquellas en las que examinamos que la solución Data Warehouse se pueda integrar a otras soluciones o aplicaciones a través del uso de un diseño documentado y de estándares de programación.

1.4.6.3 Pruebas de Validación

Las pruebas de validación se enfocan a saber si la solución Data Warehouse que se ha construido es la misma que el cliente desea, es decir si se ha cumplido con todos los requerimientos que se acordó con el cliente.

Para estas pruebas es necesario contar con la documentación de la fase de requerimientos; de tal forma que se pueda cumplir con todas las necesidades del cliente.

1.4.6.4 Pruebas del Sistema

Las pruebas del sistema se encargan de saber si todos los componentes que permiten el funcionamiento de la solución Data Warehouse funcionan

correctamente. Y también examina que las interfaces que usaran los usuarios para visualizar la información sean apropiadas.

CAPITULO 2

2 DESARROLLO DE UN DATA MART PARA EL ÁREA DE SISMOLOGÍA DEL DEPARTAMENTO DE GEOFÍSICA DE LA ESCUELA POLITÉCNICA NACIONAL

2.1 PLANIFICACIÓN

Para la planificación inicial de la solución Data Mart del Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional se realizará una secuencia de actividades que a continuación describimos y que son primordiales para un óptimo desarrollo de dicha solución.

2.1.1 SELECCIONAR LA ESTRATEGIA DE IMPLEMENTACIÓN

Para desarrollar la solución Data Mart para el Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional se vio conveniente usar como **estrategia de implementación el enfoque conocido como de abajo hacia arriba**, ya que vamos a partir de lo particular para llegar a lo general; a continuación enumeramos los motivos por los cuales se selecciono esta implementación:

- Se comenzará con prototipos para en base a ellos ir mejorándolos y obtener un resultado óptimo.
- Este tipo de implementación es mucho más rápida ya que en la toma de decisiones participa menos gente y se puede llegar más rápidamente a consensos.
- El gasto con este tipo de implementación es mucho menor en relación a los beneficios.

2.1.2 SELECCIONAR LA METODOLOGÍA DE DESARROLLO

La metodología que se usará para desarrollar la solución de Data Mart para el Área de Sismología de la Escuela Politécnica Nacional es la **metodología de Desarrollo Espiral** ya que de acuerdo a la disponibilidad de la organización es la más adecuada debido a los siguientes motivos:

- Con esta metodología es más rápido construir una solución Data Mart con nuevos requerimientos.

- Esta metodología nos permitirá construir prototipos funcionales que se irán mejorando en las versiones sucesivas en caso de necesitar funcionalidades adicionales.

2.1.3 SELECCIÓN DEL ÁMBITO DE IMPLEMENTACIÓN

Para seleccionar el ámbito inicial de implementación para la solución Data Mart del Instituto Geofísico de la Escuela Politécnica Nacional debemos considerar varios aspectos tanto desde un punto de vista empresarial como desde el punto de vista tecnológico:

En cuanto al punto de vista empresarial vemos que el departamento que usará la solución Data Mart es el Área de Sismología del Instituto Geofísico con la finalidad de optimizar el manejo de la información de esta área, ya que dicha información es muy importante para el monitoreo y vigilancia de la actividad sísmica con el fin de disminuir el impacto de los sismos en nuestro país.

Desde la perspectiva tecnológica podemos decir que el volumen de los datos que se manejan en el Instituto Geofísico en el Data Mart son:

TIPO DE INFORMACIÓN	SISTEMA QUE LOS GENERA	SUBSISTEMA / FORMA DE ADQ	CANTIDAD DIARIA	CANTIDAD MES (30 días)
Sísmica	Earthworm	Rsam	0.2 MB	6 MB
Sísmica	Earthworm	Rsam	0.02 MB	0.6 MB
Sísmica	Earthworm	Rsam	0.03 MB	0.9 MB
Sísmica	Earthworm	Rsam	0.02 MB	0.6 MB
Sísmica	Earthworm	Ssam	44 MB	1320 MB
Sísmica	Earthworm	Ssam	6.4 MB	192 MB
Sísmica	Earthworm	Continuous	94.4 MB	2832 MB
Sísmica	Earthworm	Triggers	81.2 MB	2436 MB
Sísmica	Earthworm	Helicorders	1.81 MB	54.17 MB
Sísmica	ACQ	Triggers	80 MB	2400 MB

Sísmica	ACQ	Eventos	20 MB	600 MB
Sísmica	Marslite	Continuous	191.6 MB	5748 MB
Sísmica	Marslite	Continuous	52.33 MB	1570 MB
Sísmica	IRIS		0.08 MB	2.40 MB
Térmica	Thermalcam		284.49 MB	8534.58 MB
Laharica			10 MB	300 MB
Laharica			0.1 MB	3 MB
Grafica			100 MB	3000 MB
Bases volcánicas			0.28 MB	8.33 MB
Documental	Office		2.33 MB	70 MB
Documental	Informes		0.1 MB	3 MB
Mail	Servidor de Correo		45 MB	1350 MB

Tabla 2.1 Volúmenes de Información del Instituto Geofísico³⁶

La información documental y de usuarios:

GENERADOR DE LA INFORMACION	CANTIDAD
USUARIOS	500 GB Aproximadamente
USUARIOS (mail)	20 GB Aproximadamente
PUBLICACIONES	10 MB
BASES GEOGRAFICAS	30 GB

Tabla 2.2 Volúmenes de Información Documental

Los volúmenes que se añadirán a esta lista el próximo año por la incorporación de nuevos sistemas son:

³⁶ Fuente: IGEPN

TIPO DE INFORMACION	SISTEMA	SUBSISTEMA / FORMA DE ADQ	DIA	MES (30 D)
Sísmica	JICA	Continuous	420 MB	12600 MB
Video		Continuous	70.31 MB	2109.38 MB

Tabla 2.3 Volúmenes de Información que se añadirá próximamente³⁷

Se estima una tasa de crecimiento aproximada de 15% anual, sin contar con los saltos descritos que son producidos por nuevos sistemas o crisis.

Debido a la gran cantidad de información sísmica que debemos almacenar es recomendable el uso de una solución Data Mart en las distintas fuentes de datos que actualmente almacenan la información.

2.1.4 SELECCIÓN DE LA ARQUITECTURA DE IMPLEMENTACIÓN

La arquitectura de implementación que se selecciono para la solución Data Mart que se implementará en el Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional es la arquitectura Solo Data Mart y se decidió por esta opción de arquitectura debido al siguiente motivo:

- Esta arquitectura reconoce las necesidades específicas de cada departamento de una organización para satisfacerlas completamente, mientras que un Data Mart empresarial cumpliría con satisfacer las necesidades generales de la organización sin adentrarse en cada uno de los departamentos.

2.1.5 CRONOGRAMA Y PRESUPUESTO DEL PROYECTO

Dentro del desarrollo del Data Mart de Sismología tenemos una serie de actividades que se cumplirán y para esto se desarrollo un cronograma que nos guiará en el transcurso del desarrollo: **VER ANEXO B**

2.1.6 RECOPIACIÓN DE METADATOS

Dentro de la planeación del Data Mart de Sismología tenemos las definiciones de datos que se pueden observar en el **ANEXO D**.

³⁷ Fuente:IGEPN

2.2 ETAPA DE DEFINICIÓN DE REQUERIMIENTOS

En esta fase de definición de requerimientos se definen las características y funciones que tendrá el Data Mart que se implantara en el Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional partiendo de descripciones claras del uso que se le dará a la solución Data Mart.

2.2.1 DEFINIR LOS REQUERIMIENTOS GENERALES O PROPIOS

En el Área de sismología del Instituto Geofísico de la Escuela Politécnica Nacional se registra una gran cantidad de información de los diferentes sismos ocurridos en el Ecuador, por tal motivo es necesario construir un Data Mart para la organización, optimización y seguridad de los datos. Los resultados que el Data Mart desplegara ayudaran a obtener un mejor conocimiento de lo que ocurre en el Área de sismología.

Los principales objetivos que la empresa cumplirá serán reducir el impacto de desastres sísmicos, además de difundir los resultados de la vigilancia e investigación de las actividades sísmicas, identificando sus amenazas. Al momento de construir el Data Mart en el Área de Sismología se mejorará la organización de los datos, disminuyendo el tiempo de búsqueda y además facilitando su visualización. La información obtenida actualmente de los sismos es guardada en archivos planos dificultando su manejo; la ventaja de implementar el Data Mart radica en la mejora del manejo los datos y la reducción de búsqueda de los mismos.

Debido a la gran información de datos que se obtiene diariamente es necesario que el diseño del Data Mart sea el correcto caso contrario su administración se complicará, al momento en que existan tareas críticas por cumplir.

2.2.2 DEFINIR LOS REQUERIMIENTOS ARQUITECTÓNICOS

En esta sección recopilaremos los requerimientos que tienen que ver con los datos, con la aplicación y con la tecnología que nos permitirá implementar la solución Data Mart en el Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional.

El Departamento de Geofísica maneja un gran volumen de información de la actividad sísmica de nuestro país y actualmente dicha información se almacena

en archivos de texto, hojas de cálculos, en distintos formatos de GIS y en archivos de distintos formatos de gráficos.

La principal función de la solución Data Mart para el Área de Sismología es la de integrar toda esta información con el objetivo de tener un acceso rápido a la misma, además de incrementar la eficiencia y productividad en el proceso de toma de decisiones y prevención de desastres.

Las plataformas que se requerirán para la implementación del Data Mart las listaremos a continuación:

- Un Servidor de Base de Datos
- Clientes para el Data Mart
- Sistema Operativo Microsoft Windows XP
- Microsoft SQL Server 2000 o superior

Ya en cuanto a la implementación de la solución Data Mart se usarán estándares creados por nosotros en acuerdo con los técnicos del Instituto Geofísico ya que en la actualidad no tienen definidos estándares en dicho tema.

Además, los sistemas que proveerán de información a ser almacenada en el Data Mart son sistemas que proveen de información sísmica obtenida de las diferentes estaciones de la Red Nacional de Sismógrafos ubicadas en todo el país.

2.2.3 DEFINIR LOS REQUERIMIENTOS DEL DESARROLLADOR

Para el desarrollo del Data Mart del Área de Sismología se necesita los siguientes requerimientos mínimos para su correcto funcionamiento:

- Hardware
 - Procesador de 166MHz o superior
 - Memoria 128Mb
 - Espacio en disco Duro 20 GB
 - CD Rom
- Software
 - Sistema Operativo
 - Windows XP Professional
 - Internet Explorer 5.0 o superior
 - Java 2 SDK
 - Apache Tomcat

- Internet Information Services
- Microsoft SQL Server 2000

Para desarrollar el Data Mart se necesita tener la base de datos desde la cual se empezará con el desarrollo de la solución Data Mart.

2.2.4 DEFINIR LOS REQUERIMIENTOS DE USUARIO FINAL

Los usuarios finales necesitan visualizar los datos que se almacenarán en el Data Mart, para lo cual se necesita una herramienta desde la cual puedan acceder a estos datos.

Otro punto importante relacionado con los usuarios finales será el poder acceder a reportes de sismos por fechas, así como también reportes de sismos por regiones, y además por intensidades sísmicas.

2.3 ETAPA DE ANÁLISIS

Luego de obtenidos los requerimientos durante la fase de definición de requerimientos se deben analizar y convertirlos en especificaciones que nos ayudaran con el diseño del Data Mart para el área de sismología del Instituto Geofísico de la Escuela Politécnica Nacional.

2.3.1 ANALIZAR DE LOS FACTORES QUE CONDUCEN EL NEGOCIO

Entre los factores externos que afectan a la toma de decisiones dentro del Instituto Geofísico tenemos a los siguientes:

- El Rectorado de la EPN el 28 de noviembre del 2001, informa que Consejo Politécnico en sesión extraordinaria del 22 de noviembre de 2001, acredita en forma definitiva al Instituto Geofísico como un Departamento de la Escuela Politécnica Nacional y el Gobierno mediante Decreto No. 3593 del 13 de enero de 2003, decreta encargar al Instituto Geofísico de la Escuela Politécnica Nacional, el diagnóstico, la vigilancia y la mitigación de los peligros sísmicos y volcánicos en todo el territorio nacional.

2.3.2 ANALIZAR DE LOS OBJETIVOS DEL NEGOCIO

Dentro de los objetivos más importantes que el Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional se propone tenemos:

- Realizar la investigación científica fundamental para crear las bases necesarias para una efectiva reducción del riesgo.
- Comprender el volcanismo ecuatoriano para reducir el impacto de las erupciones.
- Entender los procesos de generación de sismos de origen tectónico y de la actividad sísmica relacionada con volcanes, para contribuir a la disminución del riesgo debido a estos fenómenos.
- Desarrollar y disponer de las herramientas tecnológicas y de capacidad técnica propia para generar y mantener el flujo de información necesaria para el monitoreo e interpretación sísmica y volcánica.
- Fortalecer la capacidad de gestión interna y de consecución de fondos externos

2.3.3 ANALIZAR LAS NECESIDADES DE INFORMACIÓN DE ALTO NIVEL

Luego de analizar las diferentes necesidades del Área de Sismología encontramos que la información de alto nivel necesaria para la toma de decisiones es la siguiente:

- Analizar la información de los sismos por cada nivel de intensidad
- Analizar la información de la concentración de sismos por cada región del país

2.3.4 DEFINIR PROCESOS Y ROLES

2.3.4.1 Procesos del Instituto Geofísico para el Área de Sismología

Dentro del Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional existen los siguientes procesos en los cuales se basará la Implementación del Data Mart de Sismología:

2.3.4.1.1 Monitoreo de los eventos Sísmicos

Actividades:

Este proceso comprende las siguientes actividades:

- Adquisición y procesamiento de la información sísmica
- Administración de Observatorios Vulcanológicos
- Administración de la RENSIG

Información de Entrada:

- Registro de datos de Sismología adquiridos automáticamente
- Registro de datos de Sismología adquiridos manualmente
- Diseño de redes de monitoreo sismología

Información de Salida:

- Registro de datos de Sismología adquiridos automáticamente
- Registro de datos de Sismología adquiridos manualmente
- Informes periódicos de monitoreo Sismología

2.3.4.1.2 Análisis y estudios de Eventos Sísmicos

Actividades:

Este proceso comprende las siguientes actividades:

- Estudio de eventos premonitores
- Elaboración de tomografías
- Análisis de Sismos LP.
- Análisis estadísticos de sismicidad
- Estudios de deformación costera

Información de Entrada:

- Registro de datos de Sismología adquiridos automáticamente
- Registro de datos de Sismología adquiridos manualmente
- Documentación de las áreas del IG

Información de Salida:

- Registros históricos
- Registro de análisis y estudios de Sismología

2.3.4.1.3 Evaluación de peligros Sísmicos - Volcánicos

Actividades:

Este proceso comprende las siguientes actividades:

- Evaluación de la Amenaza Volcánica para Proyectos de Infraestructura
- Elaboración y evaluación de modelos de tránsito de lahares.
- Modelación Teórica del Cotopaxi.
- Evaluación de peligros de volcanes.

Información de Entrada:

- Registro de datos de Sismología adquiridos automáticamente

- Registro de datos de Sismología adquiridos manualmente
- Registro de datos de Vulcanología adquiridos automáticamente
- Registro de datos de Vulcanología adquiridos manualmente
- Registro de análisis y estudios de Sismología
- Registro de análisis y estudios Vulcanología

Información de Salida:

- Informes de evaluaciones de peligros Sísmico-Volcánicos
- Registro de requerimientos de monitoreo

2.3.4.1.4 Difusión de información

Actividades:

Este proceso comprende las siguientes actividades:

- Elaboración y actualización de mapas de peligro de cada volcán.
- Elaboración de publicaciones.
- Elaboración de informes periódicos OVT.

Información de Entrada:

- Registro de análisis y estudios Sismología
- Registro de análisis y estudios Vulcanología

Información de Salida:

- Publicaciones
- Seminarios y Charlas

2.3.4.2 Roles dentro del Instituto Geofísico para el Área de Sismología

Dentro del Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional se han definido los siguientes roles de usuarios que tendrán acceso al Data Mart de Sismología:

Tipo de usuario	Nivel de uso del Data Mart
Comité técnico – Jefe del Departamento	Toma de decisiones
Científico de turno	Observación de información histórica y estadística Elaboración de informes sísmicos

Científicos y personal que maneja los datos de la RENSIG ³⁸	Nuevas consultas Elaboración de estadísticas Elaboración de informes
Personal de monitoreo	Carga de datos Elaboración de informes
Administrador de la base de datos	Actualización de la base de datos (definición, propiedades) Uso intensivo de la documentación del sistema

Tabla 2.4 Roles

2.3.5 DEFINIR LOS INDICADORES CLAVES DE RENDIMIENTO

Dentro del análisis realizado obtuvimos que los siguientes indicadores claves de rendimiento:

- Cantidad de Sismos por Región
- Cantidad de Sismos por Provincia
- Cantidad de Sismos por Cantón
- Cantidad de Sismos por Magnitud
- Cantidad de Sismos por Región que se han producido mensualmente.
- Cantidad de Sismos por Región que se han producido trimestralmente.
- Cantidad de Sismos por Región que se han producido anualmente.
- Regiones con mayor porcentaje de sismos de alta intensidad sísmica
- Regiones con mayor porcentaje de sismos de baja intensidad sísmica

2.3.6 DIMENSIONES Y HECHOS

Luego de los diferentes análisis obtuvimos las siguientes dimensiones y hechos para la construcción de la solución Data Mart:

DIMENSIONES

- Tiempo
- Estación
- TipoSismo
- Región

³⁸ La definición se encuentra en <http://www.igeqn.edu.ec/instrumentacion/monitoreo.htm>

HECHOS

- Sismos

2.3.7 ANALIZAR LAS FUENTES DE DATOS

Entre las fuentes de datos de la solución Data Mart del Instituto Geofísico tenemos las siguientes:

2.3.7.1 Base de datos de sismos localizados e históricos

Esta fuente de datos se encuentra actualmente implementada en Microsoft Access y contiene la información de sismos que generan las distintas estaciones de la Red Nacional de Sismógrafos, en esta base de datos esta almacenada información de sismos desde hace 15 años aproximadamente.

2.3.7.2 Mecanismos focales

Esta fuente de datos se creara a partir de información que se generan de los distintos programas que la obtienen de la Red Nacional de Sismógrafos y contiene tanto archivos de Microsoft Excel y Archivos Gráficos de Mapas de Bits desde hace 10 años.

2.4 DISEÑO

Luego de realizar el procesamiento y análisis de toda la información obtenida procedemos a realizar el diseño que nos permitirá implementar tanto la arquitectura de los datos de la solución Data Mart como la arquitectura de la Aplicación.

