



UNIVERSIDAD NACIONAL DE LA MATANZA

SECRETARIA DE POSGRADO

MAESTRIA EN INFORMÁTICA

Tesis

**“Data warehouse e inteligencia de negocios aplicados a
organismos del Estado”**

Autor: Leonardo Moffa

Director: Mag. Cristóbal Santa María

Codirector: Esp. Hugo Castro

Buenos Aires, *Noviembre, 2015*

Índice

RESUMEN.....	5
1-INTRODUCCIÓN.....	6
1.1. – Selección del Tema y Definición del Problema.....	6
1.2. – Antecedentes.....	7
1.2.1 – Marco Teórico.....	8
1.2.1.1 – La Información en la Organización.....	9
1.2.1.2 – Toma de Decisiones.....	11
1.2.1.2.1 - Proceso de Toma de decisiones.....	11
1.2.1.3 – Sistemas de Soporte a las Decisiones.....	15
1.2.1.4 – Inteligencia de Negocios.....	17
1.2.1.5 – Sistemas de Información Ejecutiva.....	20
1.2.1.6 – Data Warehouse.....	22
1.2.1.6.1 - Características de un DW.....	26
1.2.1.6.2 - Construcción e Implementación de un Data Warehouse.....	27
1.2.1.6.3 - Personal involucrado.....	31
1.2.1.6.4 - Madurez de un Data Warehouse en la Organización.....	32
1.2.1.6.5 - Proceso ETL.....	34
1.2.1.6.6 - Ventajas del uso de un Data Warehouse.....	35

1.2.1.7 – Data Marts	36
1.2.1.7.1 - Construcción de un Data Mart	37
1.2.1.7.2 - Requerimientos para la Construcción de un Data Mart	39
1.2.1.7.3 - Data Mart Distribuido	40
1.2.1.8 – Modelo Dimensional	41
1.2.1.8.1 - Modelo de Estrella	44
1.2.1.8.2 - Modelo de Copo de Nieve	45
1.2.1.9 – Proceso OLAP	46
1.2.1.10 – Data Mining.....	52
1.3. – Justificación del Estudio	56
1.4. – Limitaciones de la Investigación	56
1.5. – Alcance de la Investigación	56
1.6. – Objetivos de la Investigación	57
1.6.1 -- Objetivo General	57
1.6.2 -- Objetivos Específicos	58
1.7. – Hipótesis	58
2-MATERIAL Y MÉTODOS.....	59
2.1. – Material.....	60
2.1.1 – Lugar y Tiempo de Trabajo de la Investigación	60
2.1.2 – Descripción del Objeto de Estudio	60

2.1.3 – Descripción Población y Muestra	61
2.2. – Métodos	62
2.2.1 –Diseño de la Investigación	62
2.2.1.1 - Introducción	62
2.2.1.2 – Parámetros del Estudio.....	63
2.2.1.2.1 - La delimitación de la actividad del Estado	64
2.2.1.2.2 - Estructura del Sisio Red Federal y Entes.....	65
2.2.1.2.3 - El ámbito geográfico de las auditorías.....	70
2.2.1.2.4 - El intervalo temporal de captación de datos	71
2.2.1.2.5 – Muestra de Organismos.....	73
2.2.1.2.5 – Muestra de Programas.....	76
2.2.1.2.6 – Las auditorías observadas, tipos de auditorías, estados de auditorías y la cantidad de horas observadas	78
2.2.1.3 – Diseño de la solución	81
2.2.1.3.1 – Planeamiento del problema	82
2.2.1.3.2 – Objetivos del proyecto	83
2.2.1.3.3 – Requerimientos de la solución	84
2.2.1.3.4 – Descripción funcional de la solución	88
2.2.1.3.5 – Modelo Dimensional.....	90
2.2.3.6 - Descripción de los procesos desarrollados	93