2.4.1 DISEÑO DETALLADO DE LA ARQUITECTURA DE DATOS

A continuación tenemos el diseño de los modelos físicos de datos que conformaran el Data Mart de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional, una descripción de cada una de las entidades y atributos la tenemos en el ANEXO D:

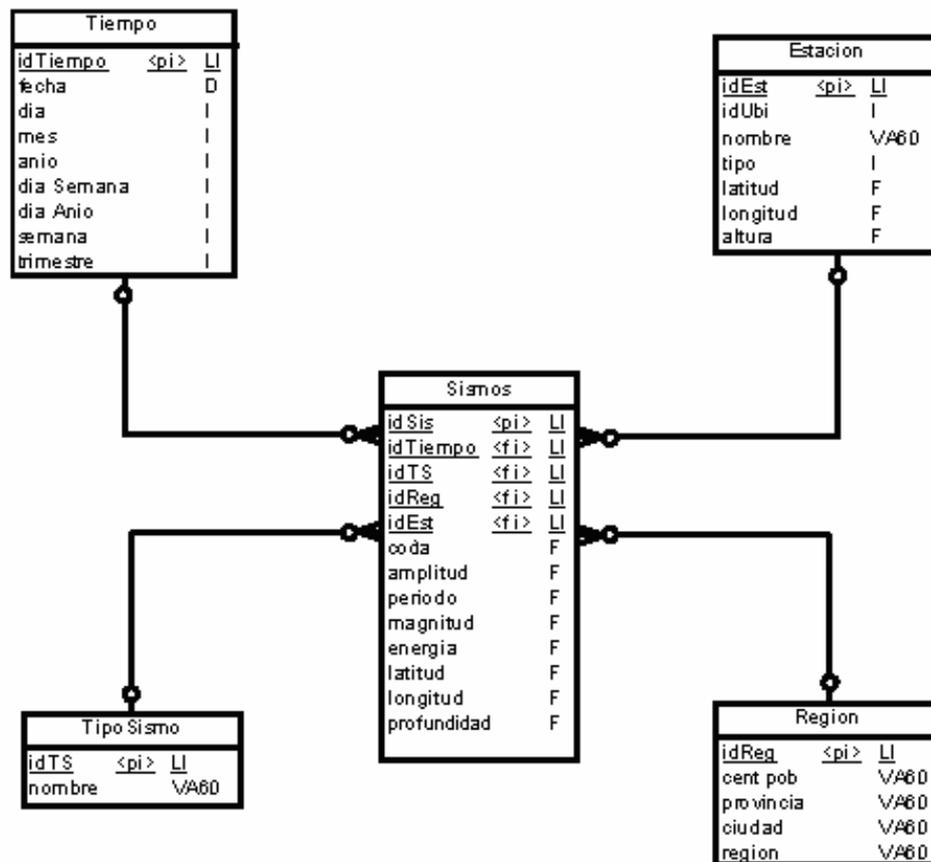


Figura 29 Diseño de la Arquitectura de Datos³⁹

2.4.2 DISEÑO DETALLADO DE LA ARQUITECTURA DE LA APLICACIÓN

A continuación tenemos el diseño de las distintas aplicaciones que forman parte del Data Mart de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional; una descripción de cada una de las aplicaciones la tenemos en el ANEXO E:

³⁹ Fuente: Autores

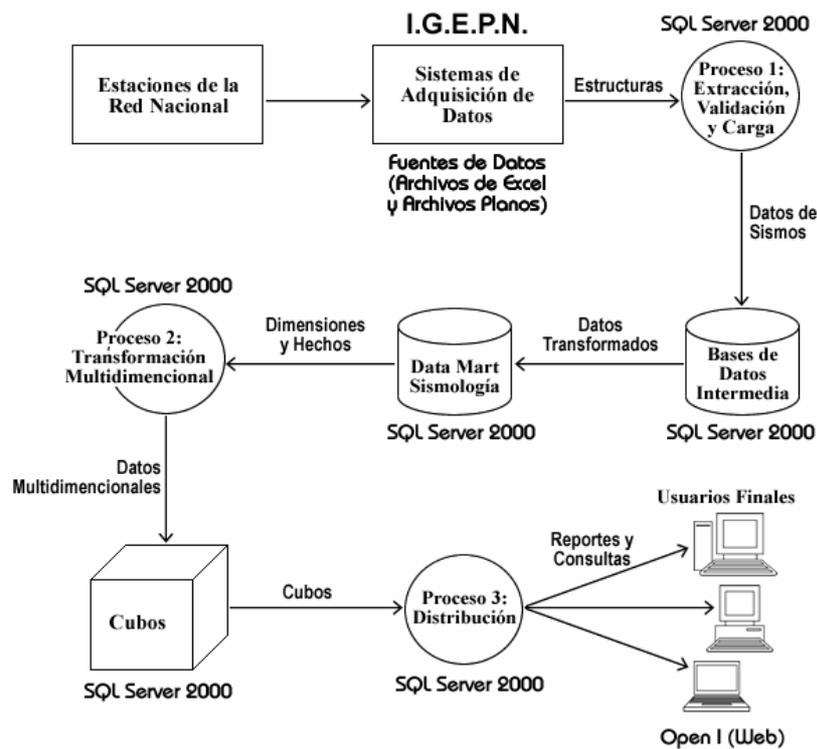


Figura 30 Diseño de la Arquitectura de la Aplicación⁴⁰

2.5 IMPLEMENTACIÓN

Luego de haber recolectado la información necesaria y concluido con los distintos diseños procedemos a implementar la solución Data Mart para el Área de Sismología del Instituto Geofísico utilizando las diferentes herramientas disponibles tanto en hardware como software.

2.5.1 CONSTRUCCIÓN

2.5.1.1 Selección de Plataformas de Software

Para la Solución Data Mart del Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional debemos seleccionar los distintos programas que se utilizarán para su construcción; esto se hará de acuerdo a la disponibilidad física y a un óptimo funcionamiento sobre las plataformas existentes. A continuación

⁴⁰ Fuente: Autores

detallaremos los diferentes programas que se han seleccionado para la construcción del Data Mart de Sismología del Instituto Geofísico:

2.5.1.1.1 Sistema Operativo

Se selecciono Microsoft Windows XP Professional Versión 5.1 con Service Pack 2 debido a que brinda el soporte necesario en cuanto a seguridad ya que tiene nuevas herramientas que se pueden usar para mantener seguro el equipo y además posee nuevas tecnologías que se ejecutan en segundo plano y que consiguen que el equipo trabaje de forma más eficiente y confiable.

2.5.1.1.2 Base de Datos

Para construir la Base de Datos utilizaremos Microsoft SQL Server 2000 ya que nos ofrece un óptimo rendimiento, gran fiabilidad y facilidad de uso. Además Microsoft SQL Server 2000 posee características que le convierten en una excelente plataforma para bases de datos de proceso transaccional en línea (OLTP) y la función Analysis Services de SQL Server 2000 nos brinda excelencia en Servicios OLAP.

2.5.1.1.3 Herramientas de Carga de Datos

Luego de construir la Base de Datos necesitamos Cargar Datos desde las fuentes para esto se selecciono la herramienta de transformación de datos DTS (Data Transformation Services) que esta disponible con Microsoft SQL Server 2000 porque nos ofrece la Transformación de datos para cargarlos a una base de datos existente o para crear una nueva. Además se pueden ejecutar tareas de carga a través de un paquete DTS programado con horarios determinados.

2.5.1.1.4 Transformación Multidimensional

La transformación Multidimensional de los datos del Data Mart de Sismología se hará a través de la herramienta Analysis Services disponible en Microsoft SQL Server 2000 ya sus características ofrecen facilidad de generación, análisis y seguridad de los cubos. Además los usuarios finales pueden hacer uso de muchas de las nuevas características y mejoras, como minería de datos y conexiones HTTP.

2.5.1.1.5 Herramienta OLAP

La herramienta utilizada para obtener la información almacenada en la base de datos Multidimensional es el Open I debido a que es una herramienta de software de libre difusión con lo cual se evita el tener que comprar licencias para utilizar la herramienta, facilitando así el uso para el Instituto, ya que no se tienen los recursos suficientes para adquirir licencias de nuevas herramientas.

2.5.1.1.6 Herramientas de Metadatos

Para la recolección de los metadatos usamos Power Designer 9 y Microsoft Excel 2003 ya que nos prestan facilidades para el diseño de los diferentes diagramas modelos y también nos facilitan la documentación de cada uno de los diagramas realizados con sus distintas herramientas.

2.5.1.2 Construcción de la Base de Datos

Para la construcción de la base de datos se necesitan requerimientos mínimos que aseguren su óptimo funcionamiento y son los siguientes:

2.5.1.2.1 Requerimientos Físicos de la Base de Datos

Inicialmente necesitamos disponer de un espacio en disco para la generación de la base de datos:

- Para el log de transacciones se ocupara 100 MB en disco.
- Para los datos iniciales que se almacenaran en la base de datos del Data Mart de Sismología necesitamos un espacio de 2,5 GB ya que se generarán aproximadamente 20000 registros con una longitud promedio de 128 bytes por cada registro.

2.5.1.2.2 Construcción de los Objetos del Data Mart de Sismología

A continuación procedemos a generar los distintos objetos que integraran el Data Mart corriendo el script que tenemos en el ANEXO F.

2.5.1.3 Extracción, Transformación y Carga de Datos

Ya generada la base de datos debemos proceder a extraer los datos de las distintas fuentes, luego los transformamos y cargamos en los diferentes objetos de la base de datos del Data Mart de Sismología.

2.5.1.3.1 Acceso a las fuentes y Transformación de Datos

Los datos que tenemos almacenados en la base de datos fuente se deben cargar hacia el Data Mart y para esto se accede a los datos usando DTS de Microsoft SQL Server 2000 y luego se toman los datos del origen se adaptan a la definición de las dimensiones y hechos del Data Mart a través de paquetes DTS de Microsoft SQL Server 2000 y a continuación explicamos este proceso:

Los datos que tenemos almacenados en la base de datos fuente se deben cargar hacia el Data Mart y para esto se accede a los datos usando DTS de Microsoft SQL Server 2000 y luego se toman los datos del origen se adaptan a la definición de las dimensiones y hechos del Data Mart a través de paquetes DTS de Microsoft SQL Server 2000 y a continuación explicamos este proceso:

- Creamos un nuevo paquete de transformación de datos, seguido de esto se crea conexiones a las bases de datos del tipo OLE DB para SQL Server, tanto para la Bases de datos intermedia como para el Data Mart

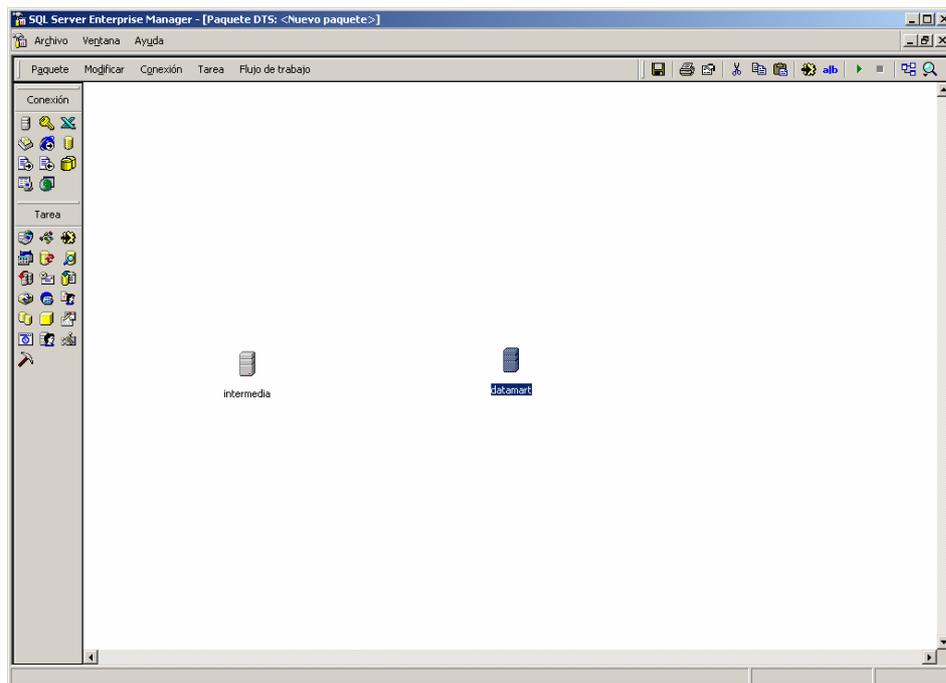


Figura 31 Pantalla de creación del paquete de transformación de datos

- Borramos los que almacenados en la base de datos del Data Mart, y además creamos tareas de transformación de datos para cargar las diferentes tablas desde la base de datos intermedia hacia la base de datos del Data Mart, así

por ejemplo para cargar la tabla de Dimensiones Estacion utilizamos el siguiente proceso:

- o Primero definimos la fuente desde la cual se cargarán los datos así:

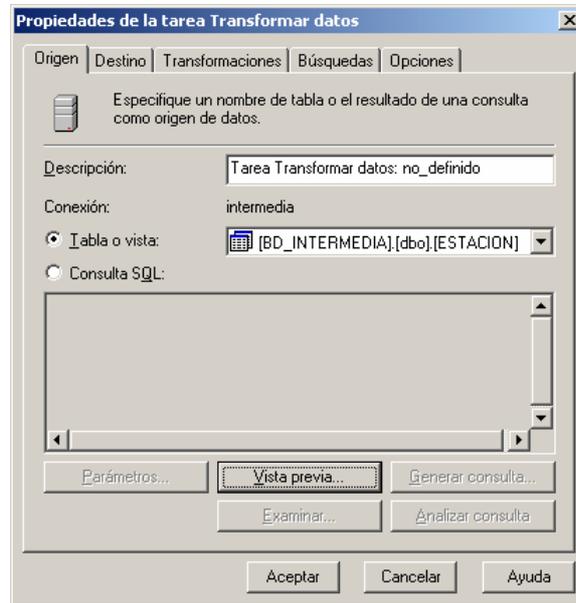


Figura 32 Pantalla de definición del origen de datos

- o Se define el destino de los diferentes datos que fueron tomados en el paso anterior.

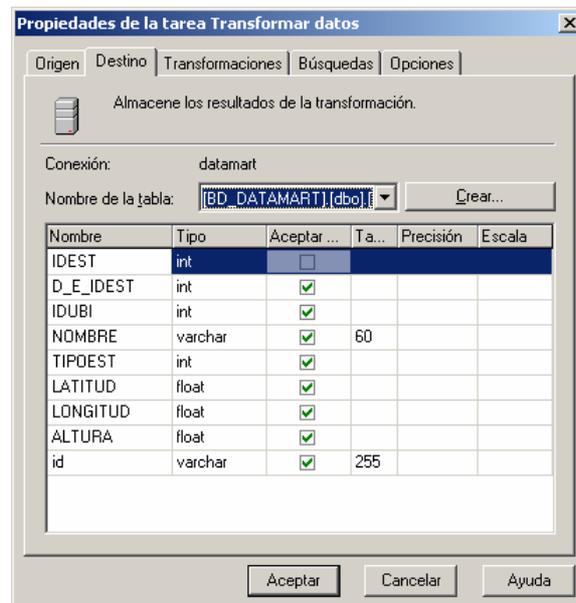


Figura 33 Pantalla de definición del origen de los datos

- Escogemos las diferentes transformaciones que van a surtir efecto desde los datos fuente hacia los datos de destino

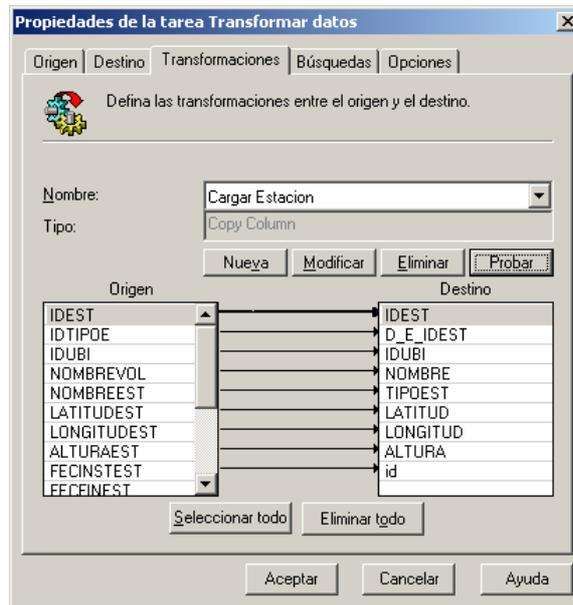


Figura 34 Pantalla de definición de las transformaciones de los datos

- Se prueba la consistencia de las transformaciones, con lo cual nos aseguramos que éstas son correctas y los datos cargados son válidos.
- Se deben crear tareas para todas y cada una de las transformaciones que se van a implementar para cargar los datos desde la base de datos intermedia hacia nuestro Data Mart, teniendo como resultado la base de datos final.

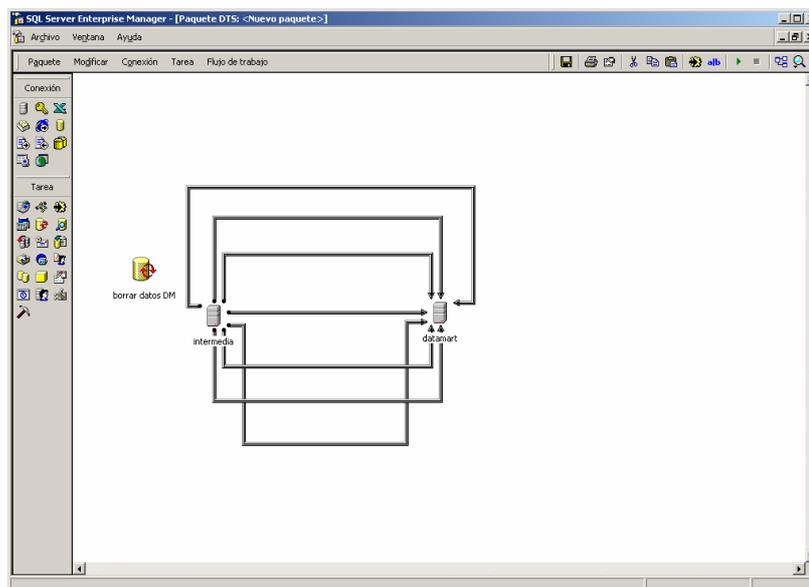


Figura 35 Vista del Paquete DTS

2.5.1.4 Transformación Multidimensional

Luego de tener los datos cargados en las tablas de dimensiones y hechos del Data Mart procedemos a generar los diferentes cubos con la herramienta Analysis Services de Microsoft SQL Server 2000 de tal forma que cubran las necesidades de información que tiene el Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional.

Lo que primero se debe hacer es generar los cubos de la información necesaria para la construcción del Data Mart, para lo cual seguimos los siguientes pasos:

- Ingresamos al asistente para generar cubos



Figura 36 Pantalla del Asistente para la generación de cubos

- Seleccionamos la tabla de hechos.

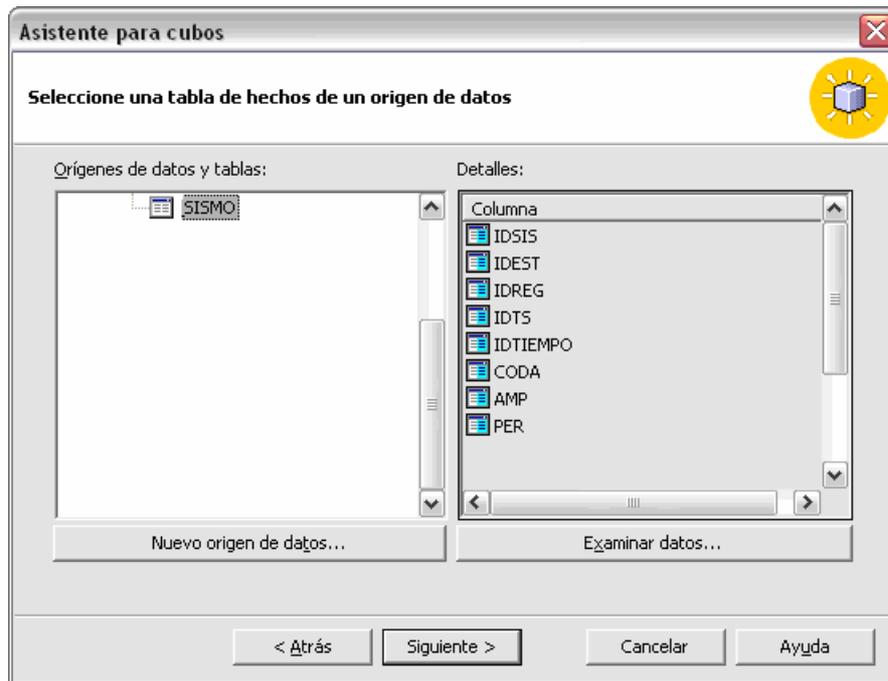


Figura 37 Pantalla de selección de la tabla de hechos

- Seleccionamos las medidas que van a definir al cubo



Figura 38 Pantalla de selección de la columnas numéricas de hechos

- Luego lo que debemos hacer es crear las dimensiones del cubo
 - Creación de la dimensión Tiempo



Figura 39 Pantalla del asistente para generar las dimensiones

- Definimos que tipo de dimensión vamos a crear

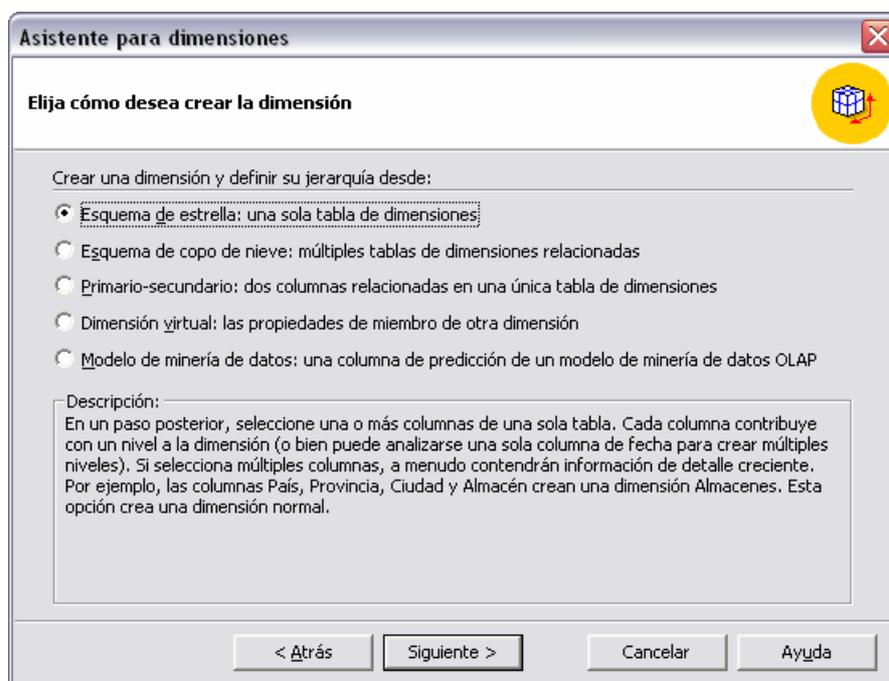


Figura 40 Pantalla de selección de tipo de dimensiones

- Se selecciona la tabla de dimensiones

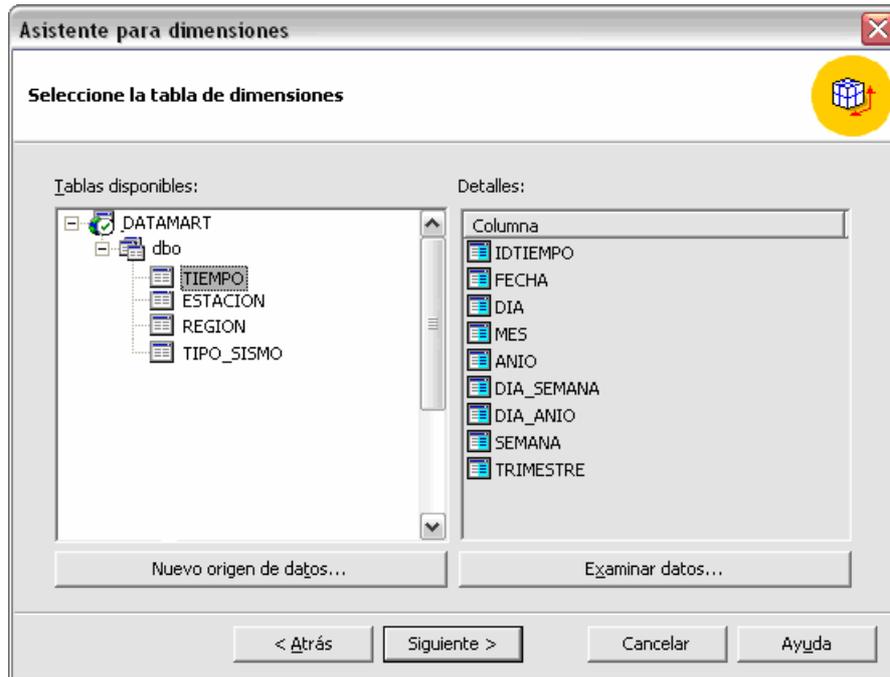


Figura 41 Pantalla de selección de las tablas de dimensiones

- En este caso como la dimensión escogida es el tiempo, por defecto el Análisis Services de SQL 200 permite trabajar con la dimensión tiempo

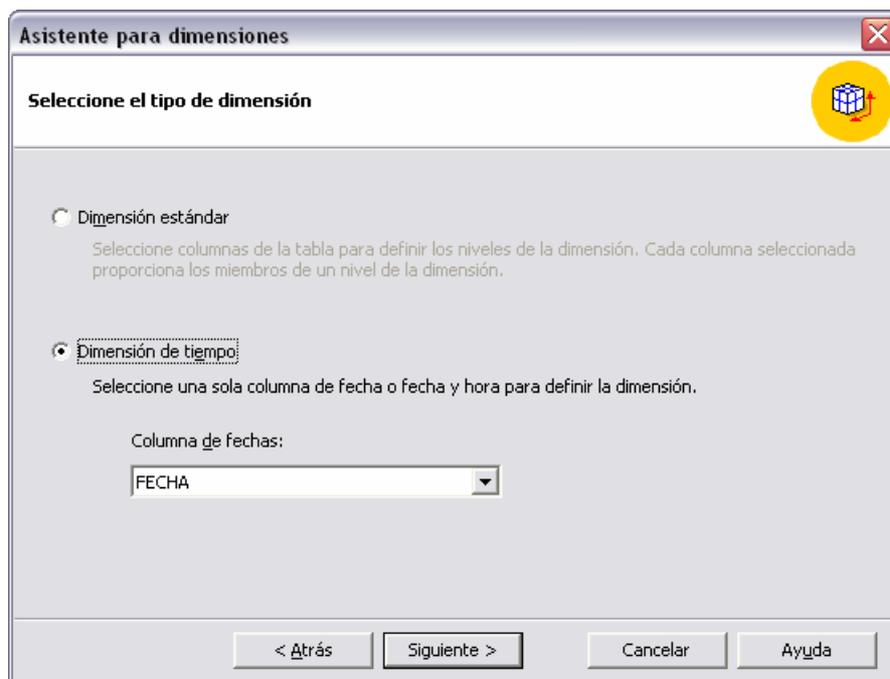


Figura 42 Pantalla de selección de la dimensión sobre la que trabajaremos

- o Se debe especificar los niveles de la dimensión los mismos que nos permitirán hacer un filtrado de los datos por niveles de jerarquía.

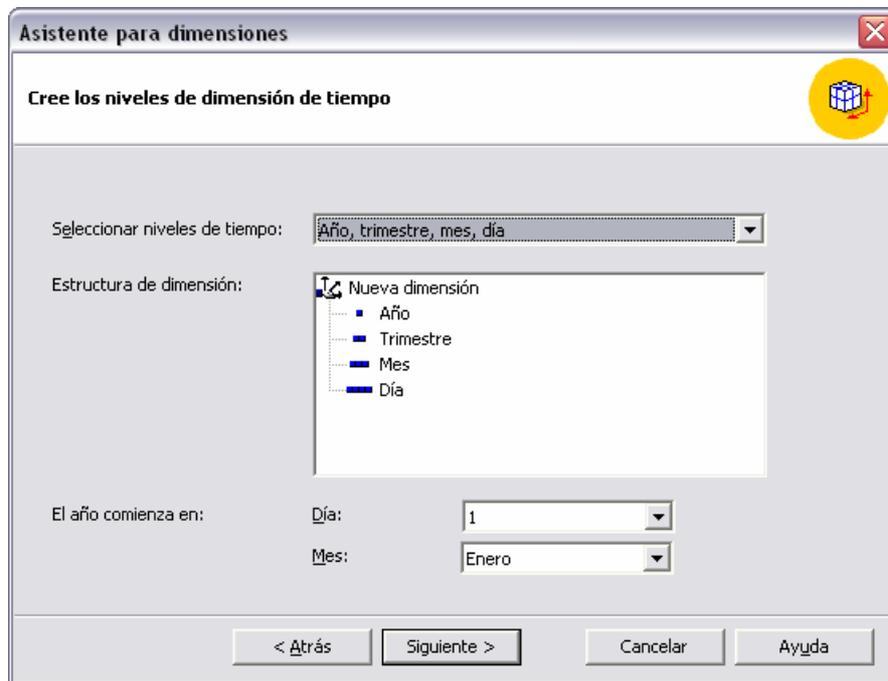


Figura 43 Pantalla de selección de grado de granularidad de la dimensión tiempo

- o Por último se nombra a la dimensión.



Figura 44 Pantalla de presentación de la dimensión Tiempo

- Creación de la Dimensión Región
 - Seleccionamos la tabla de dimensiones

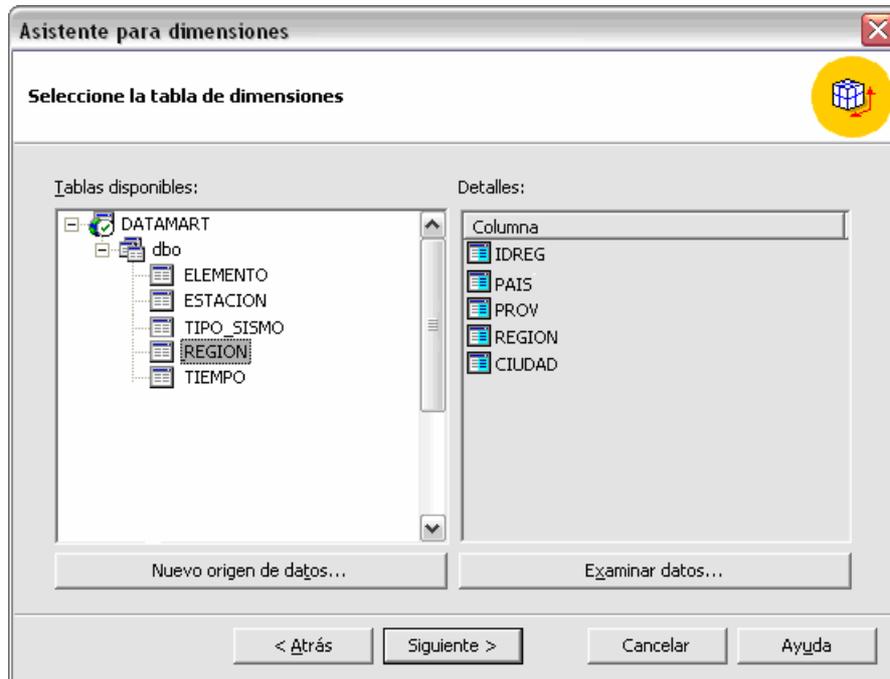


Figura 45 Pantalla de selección de la dimensión sobre la que trabajaremos

- Seleccionamos los niveles para los campos de la dimensiones

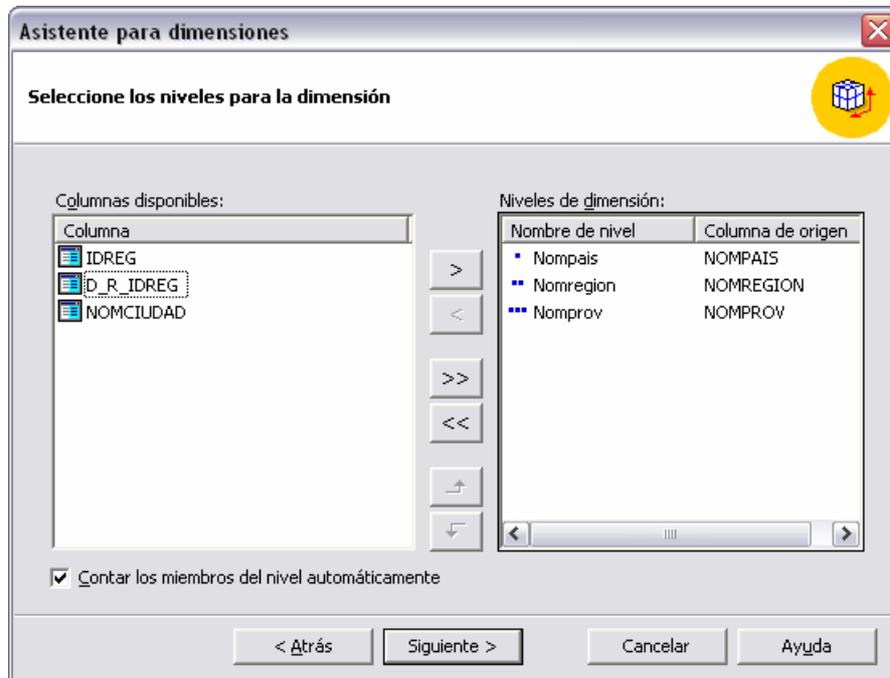


Figura 46 Pantalla de selección de grado de granularidad de la dimensión
Región

- Finalmente damos un nombre a la dimensión

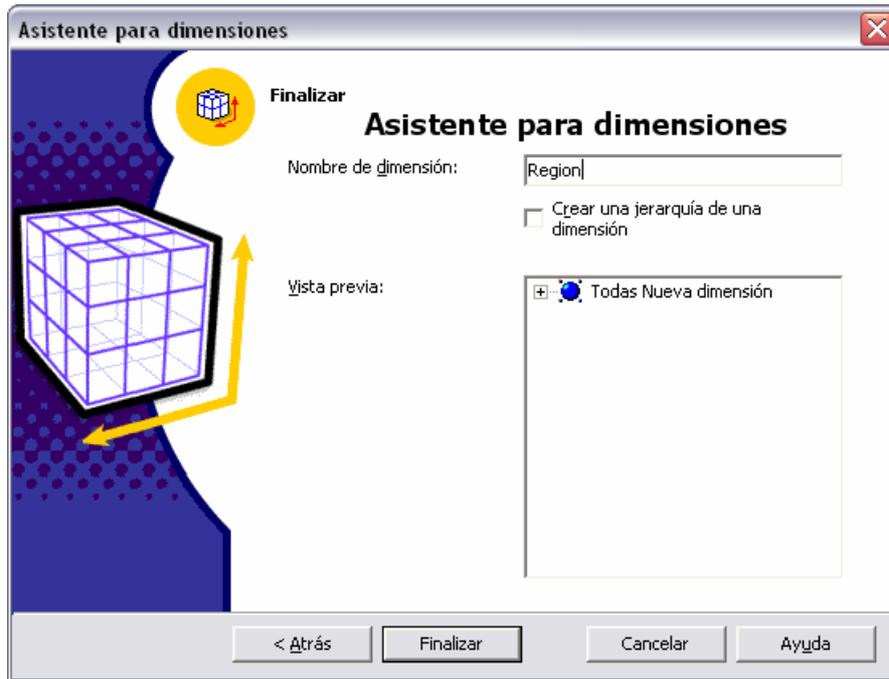


Figura 47 Pantalla de presentación de la dimensión Región

- Luego de haber terminado con las dimensiones, se procede a nombrar al cubo

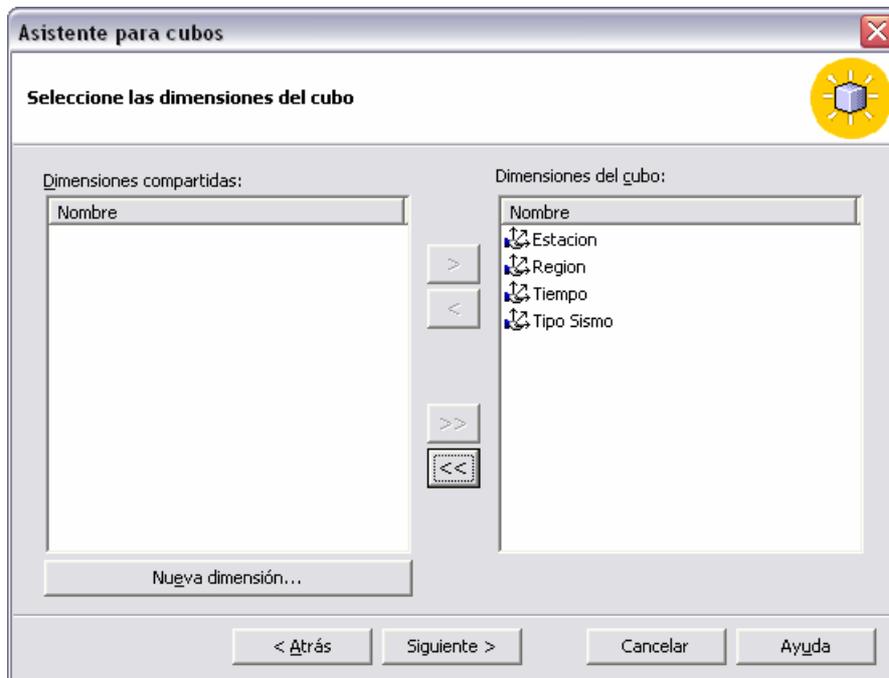


Figura 48 Pantalla de selección de las dimensiones que contendrá el cubo



Figura 49 Pantalla de presentación del Cubo de Sismología

2.5.1.5 Visualización de Datos

Para la visualización de los datos se utilizó la herramienta Open I, la misma que se la utiliza vía web, ingresando en un browser la dirección en la que se encuentra el servidor y haciendo referencia al puerto 8080, como se muestra en la figura.

Es aquí donde se debe ingresar ya sea como administrador para crear nuevos módulos, o como un usuario que tenga privilegios solo para desplegar información. Para el caso se procederá a ingresar como administrador.



Figura 50 Pantalla de inicio de la herramienta OPEN I

Luego de haber ingresado como administrador, se puede observar los diferentes módulos que se tiene en el servidor Olap.