2.2.2 - Instrumentos de recolección y medición de datos	97
2.2.3 - Confiabilidad y validez de la medición	99
2.2.3.1 – Pruebas de Hipótesis	99
2.2.3.2 - Intervalos de Confianza	100
2.2.4 - Métodos de análisis estadístico.....	101
2.2.4.1 - Distribuciones de frecuencias	102
3-RESULTADOS.....	106
3.1. – Modelo Resultante.....	106
3.2. – Cubo Resultante.....	108
4-DISCUSIÓN	128
5-CONCLUSIONES	134
LISTA DE ILUSTRACIONES	136
LISTA DE TABLAS.....	138
BIBLIOGRAFÍA.....	140

RESUMEN

En el presente trabajo de tesis de magister se realizó una investigación para el apoyo a la toma de decisiones del Estado. El objetivo principal es poner en evidencia que el Estado puede sacar provecho de la Inteligencia de Negocios para la toma de decisiones sobre cuestiones actuales que afectan a la sociedad. Para ello se proponen técnicas para la obtención de objetivos claros y para poder definir indicadores claves de desempeño útiles para la toma de decisiones de los organismos estatales. Además se determinan los principales elementos de los requerimientos que el Estado busca en una modelo dimensional de Inteligencia de Negocios.

Para el desarrollo de esta propuesta de investigación, se desarrolla un caso práctico con datos de entrada perteneciente a los generados por los sistemas y registros de la Sindicatura General de la Nación.

1-INTRODUCCIÓN

1.1. – Selección del Tema y Definición del Problema

El data warehouse es un depósito donde se almacenan los datos que la organización utiliza para saber cómo está funcionando. Es un componente del modelo de Inteligencia de Negocios, en cuya base se encuentra el almacén, la infraestructura conformada por la base de datos donde se integra toda la información; sobre él están las herramientas para presentar informes y hacer consultas; más arriba esta la estrategia de inteligencia de negocios formada metodologías que se ciñen a las mejores prácticas, la cultura organizacional y las habilidades de desarrollo con las que se cuenta para sacar mayor provecho de la infraestructura y de las aplicaciones.

La Inteligencia de Negocios consiste en convertir datos en información útil, en utilizar datos para conocer el desempeño de una organización. El vínculo entre la inteligencia de Negocios y la tecnología es fundamental, porque es la zona donde se traducen los datos en información útil para la organización. Es un puente que transforma, depura e integra datos.

El Estado cuenta cada vez con más información en bases de datos almacenada por sistemas computarizados. Para ello es necesario contar con una Inteligencia de Negocios que permita alcanzar los objetivos propuestos por la organización facilitando la consolidación y el análisis del gran volumen de datos que llegan a manejar.

La diferencia más importante entre los objetivos de empresas con fines financieros y en organismos del Estado es que para los segundos no es únicamente un balance de resultados financieros, sino que su medición tiene indicadores tales como: más y mejor educación, más y mejores empleos, o reducción de la pobreza, por citar algunos ejemplos. De ahí que los datos que

sirven para medir el desempeño del Estado se encuentra en la satisfacción del ciudadano y en una mejor calidad de vida para todos.

Por ello la obtención de datos para una Inteligencia de Negocios en el Estado es mucho más compleja que en una empresa ya que las fuentes son mucho más diversas. El Estado es la suma de una gran variedad de dependencias, organismos, empresas estatales, organizaciones sociales y ciudadanas, entre otras, que trabajan en distintos ámbitos territoriales, y es necesario poder encapsular todos esos datos dispersos para obtener una eficaz medición.

1.2. – Antecedentes

Las expectativas de los ciudadanos sobre cómo el gobierno debería aumentar la eficiencia y la eficacia de la gestión pública, haciéndola más efectiva están evolucionando con rapidez.

La velocidad y alcance de las TIC hacen posible que el gobierno incremente su capacidad de comprometerse con los ciudadanos y con otros actores que hacen a la gestión de gobierno. Ha utilizado la tecnología de información por muchos años, algunos organismos de manera muy efectiva. Muchos utilizan herramientas digitales para administrar sus operaciones básicas, como presupuestos o rentas; pero muy pocos utilizan las mismas para administrar el flujo de información dentro del gobierno como un todo, especialmente entre diferentes niveles y jurisdicciones o en la entrega directa de servicios a los ciudadanos. El rápido crecimiento del flujo de datos demanda un preciso gestionamiento de manera que ayude a los ejecutivos de un organismo a tomar decisiones eficaces, con base en la información almacenada.