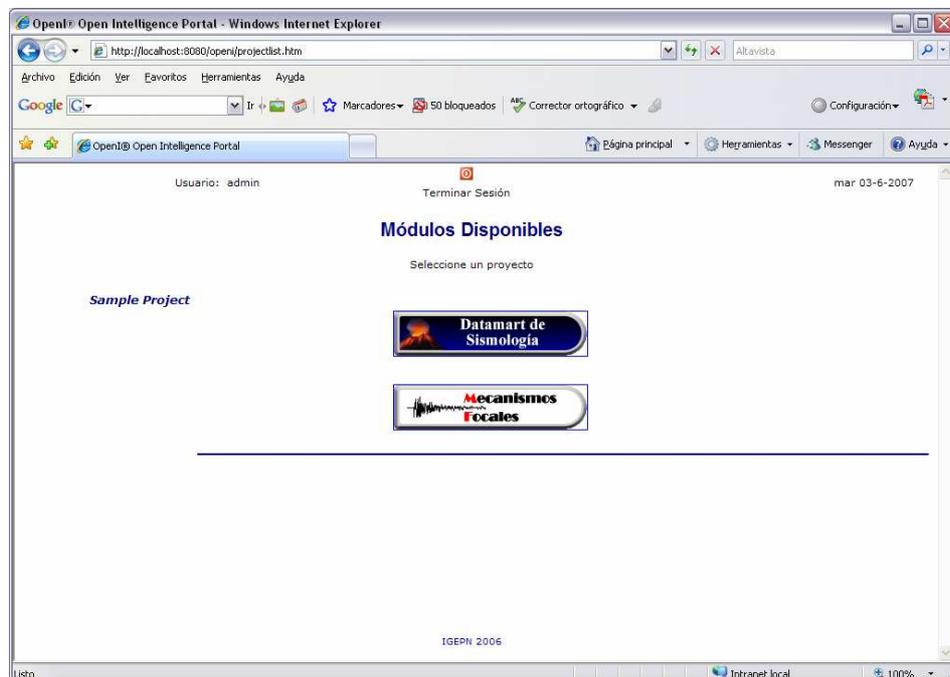


Figura 51 Pantalla de presentación de los proyectos del servidor OLAP

Se procede a configurar la administración del Proyecto, para lo cual se ingresa al link *Administration -> Manage Project* como se muestra en la figura.

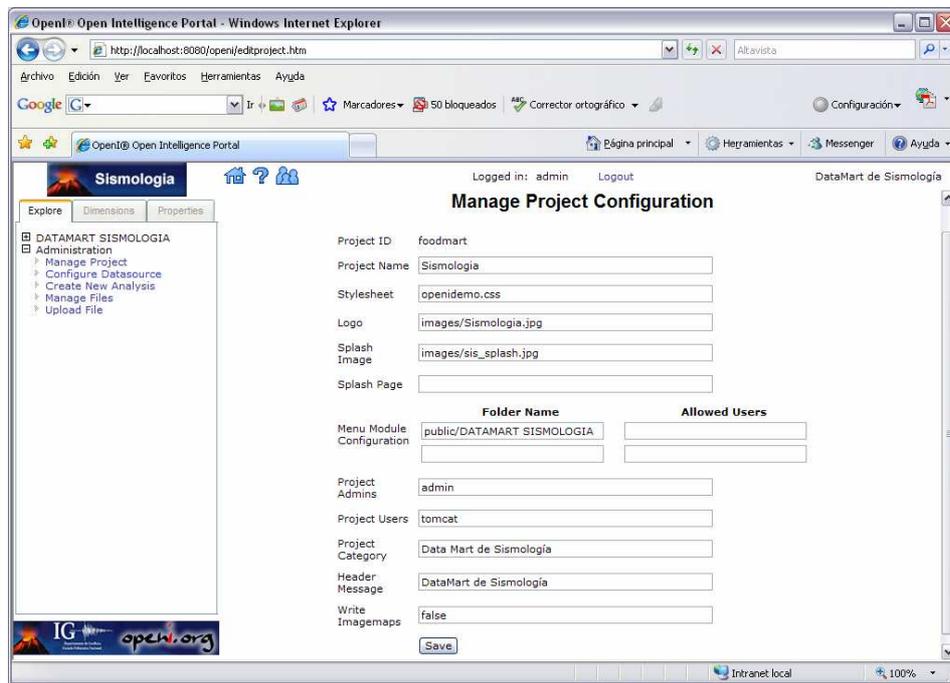


Figura 52 Pantalla de ingreso para administrar un proyecto

Se edita desde donde se va a obtener los datos, para lo cual ingresamos a la pantalla que se muestra a continuación.

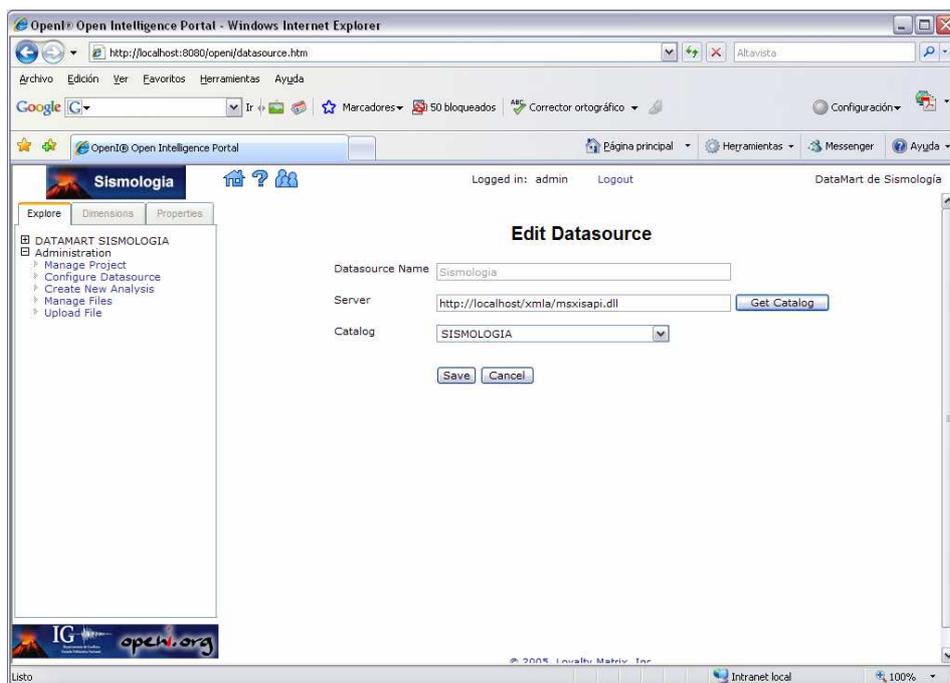


Figura 53 Pantalla de configuración de la fuente de datos

Luego de haber configurado la fuente desde la cual se obtendrán los datos, se procede a configurar las tablas de dimensiones y de hechos, indicando los diferentes filtros de búsqueda de la siguiente manera:

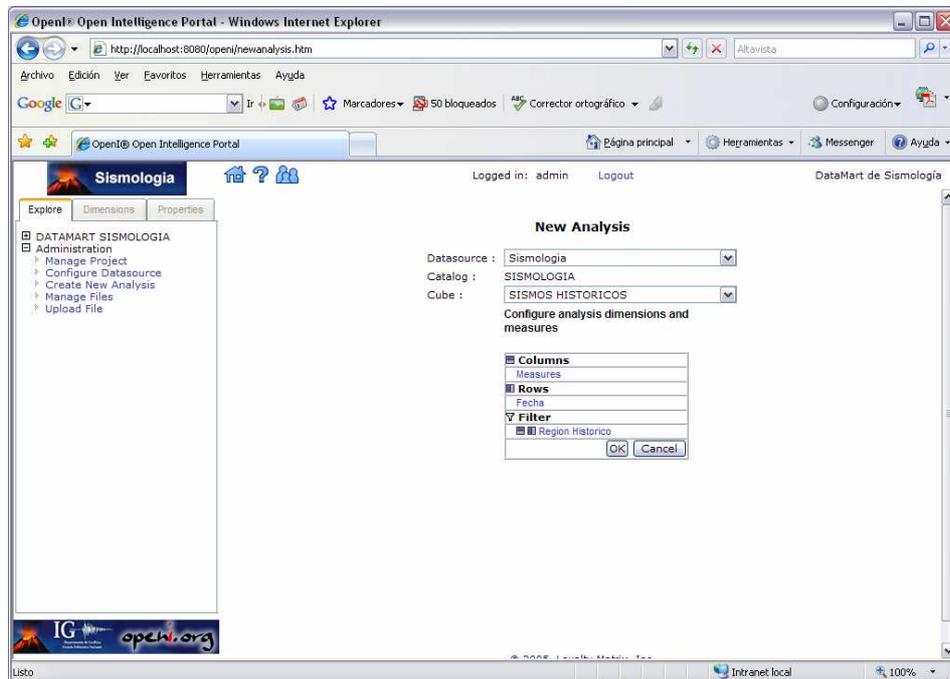


Figura 54 Pantalla de creación de filtros para generación de datos

Por ultimo, teniendo configurado absolutamente todo lo que concierne a las fuentes de datos y las dimensiones, se observan los datos como se lo configuró así:

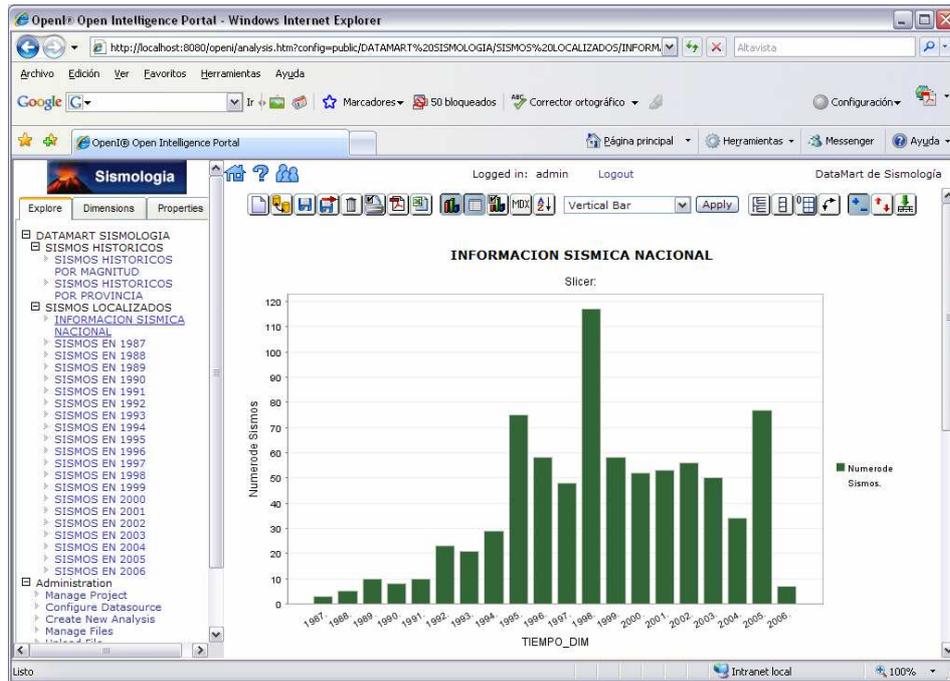


Figura 55 Pantalla inicial de visualización de la información del Data Mart de Sismología

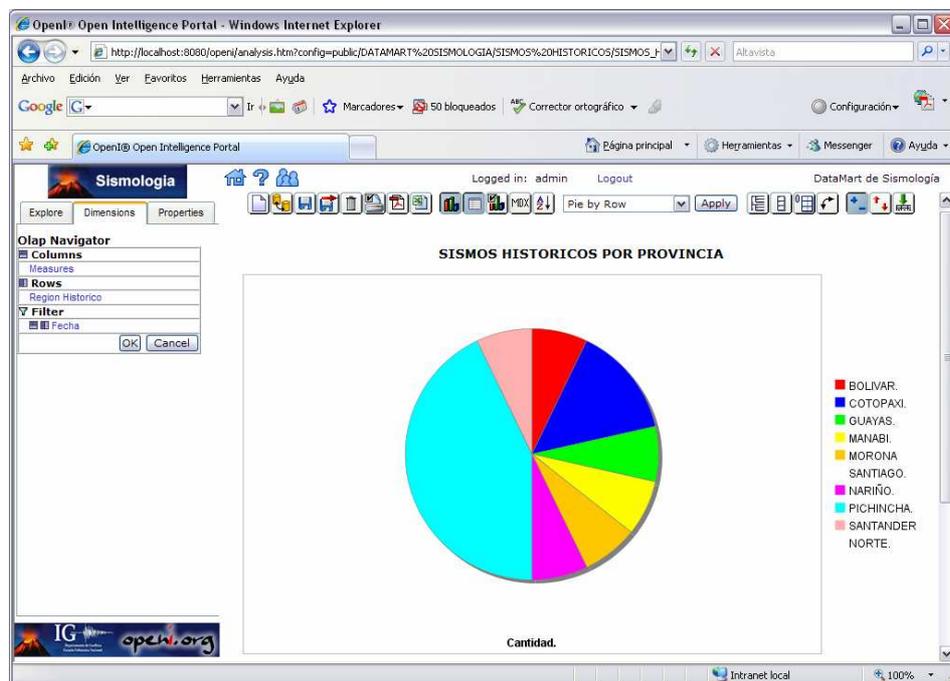


Figura 56 Pantalla de visualización de la información del Data Mart de Sismología con barras de herramientas para modificarlas consultas

Una de las opciones que nos brinda el Open I es el poder observar las consultas MDX que se realizan al servidor Olap, para lo cual ingresamos a la opción Properties del menú principal como se lo muestra a continuación

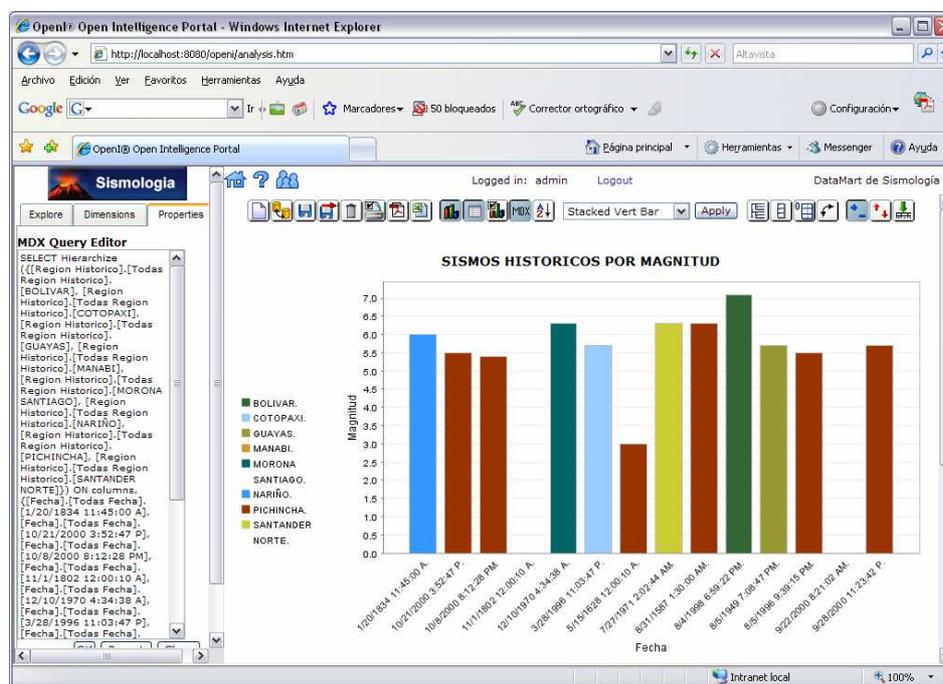


Figura 57 Pantalla de visualización de la información del Data Mart de Sismología con visualización del texto de la consulta que se visualiza

2.5.2 IMPLANTACIÓN

2.5.2.1 Instalación Inicial de las Herramientas

Instalación inicial de las herramientas necesarias:

Para iniciar con la implantación instalaremos Microsoft SQL Server 2000 que será nuestra principal herramienta para la construcción del Data Mart de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional, proceso que lo describimos a continuación:

- Insertamos el CD de instalación de Microsoft SQL Server 2000; en caso de que no se presente directamente la pantalla de instalación se procederá a correr el archivo AUTORUN que tenemos en el CD. Y a continuación seleccionamos el vínculo Componentes de SQL Server 2000.



Figura 58 Menú de Instalación de SQL Server 2000

- Al obtener la pantalla siguiente procedemos a instalar los componentes de SQL Server que necesitamos; en nuestro caso es Servidor de Base de Datos y el Análisis Services:



Figura 59 Menú de componentes de SQL Server

- Procedemos con la instalación del Servidor de Bases de Datos utilizando el Asistente:



Figura 60 Pantalla de inicio de la Instalación de SQL Server

- Seleccionamos el tipo de equipo en el que instalaremos SQL Server 2000 en nuestro caso Equipo local.

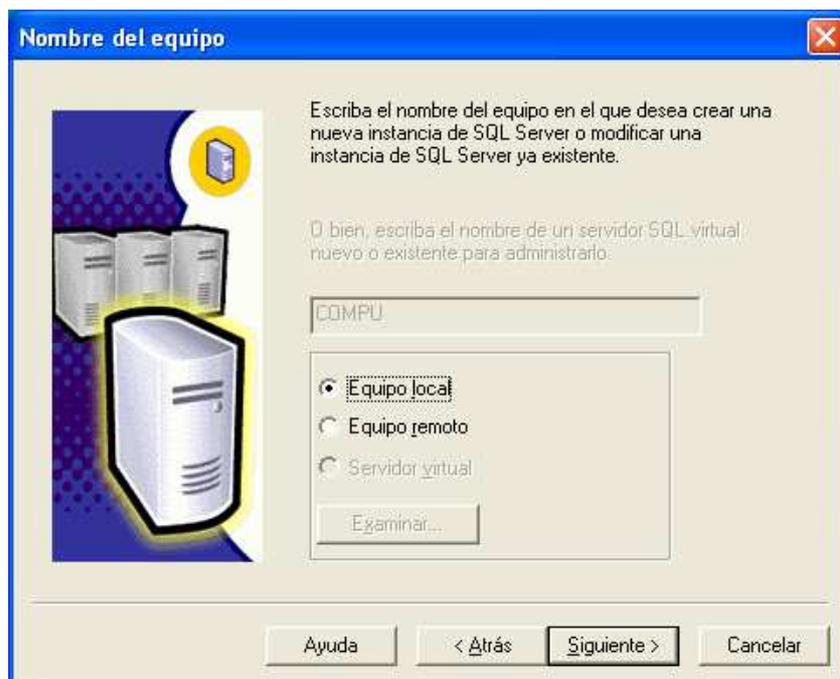


Figura 61 Pantalla de selección del equipo a instalar

- A continuación procedemos a instalar las herramientas necesarias en nuestro caso la opción Crear una nueva instancia de SQL Server:



Figura 62 Pantalla de selección de la instalación

- Para instalar SQL Server 2000 debemos especificar el nombre del propietario:



Figura 63 Pantalla de ingreso de la información de usuario

- Debemos aceptar el contrato de licencia de software para continuar con la instalación:

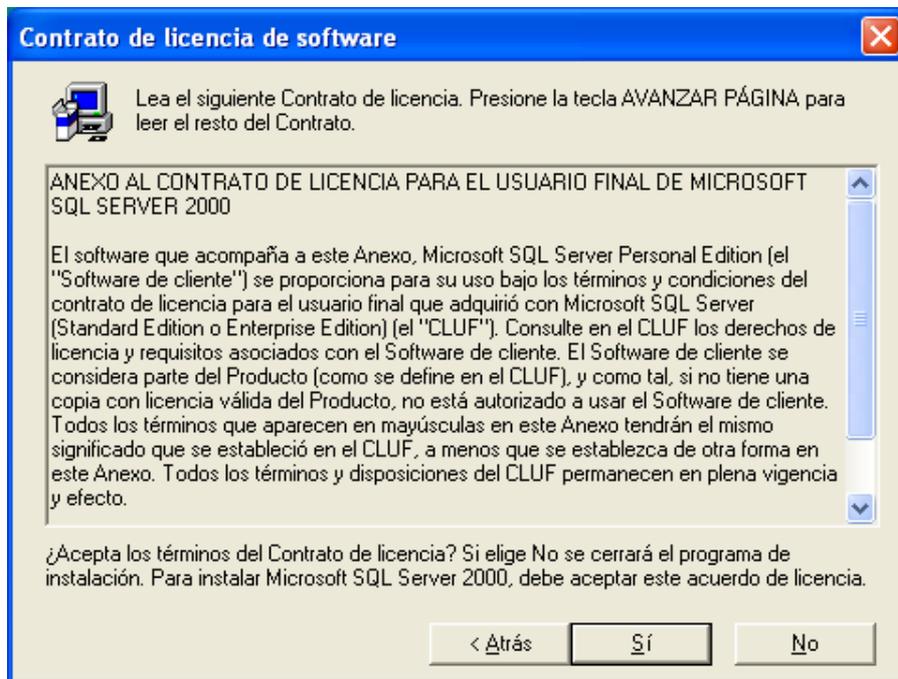


Figura 64 Pantalla del Contrato de Licencia de SQL Server 2000

- Luego seleccionaremos lo que deseamos instalar en nuestro caso instalaremos herramientas de cliente y servidor:



Figura 65 Pantalla de selección del tipo de instalación

- Podemos agregar nombres a las instancias de bases de datos que se instalaran o generaran o seleccionar como nosotros la instalación predeterminada:



Figura 66 Pantalla para configurar el nombre de la instancia de SQL Server

- Ahora seleccionaremos el tipo de instalación que se realizará en nuestro caso personalizada ya que instalaremos todos los componentes de bases de datos:

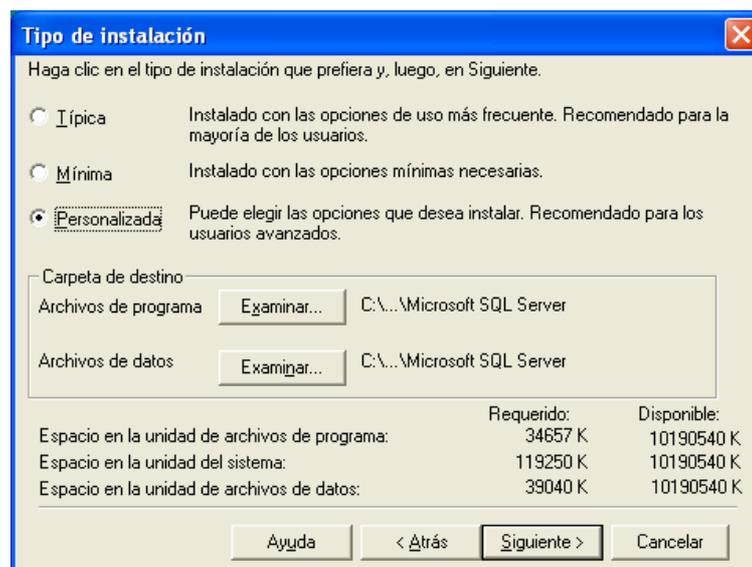


Figura 67 Pantalla para seleccionar el tipo de instalación

- A continuación debemos indicar los detalles de las cuentas de cada uno de los servicios de SQL Server o como en nuestro caso seleccionar que todos los servicios se inicien con la misma cuenta y utilizando la cuenta del sistema local:

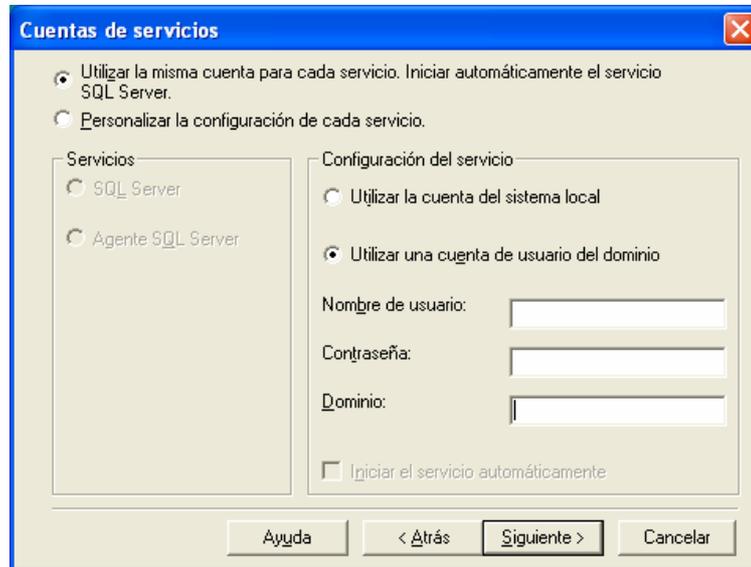


Figura 68 Pantalla para especificar las cuentas de los servicios de SQL Server

- Debemos también especificar el tipo de autenticación que se usará para nuestro servidor en nuestro caso tenemos distintos tipo de usuarios y además la autenticación de Windows:

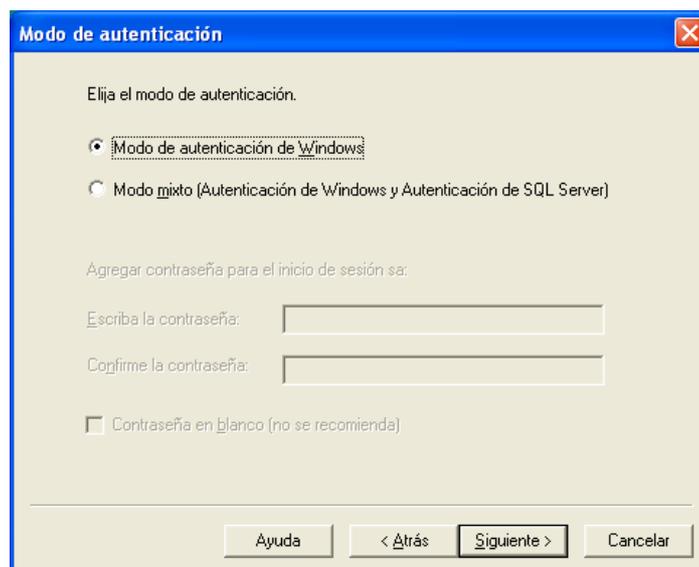


Figura 69 Pantalla para especificar el modo de autenticación

- Luego de la pantalla que tenemos a continuación se procede con la copia de los archivos para la instalación de Microsoft SQL Server 2000:



Figura 70 Pantalla de información del inicio de la copia de archivos

- Luego se copian los diferentes archivos de SQL Server 2000:

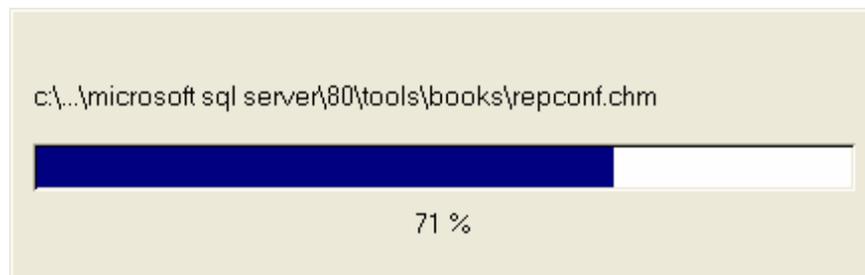


Figura 71 Indicador de la progresión de la copia de archivos

- Y el Servidor de Bases de Datos de SQL Server esta instalado como nos informa la imagen siguiente:



Figura 72 Pantalla de información de un exitoso proceso de instalación del servidor de bases de datos de SQL Server

- Y a continuación procedemos a instalar el Análisis Services de SQL Server con el que generaremos los cubos del Data Mart:



Figura 73 Pantalla de Inicio de instalación del Análisis Services de SQL Server 2000

- Debemos aceptar el contrato de licencia de software para continuar con la instalación:

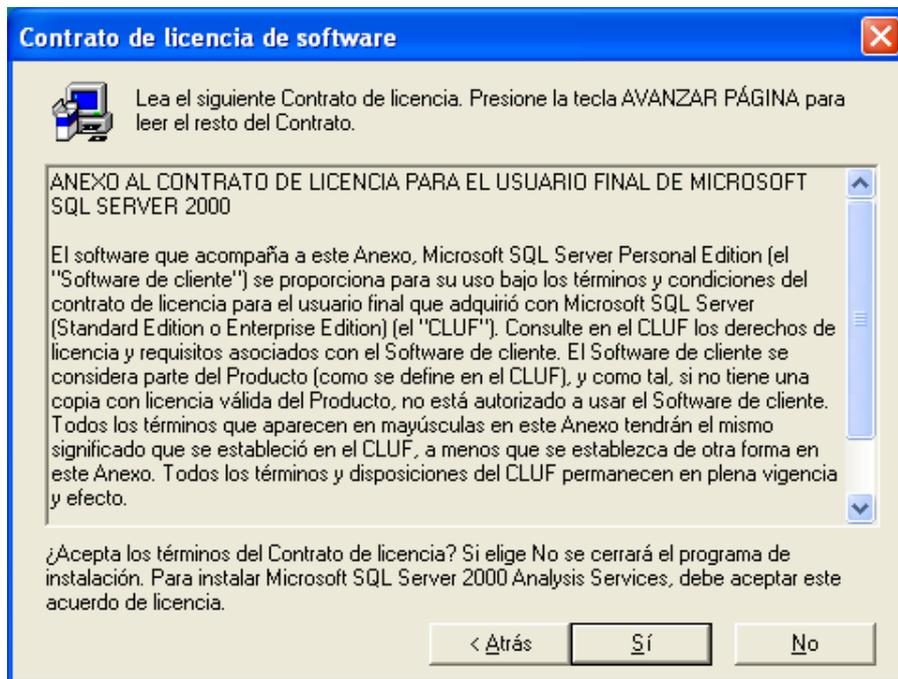


Figura 74 Pantalla del Contrato de Licencia de SQL Server 2000

- Luego seleccionaremos los componentes que instalaremos:



Figura 75 Pantalla de selección de componentes a instalar

- Ahora especificaremos la carpeta en el menú de inicio en el que se agregará los accesos al Análisis Services:



Figura 76 Pantalla de selección de la carpeta donde se desplegará los accesos al Análisis Services

- Y el Análisis Services de SQL Server esta instalado como nos informa la imagen siguiente:



Figura 77 Pantalla de información de un exitoso proceso de instalación del Análisis Services de SQL Server

- Ahora podemos observar en el Menú Inicio los accesos a las distintas herramientas de Microsoft SQL Server 2000:

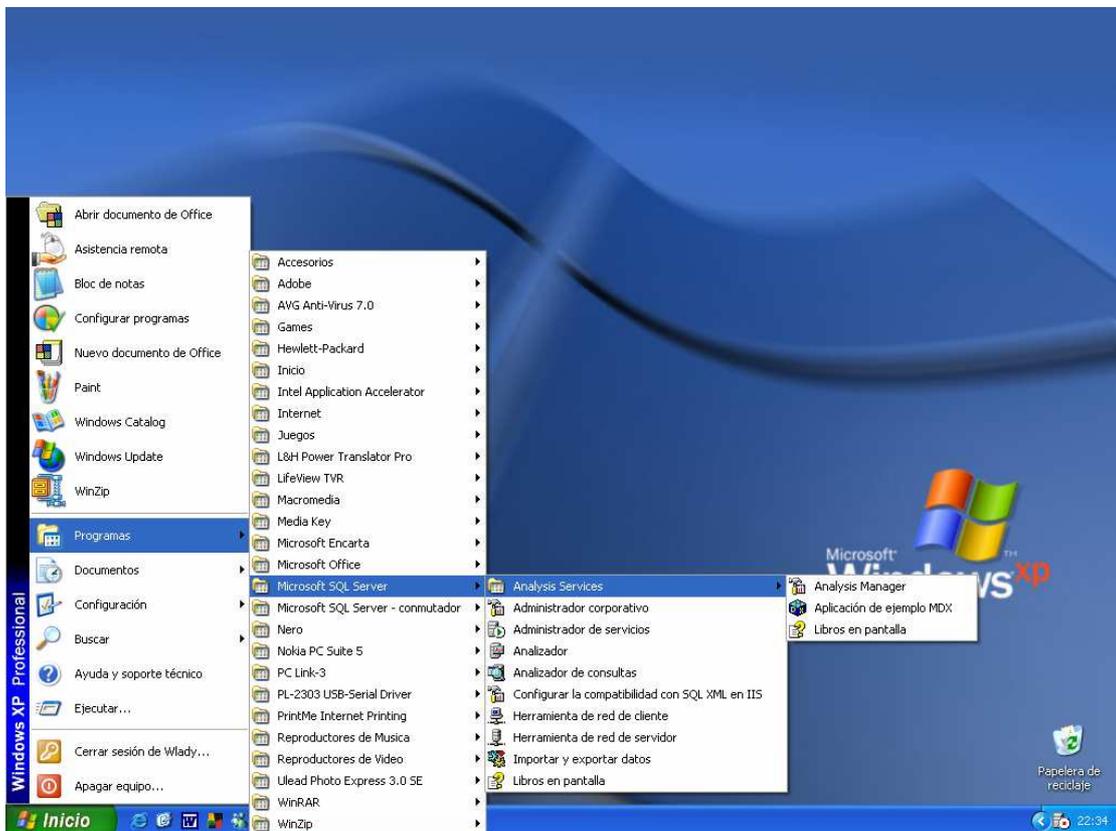


Figura 78 Menú Inicio con los nuevos accesos a SQL Server

- Para comenzar a utilizar SQL Server debemos iniciar los servicios de SQL y se hace seleccionando en el menú inicio el Administrador de Servicios:



Figura 79 Pantalla del Administrador de Servicios de SQL Server

- Ya en el Administrador de Servicios de SQL Server iniciamos cada uno de los servicios dando clic sobre iniciar o continuar:



Figura 80 Pantalla ya con los Servicios iniciados

También debemos proceder a instalar la herramienta Open I que se usará para la de visualización de la información del Data Mart de Sismología, la misma que tiene como requerimiento previo la instalación de aplicaciones adicionales para su funcionamiento, las que se detallan a continuación:

- *Java 2 SDK* es una herramienta de libre difusión lo misma que nos permitirá ejecutar el servidor Apache.
 - Al ejecutar el instalador de Java 2 SDK, se despliega un asistente el mismo que nos guiará en el transcurso de la instalación.



Figura 81 Asistente de instalación de Java 2 SDK

- Debemos escoger cuales son las aplicaciones que vamos a instalar

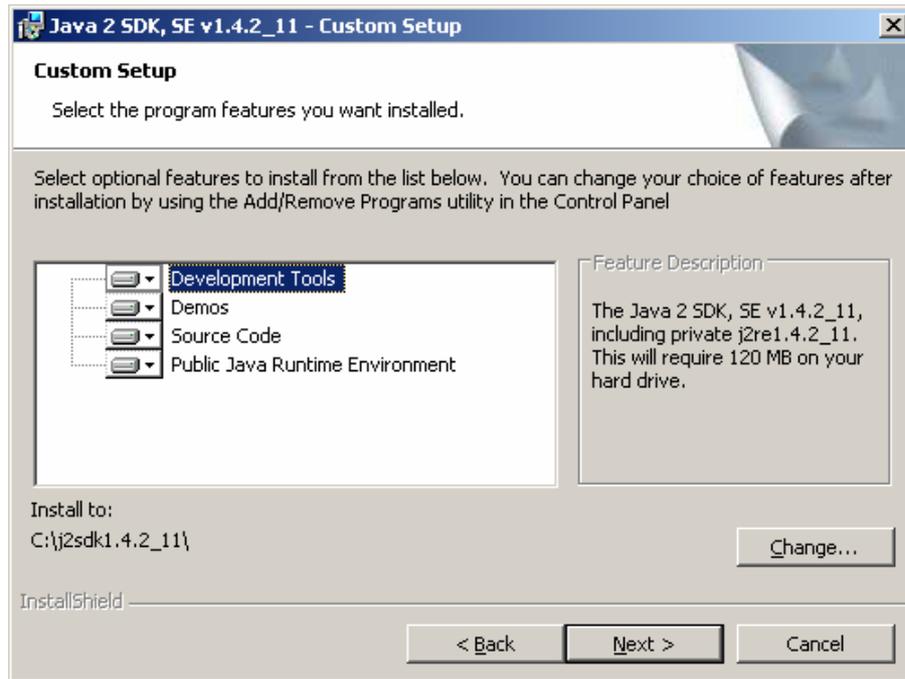


Figura 82 Pantalla de selección de aplicaciones a instalar

- Luego se descarga la información escogida a nuestro disco

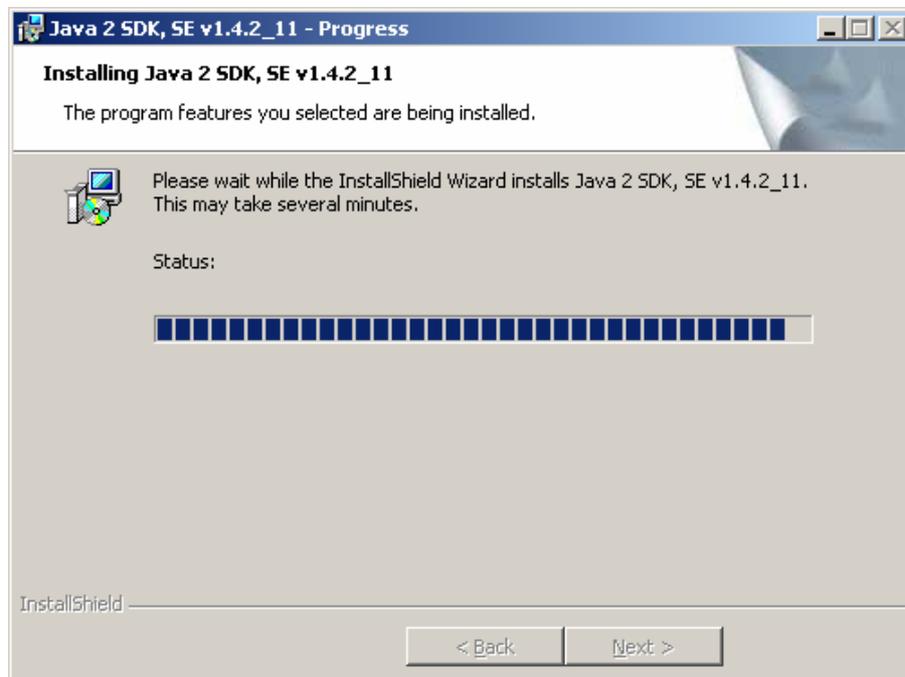


Figura 83 Pantalla de copia de archivos de instalación

- *Apache Tomcat* servidor de aplicaciones que permite publicar paginas jsp, formato en el cual se encuentra desarrollado el openi
 - Al igual que con el Java, se tiene un asistente para la instalación de Apache Tomcat.

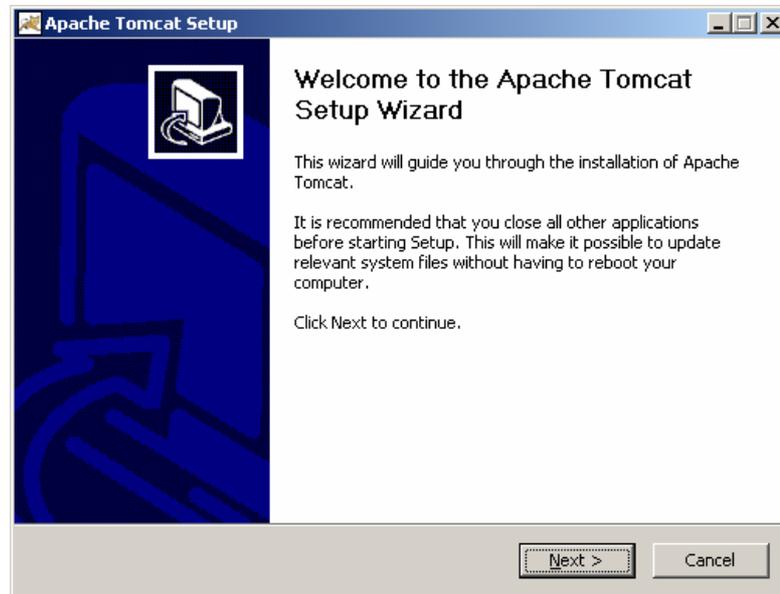


Figura 84 Asistente de instalación de Apache Tomcat

- Es muy importante el ingresar la clave de administrador para apache, debido a que openi hace referencia a esta para el ingreso a sus módulos

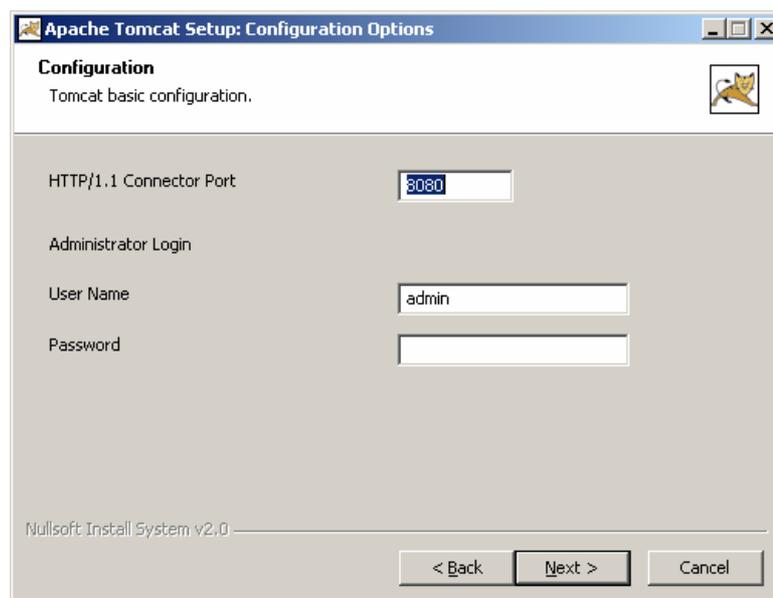


Figura 85 Pantalla de Configuración de usuario de Apache

- *MSXML* es un protocolo que permite conectar a openi con servidores OLAP Microsoft, es nuestro caso Microsoft Analysis Services (msas)

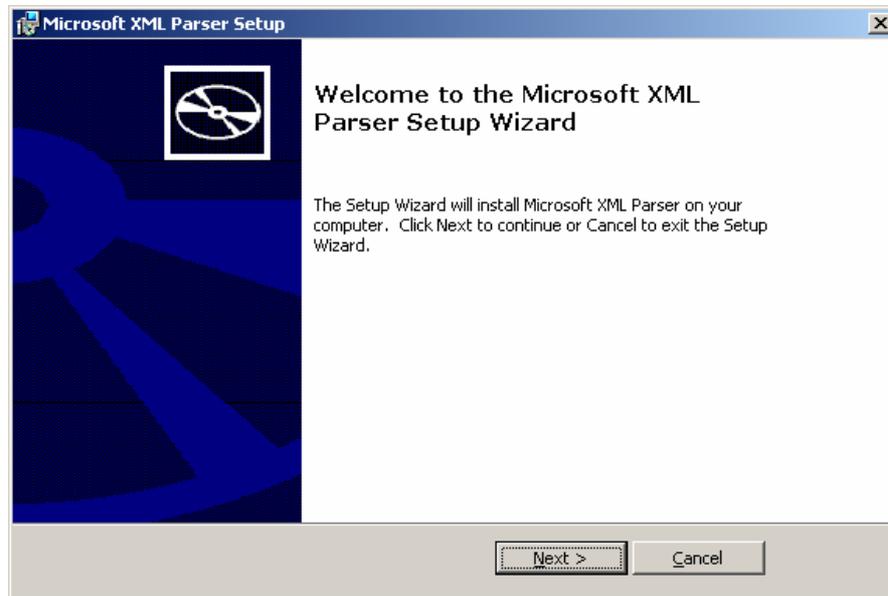


Figura 86 Asistente de instalación de Microsoft XML

Por ultimo se instala el Open I, el primer paso es obtener el código fuente. Luego se debe copiar los siguientes archivos con la extensión war:

- openi.war
- openi-projects.war
- mondrian.war

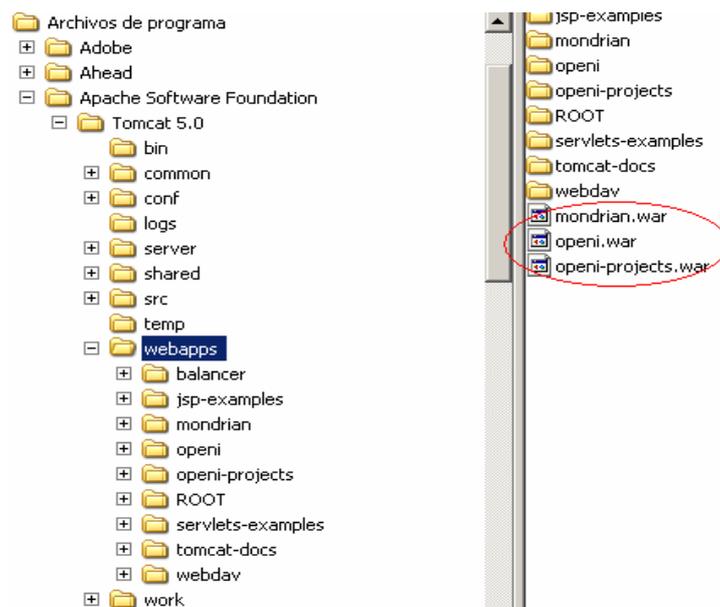


Figura 87 Muestra de Archivos copiados

Estos archivos deben ser colocados en la carpeta *webapps* la misma que podemos ubicarla en el sistema de archivo en el cual instalamos el servidor de aplicaciones Apache.

Seguido de esto, se procede a iniciar el servidor de aplicaciones si este no se está ejecutando, caso contrario se lo reinicia, con lo cual Apache genera el código necesario para que el Open I funcione.

2.6 PRUEBAS

Para evaluar el funcionamiento del Data Mart de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional vamos a realizar los siguientes tipos de pruebas:

- Pruebas de Unidad
- Pruebas de Integración
- Pruebas de Validación
- Pruebas del Sistema

2.6.1 PRUEBAS DE UNIDAD

PRUEBA DE UNIDAD 1

Fecha de Realización: 14 de Marzo del 2006

Responsable: Michael Vizuite

Etapas de Prueba: Prueba de Unidad

Nombre del Proceso: Verificar que todos los paths posibles en las distintas conexiones sean ejecutados

Tipo de Prueba: Pruebas de Path

Técnicas y Herramientas Utilizadas: Se evaluarán revisando que los paths en cada uno de los procedimientos sean correctos.

Entradas: Paths

Salidas: Conexiones exitosas

Resultados Esperados: Que todos los paths sean ejecutados.

Porcentaje de Aprobación: 100%

Observaciones: Todos los paths fueron ejecutados correctamente de tal manera que el funcionamiento es óptimo.

PRUEBA DE UNIDAD 2

Fecha de Realización: 14 de Marzo del 2006

Responsable: Michael Vizuite

Etapas de Prueba: Prueba de Unidad

Nombre del Proceso: Verificar que en el Data Mart no existan valores erróneos.

Tipo de Prueba: Pruebas de Manejo de errores

Técnicas y Herramientas Utilizadas: Se evaluarán revisando mediante consultas la consistencia de los datos

Entradas: Consultas SQL

Salidas: Datos consultados

Resultados Esperados: Que los datos sean consistentes y sin errores

Porcentaje de Aprobación: 100%

Observaciones: Todos los datos cargados en el Data Mart son correctos y consistentes lo cual asegura que no existirán errores en las consultas desplegadas.

2.6.2 PRUEBAS DE INTEGRACIÓN**PRUEBA DE INTEGRACION 1**

Fecha de Realización: 14 de Marzo del 2006

Responsable: Michael Vizuite

Etapas de Prueba: Pruebas de Integración

Nombre del Proceso: Probar cada una de las consultas y procedimientos para asegurarse que hacen lo que deben hacer.

Tipo de Prueba: Pruebas de Escritorio

Técnicas y Herramientas Utilizadas: Se evaluarán revisando que las consultas entreguen lo que se solicita

Entradas: Consultas SQL

Salidas: Datos consultados

Resultados Esperados: Que todos los procedimientos y las consultas hagan lo que adecuado para el correcto funcionamiento del sistema.

Porcentaje de Aprobación: 100%

Observaciones: Todo los procedimientos y consultas hacen lo solicitado asegurando un correcto funcionamiento de la solución Data Mart.

2.6.3 PRUEBAS DE VALIDACIÓN

PRUEBA DE VALIDACION 1

Fecha de Realización: 14 de Marzo del 2006

Responsable: Michael Vizuite, Ing. Mónica Segovia e Ing. Jorge Aguilar

Etapas de Prueba: Pruebas de Validación y Configuración

Nombre del Proceso: Definir todos los procesos que realiza el sistema y detallar cada requerimiento a cumplirse

Tipo de Prueba: Pruebas de Validación

Técnicas y Herramientas Utilizadas: Se evaluarán revisando que la solución Data Mart no permita ingresar valores nulos o que no cumplan con los formatos establecidos en su almacenamiento.

Entradas: Procesos

Salidas: Mensajes de error debido a que no se permite el ingreso de dichos valores en las tablas

Resultados Esperados: Que no se pueda ingresar valores incorrectos en sus Bases de datos

Porcentaje de Aprobación: 100%

Observaciones: Todas las tablas no permiten ingresar valores que no cumplan con sus formatos lo cual nos asegura la consistencia del Data Mart.

PRUEBA DE VALIDACION 2

Fecha de Realización: 15 de Marzo del 2006

Responsable: Carlos Yela e Ing. Jorge Aguilar

Etapas de Prueba: Prueba de Validación y Configuración

Nombre del Proceso: Verificar la existencia de todos los documentos mínimos necesarios después de haber culminado cada fase de implementación.

Tipo de Prueba: Pruebas de Configuración

Técnicas y Herramientas Utilizadas: Se verificará que cada uno de los documentos creados en el proceso de desarrollo de la solución Data Mart de Sismología sean los correspondientes para la aplicación que se está evaluando.

Entradas: Documentos

Salidas: No existen

Resultados Esperados: Que todos los documentos que tenemos sean los que se crearon en el proceso de desarrollo del Data Mart que se está evaluando.

Porcentaje de Aprobación: 80%

Observaciones: Varios de los documentos que se han generado durante el proceso de desarrollo tuvieron que ser desechados debido a cambios de última hora en el Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional.

2.6.4 PRUEBAS DEL SISTEMA

PRUEBA DEL SISTEMA 1

Fecha de Realización: 15 de Marzo del 2006

Responsable: Carlos Yela

Etapas de Prueba: Prueba del Sistema

Nombre del Proceso: Instalación en el Servidor

Tipo de Prueba: Pruebas de Verificación de la Instalación

Técnicas y Herramientas Utilizadas: Esta prueba se instalará nuevamente en otra máquina similar al servidor todos los componentes que aseguren el correcto funcionamiento de la solución Data Mart de Sismología.

Entradas: Como entrada se van a tener los instaladores del software que se usara para el funcionamiento del Data Mart

Salidas: La instalación exitosa del software que se usara para el funcionamiento del Data Mart

Resultados Esperados: Se espera obtener la instalación completa del software en el servidor.

Porcentaje de Aprobación: 100%

Observaciones: El total del software se instalo correctamente en el servidor.

PRUEBA DEL SISTEMA 2

Fecha de Realización: 16 de Marzo del 2006

Responsable: Carlos Yela

Etapas de Prueba: Prueba del Sistema

Nombre del Proceso: Visualización de datos a través de la herramienta OLAP

Tipo de Prueba: Pruebas del Cliente

Técnicas y Herramientas Utilizadas: Para la realización de esta prueba se utilizará la herramienta Open I la cual nos permite visualizar los datos de los diferentes cubos del Data Mart.

Entradas: Como entrada se va a tener un browser

Salidas: Información generada desde cada uno de los cubos visualizada a través de los browsers

Resultados Esperados: Se espera obtener una clara visualización de la información requerida.

Porcentaje de Aprobación: 100%

Observaciones: Dentro de los resultados que se han podido obtener acerca de las interfaces de usuario tenemos:

- El sistema permite que el usuario pueda visualizar la información en distintas formas
- La navegación dentro del browser es fácil.
- Todo el contenido de la información se encuentra dentro de las ventanas que se encuentra utilizando.

PRUEBA DEL SISTEMA 3

Fecha de Realización: 16 de marzo del 2006

Responsable: Carlos Yela

Etapas de Prueba: Prueba del Sistema

Nombre del Proceso: Pruebas de Volumen

Tipo de Prueba: Pruebas del Servidor - Pruebas de Volumen

Técnicas y Herramientas Utilizadas: Para la realización de estas se realizarán sobre el servidor múltiples transacciones para que el servidor procese.

Entradas: Transacciones múltiples hacia el servidor.

Salidas: No hay salidas

Resultados Esperados: Se espera que el servidor funcione adecuadamente.

Porcentaje de Aprobación: 100%

Observaciones: Dentro de los resultados que se han podido tener:

Luego de realizar múltiples transacciones se puede notar que no se presentan cuellos de botella dentro del servidor.

Los tiempos de respuesta no disminuyen en comparación con tiempos de respuesta con una carga normal del sistema.

PRUEBA DEL SISTEMA 4

Fecha de Realización: 16 de Marzo del 2004

Responsable: Carlos Yela

Eta de Prueba: Prueba del Sistema

Nombre del Proceso: Claves Primarias

Tipo de Prueba: Pruebas de Bases de Datos

Técnicas y Herramientas Utilizadas: Se utilizó la base de datos en el servidor de SQL Server 2000.

Entradas: Consulta en SQL

Salidas: Claves primarias de cada una de las tablas

Resultados Esperados: Como se esperaba, cada tabla tiene su clave primaria. Al ingresar valores duplicados o violar la característica de la clave primaria, siempre el sistema emite mensajes para que el usuario lo haga de forma correcta. El diseño de la base de datos, esta correctamente realizado y además ubicadas las claves primarias en las tablas que se necesita.

Porcentaje de Aprobación: 100%

Observaciones: Al ingresar un valor erróneo, el sistema le emite mensajes para que ingrese correctamente los datos.

PRUEBA DEL SISTEMA 5

Fecha de Realización: 16 de Marzo del 2006

Responsable: Carlos Yela

Eta de Prueba: Prueba del Sistema

Nombre del Proceso: Claves Foráneas

Tipo de Prueba: Pruebas de Bases de Datos

Técnicas y Herramientas Utilizadas: Se utilizó la base de datos en el servidor de SQL Server 2000.

Entradas: Consulta en SQL

Salidas: Claves foráneas de cada una de las tablas

Resultados Esperados: El diseño de la base de datos, esta correctamente realizado y además se identificó todas las tablas que tienen relación mediante las claves primarias.

Las tablas están correctamente relacionadas.

Porcentaje de Aprobación: 100%

Observaciones: Al tener claves foráneas, se fuerza a que los valores que se encuentren en las tablas principales se encuentren también en todas sus tablas relacionadas.

PRUEBA DEL SISTEMA 6

Fecha de Realización: 16 de Marzo del 2006

Responsable: Carlos Yela

Eta de Prueba: Prueba del Sistema

Nombre del Proceso: Carga con DTS

Tipo de Prueba: Pruebas de Bases de Datos

Técnicas y Herramientas Utilizadas: DTS generados para cargar datos al Data Mart

Entradas: DTS generados

Salidas: Datos hacia las tablas destino

Resultados Esperados: Que los datos se carguen en las tablas destino

Porcentaje de Aprobación: 100%

Observaciones: Todos los datos se cargaron a cada una de las tablas del Data Mart.

PRUEBA DEL SISTEMA 7

Fecha de Realización: 16 de Marzo del 2006

Responsable: Michael Vizueté

Eta de Prueba: Prueba del Sistema

Nombre del Proceso: Generación de Cubos

Tipo de Prueba: Pruebas de Bases de Datos

Técnicas y Herramientas Utilizadas: Análisis Services de SQL Server 2000

Entradas: Consultas empresariales

Salidas: Cubos generados

Resultados Esperados: Que los cubos se generen exitosamente

Porcentaje de Aprobación: 100%

Observaciones: Los cubos se generaron de manera exitosa permitiendo la transformación Multidimensional de los datos.

CAPITULO 3

3 CONCLUSIONES Y RECOMENDACIONES

3.1 CONCLUSIONES

- En la actualidad existen muchas y mejores herramientas tanto de libre difusión como registradas para utilizarlos en soluciones Data Mart.
- Los datos se filtran cuando pasan desde el ambiente operacional al de depósito Data Mart. Existen muchos datos que nunca salen del ambiente operacional. Sólo los datos que realmente se necesitan ingresarán al ambiente de Data Mart.
- El horizonte de tiempo de los datos es muy diferente de un ambiente al otro. La información en el ambiente operacional es más reciente con respecto a la del Data Mart. Desde la perspectiva de los horizontes de tiempo únicos, hay poca superposición entre los ambientes operacional y de Data Mart.
- El Data Mart contiene un resumen de la información que no se encuentra en el ambiente operacional.
- Los datos experimentan una transformación fundamental cuando pasa al Data Mart. La mayor parte de los datos se alteran significativamente al ser seleccionados y movidos al Data Mart. Dicho de otra manera, la mayoría de los datos se alteran física y radicalmente cuando se mueven al depósito. No son los mismos datos que residen en el ambiente operacional desde el punto de vista de integración.
- En vista de estos factores, la redundancia de datos entre los dos ambientes es una ocurrencia rara, que resulta en menos de 1% aproximadamente.
- La metodología de desarrollo de aplicaciones utilizada cubre todas las etapas, evitando pasar por alto detalles que causen una mala implementación de la solución Data Mart.

3.2 RECOMENDACIONES

- Se debe mejorar la disponibilidad de herramientas computacionales en los distintos Departamentos de la Escuela Politécnica Nacional para que los proyectos desarrollados internamente tengan facilidades en su ejecución.
- Se recomienda el uso de herramientas de libre difusión en los distintos proyectos que se realizan, ya que muchas de estas herramientas facilitan y mejoran la realización de dichos proyectos.
- Se recomienda recolectar la información de requerimientos directamente con los usuarios y propietarios de la solución Data Mart, ya que si se hace a través de intermediarios dicha información puede resultar errada y traer problemas en el desarrollo.

BIBLIOGRAFÍA

LIBROS

- KIMBALL Ralph, INMON W. H.; The Data Warehouse toolkit; Editorial John Wiley & Sons Inc.; 1era. Edición; 1996.
- INMON W. H.; Building the Data Warehouse; Editorial Wiley Computer Publishing; 2da. Edición; 1996.
- GILL Harjinder S., RAO Prakash C.; Data Warehouse la integración de la información para la mejor toma de decisiones; Editorial Prentice Hall Hispanoamericana; 1era. Edición; 1996.
- JARKEMAURIZIO Lenzerini, Mattian, VASSILIOU Yannis; Fundamentals of Data Warehouse; Editorial Springer, 1era. Edición; 2000.
- Microsoft Corporation; Microsoft SQL Server 7.0 Data Warehouse Training; Editorial Microsoft Press; 1era. Edición; 2000.
- STURM Jake; DataWarehouse with SQL Server 7.0 Technical Reference; Editorial Microsoft Press; 1era. Edición; 2000.
- ANAHORY Sam, MURRAY Dennis; Data Warehouse in the real world; Editorial Adisson Wesley; 1era. Edición; 1991.
- LEWIS William J.; Data Warehouse and E/commerce; Editorial Prentice Hall; 1era. Edición; 2001.

PAGINAS WEB

- GROS Jorge; DATAMART ¿Cuanto se ha de cavar?; 2001
<http://www.w2000mag.com/atrasados/1997/12sept97/Revista/DataMart.htm>

- Instituto Geofísico de la Escuela Politécnica Nacional; 2004
<http://www.igepn.edu.ec/sismologia/sismologia.htm>
- ESPINDOLA CASTRO Juan Manuel, JIMÉNEZ Zenón, Instituto de Geofísica de la UNAM; Segunda versión de terremotos y ondas sísmicas; 2da. Edición; 2002
<http://www.ssn.unam.mx>
- Instituto Andaluz de Geofísica y Prevención de Desastres Sísmicos; 2004.
<http://www.ugr.es/iag/iag.html>

ANEXO A

**SISTEMAS DE ADQUISICIÓN DE INFORMACIÓN
SÍSMICA**

SISTEMAS DE ADQUISICIÓN DE INFORMACIÓN SÍSMICA

ACQ

Logiciel d'acquisition de donnees sismiques

Autores: Julien Fréchet y Jean-Paul Glot con la colaboración de Francois Thouvenot

*Université Joseph Fourier et Centre National de la Recherche Scientifique

*Observatoire de Grenoble

*Laboratoire de Géophysique Interne et de Tectonophysique

ACQ3

Versión: V 2.70

Adquisición: 3 tarjetas DT 2814

Estaciones de la Red Nacional, Red Tungurahua, Red Cayambe, Red Reventador, Red Guagua Pichincha.

ACQ2

Versión: V 2.70

Adquisición: 2 tarjetas DT 2814

Estaciones de la Red Cotopaxi + otras estaciones útiles en la caracterización de la actividad sísmica.

Control y definición de parámetros: C:\acq2\acq2.ini

EARTHWORM

Estaciones: Todas

Control y definición de parámetros: C:\earthworm\run_ig\params

Para hacer cambios, hay que tener en cuenta, TODOS ESTOS ARCHIVOS:

1. adsend.d (canales)

2. vhseismicscns.d (canales)

3. waveserverprivate.d *definición de archivos "tanques"*

4. waveserverpublic.d *definición de archivos "tanques"*

5. ecuador.sta *definición de códigos*

Los archivos 1 y 2 deben ser idénticos, es decir la definición de las estaciones debe ser en el mismo orden.

Archivos de definición para los diferentes procesos:

a. Grabación en forma continua

- **Contrecord.ini** → Grabación en forma continua, archivos de 5 minutos.

b. Grabación por disparo

- **Carlstatrig.d** Definición de estaciones
- **Carlsbtrig.d** Definición de subredes de disparo

Subredes definidas:	Estaciones de disparo	Archivo creado:
Cotopaxi	3	*.cotop*
Tungurahua	2	*.tungu*
Guagua	2	*.guagu*
Cayambe	2	*.cayam*

Cuando se registra un disparo en más de una subred, se graba un archivo con todas las estaciones definidas en los 5 archivos de la sección anterior y toma el nombre de:

.regio

Grabación: d:\data\triggers\seisan

c. Otros procesos

- **Rsam.ini** → **Análisis de la amplitud de las señales sísmicas**

c:\earthworm\run_ig\rsam

Grabación: d:\data\rsam\10min

d:\data\rsam\1min

- **Ssam.ini** → **Análisis de la amplitud y frecuencia de las señales sísmicas**

c:\earthworm\run_ig\ssam

Grabación: d:\data\ssam\128channel

d:\data\ssam\16channel

d. Programas “accesorios”

sismogramas digitales → helicorders → heli.d

espectrogramas → spectrograms → sgram.d

c:\earthworm\run_ig\accessories

Grabación: f:\web\heli\final

f:\web\sgram\final

Sistema ACQ-SISMALP

- *FORMATO DE ARCHIVOS*

El ACQ genera archivos cuyo nombre es ddmhmi.ss (dia/mes/hora/minuto/segundo) y este listado es escrito en un archivo llamado events.eff y events.lst (bajo el directorio del programa de adquisición).

- *RECUPERACIÓN DE INFORMACIÓN*

Es posible el traspaso de información a la computadora de procesamiento, sin detener el programa de adquisición:

Menú→ Eventos/Vaciar

Se generan dos archivos *.flg y *.lst y se copian junto con los archivos ddmhmi.ss

- *TRANSFORMACIÓN DE FORMATO*

Archivos necesarios: acq2sis.exe y acq.lng + eventos (ddmhmi.ss) + *.flg + *.lst

Al correr el programa se generan 2 archivos por cada evento:

aammddab.ndx

aammddab.sis con ab → código

- *INCORPORACIÓN DE LA INFORMACIÓN Y VISUALIZACIÓN DE LA MISMA*

Transformación de formato y concatenación de datos:

ICONO: TRIEV→ triev.bat

Copia los archivos acq2sis.exe y acq.lng al directorio DUMPING

Corre el programa acq2sis.exe

Crea un directorio DUMP y mueve los datos convertidos a ese directorio

Realiza la concatenación de eventos (los organiza por año y mes)

Mensajes de error: Si el directorio DUMP no ha sido renombrado

Si los datos no son convertidos

Si hay un error en el archivo events.tbl (administrador de archivos)

- *PROGRAMA DE VISUALIZACIÓN*

ICONO: SISMALP→pickev.exe

Bajo DOS se pueden ver los eventos con el pick.exe, indicando toda la ruta del directorio donde están los archivos transformados.

- *ADMINISTRACIÓN DE INFORMACIÓN*

Definida en el archivo c:\sismalp\dirs.ini

EventsDir = c:\sismalp\events (p.e.)

- *ORGANIZACIÓN DE LA INFORMACIÓN*

Directorios organizados por año-mes/día y un archivo con el listado de eventos.

El archivo events.tbl tiene el listado de los eventos que fueron concatenados. Si este archivo está mal, no desplegará la información.

ANEXO B

CRONOGRAMA Y RECURSOS

CRONOGRAMA Y RECURSOS

CRONOGRAMA DEL DESARROLLO DEL DATA MART DE SISMOLOGIA DEL INSTITUTO GEOFISICO DE LA ESCUELA POLITECNICA NACIONAL				
	 Task Name	Duración	Comienzo	Fin
1	 DESARROLLO DEL DATA MART DE SISMOLOGIA	117 días	vie 07/10/05	lun 20/03/06
2				
3	<input checked="" type="checkbox"/> ANALISIS INICIAL	5 días	vie 07/10/05	jue 13/10/05
4	<input checked="" type="checkbox"/> Analisis de factibilidad	5 días	vie 07/10/05	jue 13/10/05
5				
6	<input checked="" type="checkbox"/> PLANIFICACION	26 días	vie 14/10/05	vie 18/11/05
7	<input checked="" type="checkbox"/> Seleccionar la estrategia de Implementación	3 días	vie 14/10/05	mar 18/10/05
8	<input checked="" type="checkbox"/> Seleccionar metodología de desarrollo	3 días	mié 19/10/05	vie 21/10/05
9	<input checked="" type="checkbox"/> Definir el ambito de implementacion	3 días	vie 21/10/05	mar 25/10/05
10	<input checked="" type="checkbox"/> Definir el arquitectura de implementacion	3 días	mié 26/10/05	vie 28/10/05
11	<input checked="" type="checkbox"/> Presupuesto y Conograma	7 días	sáb 29/10/05	mar 08/11/05
12	<input checked="" type="checkbox"/> Recopilación de Metadatos	8 días	mié 09/11/05	vie 18/11/05
13				
14	<input checked="" type="checkbox"/> DEFINICION DE REQUERIMIENTOS	28 días	sáb 19/11/05	mié 28/12/05
15	<input checked="" type="checkbox"/> Definir los requerimientos del propietario	7 días	sáb 19/11/05	mar 29/11/05
16	<input checked="" type="checkbox"/> Definir los requerimientos arquitectónicos	7 días	mié 30/11/05	jue 08/12/05
17	<input checked="" type="checkbox"/> Definir los requerimientos del desarrollador	7 días	vie 09/12/05	lun 19/12/05
18	<input checked="" type="checkbox"/> Definir los requerimientos del usuario final	7 días	mar 20/12/05	mié 28/12/05
19				
20	<input checked="" type="checkbox"/> ANALISIS	24 días	lun 02/01/06	jue 02/02/06
21	<input checked="" type="checkbox"/> Analisis de los factores que conducen el negocio	1,15 días	lun 02/01/06	mar 03/01/06
22	<input checked="" type="checkbox"/> Analizar objetivos del negocio	2 días	mié 04/01/06	jue 05/01/06
23	<input checked="" type="checkbox"/> Analizar la información de alto nivel	4 días	vie 06/01/06	mié 11/01/06
24	<input checked="" type="checkbox"/> Definir Procesos y roles	4 días	jue 12/01/06	mar 17/01/06
25	<input checked="" type="checkbox"/> Definir indicadores claves de rendimiento	5 días	mié 18/01/06	mar 24/01/06
26	<input checked="" type="checkbox"/> Analizar fuentes de datos	7 días	mié 25/01/06	jue 02/02/06
27				
28	<input checked="" type="checkbox"/> DISEÑO	11 días	vie 03/02/06	vie 17/02/06
29	<input checked="" type="checkbox"/> Diseño de la Arquitectura de Datos	7 días	vie 03/02/06	lun 13/02/06
30	<input checked="" type="checkbox"/> Diseño de la Arquitectura de la Aplicación	4 días	mar 14/02/06	vie 17/02/06
31				
32	<input checked="" type="checkbox"/> IMPLEMENTACION	19 días	lun 20/02/06	jue 16/03/06
33	<input checked="" type="checkbox"/> Creación del Data Mart	3 días	lun 20/02/06	mié 22/02/06
34	<input checked="" type="checkbox"/> Creación de las BDD	3 días	mié 01/03/06	vie 03/03/06
35	<input checked="" type="checkbox"/> Extracción de datos	3 días	lun 06/03/06	mié 08/03/06
36	<input checked="" type="checkbox"/> Depuración de datos	2 días	jue 09/03/06	vie 10/03/06
37	<input checked="" type="checkbox"/> Carga y Transformación de datos en el Data Mart	2 días	lun 13/03/06	mar 14/03/06
38	<input checked="" type="checkbox"/> Generación de cubos	2 días	mié 15/03/06	jue 16/03/06
39				
40	<input checked="" type="checkbox"/> PRUEBAS	3 días	jue 16/03/06	lun 20/03/06
REALIZADO POR: MICHAEL W. VIZUETE N. Y CARLOS P. YELA S.				

CRONOGRAMA DEL DESARROLLO DEL DATA MART DE SISMOLOGIA DEL INSTITUTO GEOFISICO DE LA ESCUELA POLITECNICA NACIONAL

	Predecesoras	Nombres de los recursos
1		
2		
3		
4		INGENIERO ESPECIALIZADO EN BASES DE DATOS;EXPERTO EN DATAWAREHOUSE;EXPERTO EN
5		
6	3	
7	4	EXPERTO EN DATAWAREHOUSE;INGENIERO ESPECIALIZADO EN BASES DE DATOS
8	7	EXPERTO EN DATAWAREHOUSE;INGENIERO ESPECIALIZADO EN BASES DE DATOS
9	8	EXPERTO EN DATAWAREHOUSE;INGENIERO ESPECIALIZADO EN BASES DE DATOS
10	9	EXPERTO EN DATAWAREHOUSE;INGENIERO ESPECIALIZADO EN BASES DE DATOS
11	10	EXPERTO EN DATAWAREHOUSE;INGENIERO ESPECIALIZADO EN BASES DE DATOS;SERVIDOR D
12	11	EXPERTO EN SISMOLOGIA[20%];EXPERTO EN DATAWAREHOUSE[20%];INGENIERO ESPECIALIZA
13		
14	6	
15	12	INGENIERO ESPECIALIZADO EN BASES DE DATOS
16	15	INGENIERO ESPECIALIZADO EN BASES DE DATOS
17	16	INGENIERO ESPECIALIZADO EN BASES DE DATOS;EXPERTO EN DATAWAREHOUSE
18	17	INGENIERO ESPECIALIZADO EN BASES DE DATOS;ADMINISTRADOR DE BASES DE DATOS;DISEÑ
19		
20	14	
21	18	EXPERTO EN DATAWAREHOUSE;EXPERTO EN SISMOLOGIA;INGENIERO ESPECIALIZADO EN BAS
22	21	EXPERTO EN SISMOLOGIA;INGENIERO ESPECIALIZADO EN BASES DE DATOS
23	22	EXPERTO EN SISMOLOGIA;EXPERTO EN DATAWAREHOUSE
24	23	INGENIERO ESPECIALIZADO EN BASES DE DATOS
25	24	EXPERTO EN SISMOLOGIA;EXPERTO EN DATAWAREHOUSE;INGENIERO ESPECIALIZADO EN BAS
26	25	INGENIERO ESPECIALIZADO EN BASES DE DATOS;ADMINISTRADOR DE BASES DE DATOS
27		
28	20	
29	26	EXPERTO EN DATAWAREHOUSE[50%];INGENIERO ESPECIALIZADO EN BASES DE DATOS;ADMIN
30	29	HERRAMIENTAS CASE[1];ADMINISTRADOR DE BASES DE DATOS;INGENIERO ESPECIALIZADO EN
31		
32	28	
33	30	SISTEMA OPERATIVO MICROSOFT WINDOWS XP[1];SERVIDOR DE BASE DE DATOS[1];PROGRAM
34	33	SISTEMA OPERATIVO MICROSOFT WINDOWS XP[1];SERVIDOR DE BASE DE DATOS[1];MICROSOFT
35	34	SERVIDOR DE BASE DE DATOS[1];SISTEMA OPERATIVO MICROSOFT WINDOWS XP[1];MICROSOFT
36	35	SERVIDOR DE BASE DE DATOS[1];SISTEMA OPERATIVO MICROSOFT WINDOWS XP[1];MICROSOFT
37	36	SERVIDOR DE BASE DE DATOS[1];SISTEMA OPERATIVO MICROSOFT WINDOWS XP[1];MICROSOFT
38	37	SERVIDOR DE BASE DE DATOS[1];SISTEMA OPERATIVO MICROSOFT WINDOWS XP[1];MICROSOFT
39		
40	28	

REALIZADO POR: MICHAEL W. VIZUETE N. Y CARLOS P. YELA S.

ANEXO C

MODELO ENTIDAD – RELACION DE LA BASE DE DATOS INTERMEDIA

DIAGRAMA ENTIDAD-RELACION

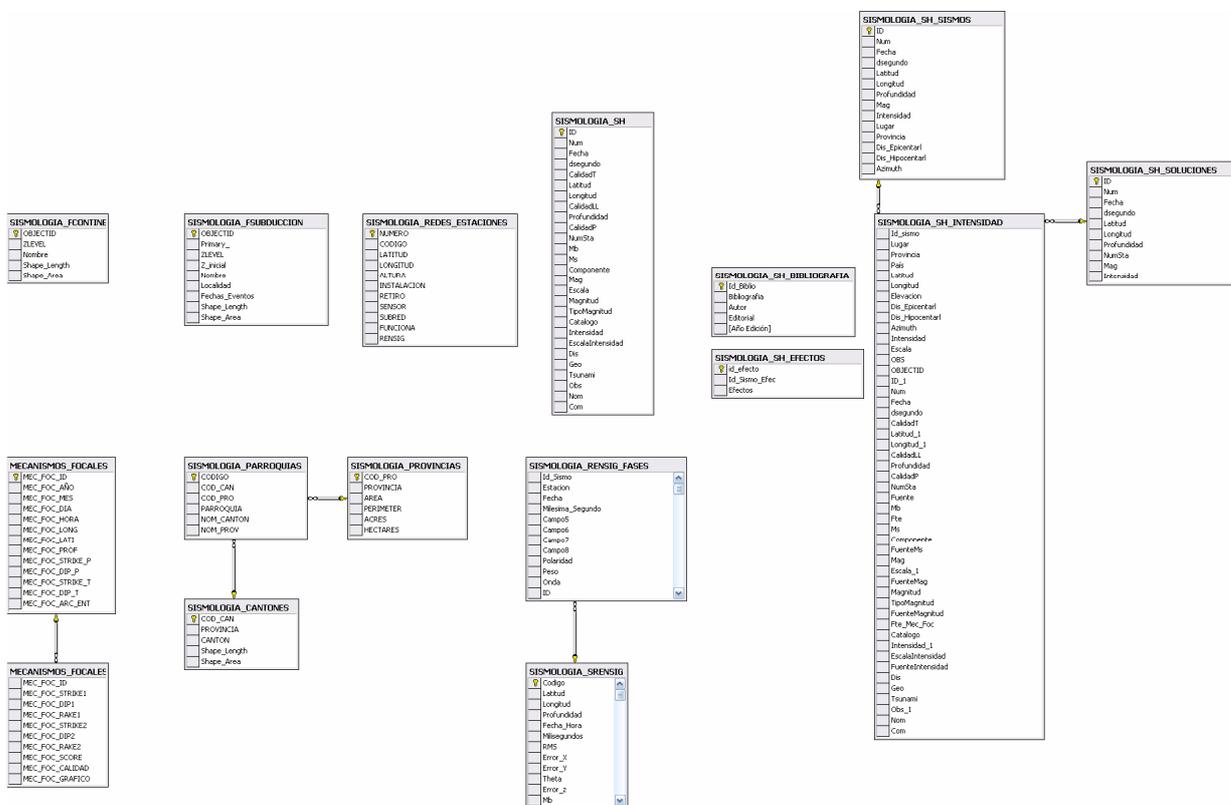


Figura 88 Diagrama Entidad Relación 41

41 Fuente: IGEPN

ANEXO D

MODELO DE DATOS DEL DATA MART DE SISMOLOGIA

MODELO DE DATOS

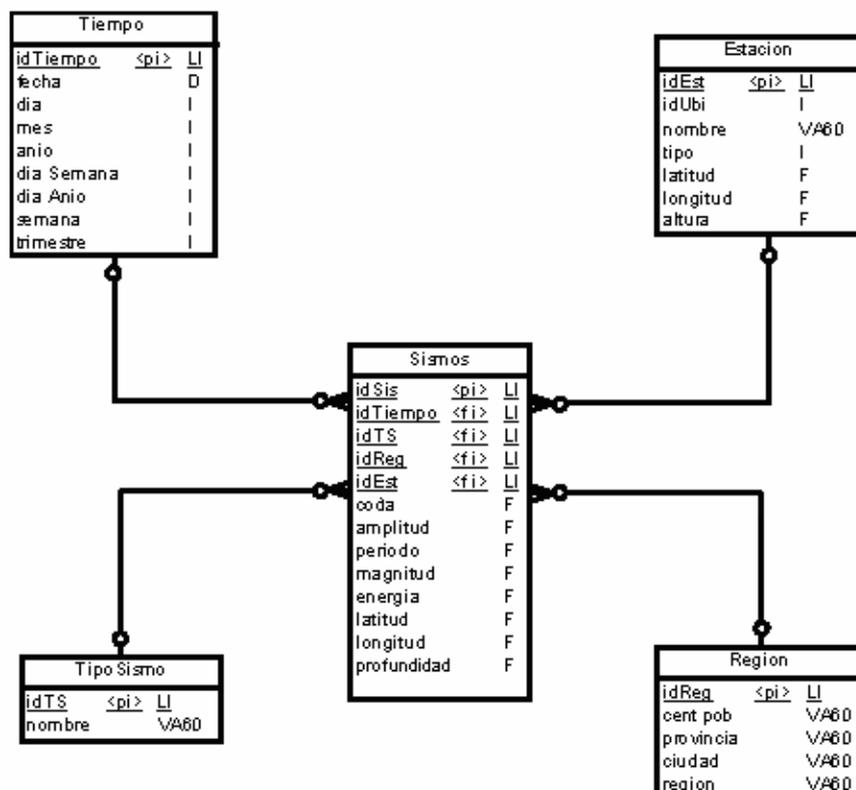


Figura 89 Diagrama Estrella ⁴²

DICCIONARIO DE DATOS DEL MODELO DE DATOS DEL DATA MART DE SISMOLOGIA

ENTIDAD	ATRIBUTO	DESCRIPCION	TIPO DE DATO	TAMAÑO	REQUERIDO	VALOR PREDETERMINADO
Tiempo	idTiempo	Identificador de tiempo	int	4	SI	Autonumerico
	fecha	Fecha	datetime	8	NO	
	dia	Día	nvarchar	20	NO	
	mes	Mes	nvarchar	20	NO	
	anio	Año	Int	4	NO	
	dia Semana	Día de la Semana	int	4	NO	
	dia Anio	Día del Año	Int	4	NO	
	semana	Semana	Int	4	NO	
	trimestre	Trimestre	int	4	NO	

⁴² Fuente: Autores

Tipo Sismo	idTS	Identificador del Tipo de Sismo	int	4	SI	Autonumerico
	nombre	Nombre del tipo de sismo	nvarchar	20	NO	
Region	idReg	Identificador de la Región	int	4	SI	Autonumerico
	cent pob	Centro poblado	nvarchar	60	NO	
	provincia	Provincia donde está ubicado	nvarchar	60	NO	
	ciudad	Ciudad donde está ubicado	nvarchar	60	NO	
	region	Región donde está ubicado	nvarchar	60	NO	
Estacion	idEst	Identificador de la Estación	int	4	SI	Autonumerico
	idUbi	Identificador de la Ubicación Geográfica	int	4	SI	Autonumerico
	nombre	Nombre de la estación	nvarchar	60	NO	
	tipo	Tipo de estación	nvarchar	60	NO	
	latitud	Latitud donde está ubicada	float	8	NO	
	longitud	Longitud donde está ubicada	float	8	NO	
	altura	Altura donde está ubicada	float	8	NO	
Sismos	idSis	Identificador del Sismo	int	4	SI	Autonumerico
Sismos	idReg	Identificador del Región (Clave foránea)	int	4	SI	Autonumerico
	idTiempo	Identificador de Tiempo (Clave foránea)	int	4	SI	Autonumerico
	idTS	Identificador del Tipo de Sismo (Clave foránea)	int	4	SI	Autonumerico
	idEst	Identificador de la Estación (Clave foránea)	int	4	SI	Autonumerico
	coda	Duración del evento sísmico en segundos	float	8	NO	
	amplitud	Amplitud del sismo	float	8	NO	
	periodo	Periodo del sismo	float	8	NO	
	magnitud	Magnitud del sismo	float	8	NO	
	energia	Energía liberada por el sismo	float	8	NO	
	latitud	Latitud donde ocurrió el sismo	float	8	NO	
	longitud	Longitud donde ocurrió el sismo	float	8	NO	
	Profundidad	Profundidad a la que ocurrió el sismo	float	8	NO	

ANEXO E

DESCRIPCION DE PROCESOS DE LA ARQUITECTURA DE LA APLICACION

DESCRIPCION DE PROCESOS DE LA ARQUITECTURA DE LA APLICACION

ARQUITECTURA DE LA APLICACION

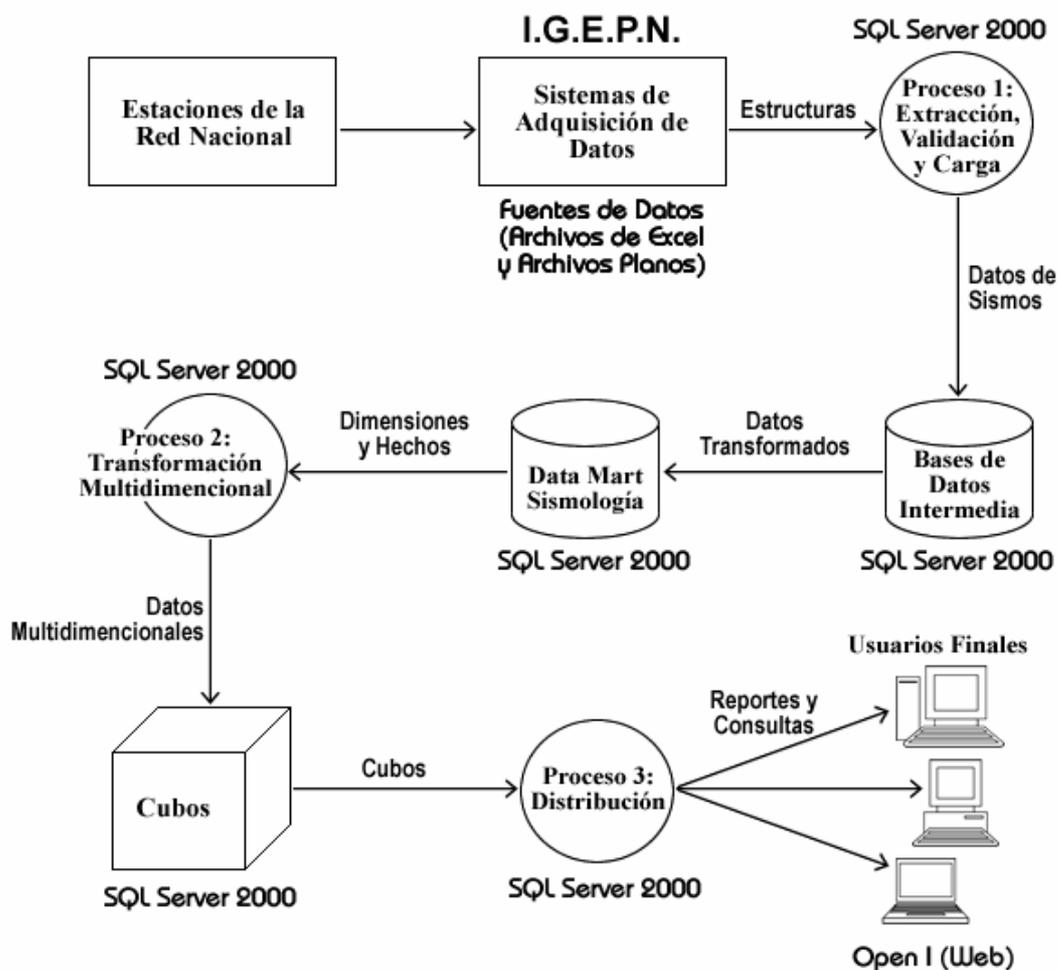


Figura 90 Arquitectura de la Aplicación ⁴³

⁴³ Fuente: Autores

PROCESO 1

NOMBRE DEL PROCESO	
Nombre:	Extracción, Validación y Carga
Código:	P1
Número:	1
DESCRIPCION DEL PROCESO	
Proceso en el cual se extrae la información desde las fuentes se valida y se carga hacia la base de datos intermedia que nos servirá como fuente para el Data Mart de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional.	
LISTA DE REFERENCIAS DEL PROCESO	
CONECTADO VIA	CONECTADO A
Estructuras	Fuentes de Datos
Datos de Sismos	Base de Datos Intermedia

PROCESO 2

NOMBRE DEL PROCESO	
Nombre:	Transformación Multidimensional
Código:	P2
Número:	2
DESCRIPCION DEL PROCESO	
Proceso en el cual se toma los datos desde las tablas de hechos y dimensiones y los procesamos transformándolos en multidimensionales para ofrecer a los usuarios del Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional mejor información.	
LISTA DE REFERENCIAS DEL PROCESO	
CONECTADO VIA	CONECTADO A
Dimensiones y Hechos	Data Mart de Sismología
Datos Multidimensionales	Cubos

PROCESO 3

NOMBRE DEL PROCESO	
Nombre:	Distribución
Código:	P3
Número:	3
DESCRIPCION DEL PROCESO	
Proceso en el cual los datos multidimensionales se distribuyen a cada uno de los usuarios del Área de Sismología del Instituto Geofísico de la Escuela Politécnica Nacional para que los puedan visualizar a través de herramientas de usuario final.	
LISTA DE REFERENCIAS DEL PROCESO	
CONECTADO VIA	CONECTADO A
Cubos	Cubos
Reportes y Consultas	Usuarios Finales

ANEXO F

SCRIPT DE CREACION DEL DATA MART DE SISMOLOGIA

SCRITP DE CREACION DEL DATA MART

```
IF EXISTS (SELECT name FROM master.dbo.sysdatabases WHERE name = N'SismologiaDM')
    DROP DATABASE [SismologiaDM]
```

```
GO
```

```
CREATE DATABASE [SismologiaDM] ON (NAME = N'SismologiaDM_Data', FILENAME =
N'C:\Archivos de programa\Microsoft SQL Server\MSSQL\Data\SismologiaDM_Data.MDF' , SIZE =
6, FILEGROWTH = 10%) LOG ON (NAME = N'SismologiaDM_Log', FILENAME = N'C:\Archivos
de programa\Microsoft SQL Server\MSSQL\Data\SismologiaDM_Log.LDF' , SIZE = 1,
FILEGROWTH = 10%)
```

```
COLLATE Modern_Spanish_CI_AS
```

```
GO
```

```
use [SismologiaDM]
```

```
GO
```

```
if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[FK_D_ESTACI_RELATIONS_D_ESTACI]') and OBJECTPROPERTY(id,
N'IsForeignKey') = 1)
```

```
ALTER TABLE [dbo].[Estacion] DROP CONSTRAINT FK_D_ESTACI_RELATIONS_D_ESTACI
```

```
GO
```

```
if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[FK_Sismos_RELATIONS_D_ESTACI]') and OBJECTPROPERTY(id,
N'IsForeignKey') = 1)
```

```
ALTER TABLE [dbo].[Sismos] DROP CONSTRAINT FK_Sismos_RELATIONS_D_ESTACI
```

```
GO
```

```
if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[FK_D_REGION_RELATIONS_D_REGION]') and OBJECTPROPERTY(id,
N'IsForeignKey') = 1)
```

```
ALTER TABLE [dbo].[D_REGION] DROP CONSTRAINT
FK_D_REGION_RELATIONS_D_REGION
```

```
GO
```

```

if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[FK_Sismos_RELATIONS_D_REGION]') and OBJECTPROPERTY(id,
N'IsForeignKey') = 1)
ALTER TABLE [dbo].[Sismos] DROP CONSTRAINT FK_Sismos_RELATIONS_D_REGION
GO

```

```

if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[FK_Sismos_RELATIONS_D_TIEMPO]') and OBJECTPROPERTY(id,
N'IsForeignKey') = 1)
ALTER TABLE [dbo].[Sismos] DROP CONSTRAINT FK_Sismos_RELATIONS_D_TIEMPO
GO

```

```

if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[FK_D_TIPO_S_RELATIONS_D_TIPO_S]') and OBJECTPROPERTY(id,
N'IsForeignKey') = 1)
ALTER TABLE [dbo].[TipoSismo] DROP CONSTRAINT FK_D_TIPO_S_RELATIONS_D_TIPO_S
GO

```

```

if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[FK_Sismos_RELATIONS_D_TIPO_S]') and OBJECTPROPERTY(id,
N'IsForeignKey') = 1)
ALTER TABLE [dbo].[Sismos] DROP CONSTRAINT FK_Sismos_RELATIONS_D_TIPO_S
GO

```

```

if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[Tiempo]') and
OBJECTPROPERTY(id, N'IsProcedure') = 1)
drop procedure [dbo].[Tiempo]
GO

```

```

if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[Estacion]') and
OBJECTPROPERTY(id, N'IsUserTable') = 1)
drop table [dbo].[Estacion]
GO

```

```

if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[D_REGION]') and
OBJECTPROPERTY(id, N'IsUserTable') = 1)
drop table [dbo].[Region]

```

GO

```
if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[D_TIEMPO]') and
OBJECTPROPERTY(id, N'IsUserTable') = 1)
```

```
drop table [dbo].[Tiempo]
```

GO

```
if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[TipoSismo]') and
OBJECTPROPERTY(id, N'IsUserTable') = 1)
```

```
drop table [dbo].[TipoSismo]
```

GO

```
if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[Sismos]') and
OBJECTPROPERTY(id, N'IsUserTable') = 1)
```

```
drop table [dbo].[Sismos]
```

GO

```
CREATE TABLE [dbo].[Estacion] (
    [IDEST] [int] IDENTITY (1, 1) NOT NULL ,
    [D_E_IDEST] [int] NULL ,
    [IDUBI] [int] NULL ,
    [NOMBRE] [varchar] (60) COLLATE Modern_Spanish_CI_AS NULL ,
    [TIPOEST] [int] NULL ,
    [LATITUD] [float] NULL ,
    [LONGITUD] [float] NULL ,
    [ALTURA] [float] NULL ,
    [id] [varchar] (255) COLLATE Modern_Spanish_CI_AS NULL
) ON [PRIMARY]
```

GO

```
CREATE TABLE [dbo].[Region] (
    [IDREG] [int] NOT NULL ,
    [D_R_IDREG] [int] NULL ,
    [NOMPAIS] [varchar] (60) COLLATE Modern_Spanish_CI_AS NULL ,
    [NOMPROV] [varchar] (60) COLLATE Modern_Spanish_CI_AS NULL ,
    [NOMREGION] [varchar] (60) COLLATE Modern_Spanish_CI_AS NULL ,
```

```

        [NOMCIUDAD] [varchar] (60) COLLATE Modern_Spanish_CI_AS NULL
    ) ON [PRIMARY]
GO

```

```

CREATE TABLE [dbo].[Tiempo] (
    [IDIEMPO] [int] IDENTITY (1, 1) NOT NULL ,
    [FECHA] [datetime] NULL ,
    [DIA] [int] NULL ,
    [MES] [int] NULL ,
    [ANIO] [int] NULL ,
    [DIA_SEMANA] [int] NULL ,
    [DIA_ANIO] [int] NULL ,
    [SEMANA] [int] NULL ,
    [TRIMESTRE] [int] NULL
) ON [PRIMARY]
GO

```

```

CREATE TABLE [dbo].[TipoSismo] (
    [IDTS] [varchar] (10) COLLATE Modern_Spanish_CI_AS NOT NULL ,
    [D_T_IDTS] [varchar] (10) COLLATE Modern_Spanish_CI_AS NULL ,
    [NOMBRETS] [varchar] (60) COLLATE Modern_Spanish_CI_AS NULL
) ON [PRIMARY]
GO

```

```

CREATE TABLE [dbo].[Sismos] (
    [IDSIS] [int] IDENTITY (1, 1) NOT NULL ,
    [IDEST] [int] NULL ,
    [IDREG] [int] NULL ,
    [IDTS] [varchar] (10) COLLATE Modern_Spanish_CI_AS NULL ,
    [IDIEMPO] [int] NULL ,
    [CODASIS] [float] NULL ,
    [AMPSIS] [float] NULL ,
    [PERSIS] [float] NULL ,
    [MAGSIS] [float] NULL ,
    [MAG1SIS] [float] NULL ,
    [MAG2SIS] [float] NULL ,

```

```
[ENERSIS] [float] NULL ,
[ENER1SIS] [float] NULL ,
[ENER2SIS] [float] NULL
) ON [PRIMARY]
GO

ALTER TABLE [dbo].[Estacion] ADD
    CONSTRAINT [PK_Estacion] PRIMARY KEY CLUSTERED
    (
        [IDEST]
    ) ON [PRIMARY]
GO

ALTER TABLE [dbo].[Region] ADD
    CONSTRAINT [PK_Region] PRIMARY KEY CLUSTERED
    (
        [IDREG]
    ) ON [PRIMARY]
GO

ALTER TABLE [dbo].[Tiempo] ADD
    CONSTRAINT [PK_Tiempo] PRIMARY KEY CLUSTERED
    (
        [IDTIEMPO]
    ) ON [PRIMARY]
GO

ALTER TABLE [dbo].[TipoSismo] ADD
    CONSTRAINT [PK_TipoSismo] PRIMARY KEY CLUSTERED
    (
        [IDTS]
    ) ON [PRIMARY]
GO

ALTER TABLE [dbo].[Sismos] ADD
    CONSTRAINT [PK_Sismos] PRIMARY KEY CLUSTERED
```

```
(
    [IDSIS]
) ON [PRIMARY]
GO

ALTER TABLE [dbo].[Estacion] ADD
    CONSTRAINT [FK_D_ESTACI_RELATIONS_D_ESTACI] FOREIGN KEY
    (
        [D_E_IDEST]
    ) REFERENCES [dbo].[Estacion] (
        [IDEST]
    )
GO

ALTER TABLE [dbo].[Region] ADD
    CONSTRAINT [FK_Region_RELATIONS_Region] FOREIGN KEY
    (
        [D_R_IDREG]
    ) REFERENCES [dbo].[Region] (
        [IDREG]
    )
GO

ALTER TABLE [dbo].[TipoSismo] ADD
    CONSTRAINT [FK_D_TIPO_S_RELATIONS_D_TIPO_S] FOREIGN KEY
    (
        [D_T_IDTS]
    ) REFERENCES [dbo].[TipoSismo] (
        [IDTS]
    )
GO

ALTER TABLE [dbo].[Sismos] ADD
    CONSTRAINT [FK_Sismos_RELATIONS_D_ESTACI] FOREIGN KEY
    (
        [IDEST]
```

```
) REFERENCES [dbo].[Estacion] (  
    [IDEST]  
)  
,  
CONSTRAINT [FK_Sismos_RELATIONS_Region] FOREIGN KEY  
(  
    [IDREG]  
) REFERENCES [dbo].[Region] (  
    [IDREG]  
)  
,  
CONSTRAINT [FK_Sismos_RELATIONS_Tiempo] FOREIGN KEY  
(  
    [IDTIEMPO]  
) REFERENCES [dbo].[Tiempo] (  
    [IDTIEMPO]  
)  
,  
CONSTRAINT [FK_Sismos_RELATIONS_D_TIPO_S] FOREIGN KEY  
(  
    [IDTS]  
) REFERENCES [dbo].[TipoSismo] (  
    [IDTS]  
)  
)
```

GO