La Inteligencia de Negocios se basa en la integración de datos históricos y actuales para determinar tendencias futuras y proponer escenarios posibles. Y en este sentido, es lógico que las

decisiones deben estar basadas en datos precisos y consistentes. Por ello existe el data warehouse como solución para sacar mayor provecho a los datos de las aplicaciones.

Hoy en día existen algunos estudios donde intentan poner en evidencia que el Estado puede sacar provecho de la Inteligencia de Negocios para la toma de decisiones siempre y cuando se cuente con objetivos claros y se disponga de datos para llegar a cumplir los mismos.

Algunos de estos estudios pueden ser observados en las reseñas referidas.

- * Revista Política digital una publicación de NEXOS: Editorial NEXOS. Ejemplar 4, Pág. 79
- * Gobierno de la provincia de Neuquén. (2003) Plan Maestro de Gobierno Electrónico.

1.2.1 – Marco Teórico

Como marco teórico se realiza una introducción a la información de las organizaciones (privadas y públicas) y su importancia para la toma de decisiones.

A su vez se describen los conceptos como La Información de la Organización, Toma de Decisiones, Sistemas de Soporte a las Decisiones, Inteligencia de Negocios, Sistemas de Información Ejecutiva, Data Warehouse, Data Marts, Modelo Dimensional, OLAP y Data Mining. Por medio de este marco teórico podremos adquirir un mejor entendimiento del objetivo del proyecto, se demostrará con qué herramientas conceptuales se cuenta, cómo se organizan e integran y cómo se utilizan estas herramientas para brindar información para la toma de decisiones.

1.2.1.1 – La Información en la Organización

Para una organización contar con información oportuna y precisa puede representar la diferencia entre ganancias o pérdidas. En un organismo público puede representar cumplir en tiempo y forma con los objetivos y metas planteadas utilizando un óptimo aprovechamiento de los recursos disponibles. Con dicha información, la alta gerencia podría conocer mejor sus problemáticas sociales, determinar mejoras, saber en qué momento debe atacar dichos aspectos y muchos más parámetros estratégicos para tomar decisiones.

Los datos por si solos no dicen nada, pero dependiendo del contexto en donde se encuentren de la correcta integración de los mismos, los datos van a permitir al ejecutivo tomar decisiones más rápidas y eficaces.

Desde que las organizaciones comenzaron a guardar los datos de sus operaciones en medios de almacenamiento físico, con el fin de permitirles una mejor administración y control de la información, ha existido una necesidad de utilizarla para atender las necesidades propias del negocio.

La información y su importancia estratégica comenzó a surgir cuando en el ámbito comercial la competencia se hizo muy fuerte, y cada vez más y más productos similares, de diferentes compañías, se ponían a la venta, en ese momento el consumidor tuvo la opción de seleccionar aquello que más le conviniera o lo que más se adecuara a sus gustos y preferencias. Surge entonces la necesidad de brindar servicios adicionales para obtener la lealtad de los clientes, quienes poco a poco comenzaron a ver, no solo el producto que compraban, sino cómo eran atendidos, qué garantías se ofrecían sobre su compra, qué ventajas habría entre diferentes productos y, en general, evaluar todo lo que genera la diferenciación entre las compras que realizan. Cuando las empresas no tienen garantizada la venta de lo que producen, realizan un cambio paulatino hacia obtener de los datos toda la información útil y estratégica para mantenerse en el mercado, dándole un lugar preponderante al cliente.

Actualmente, se le da un peso específico muy importante a la información como el principal conocimiento que sostiene el negocio. Existen empresas que, de modo predominante, ofrecen servicios y giran su negocio principal sobre el manejo de la información (bancos, aseguradoras, casas de bolsa, internet, etc.), en ellas es fácil identificar la importancia de la información, si no existiera ésta dejarían de existir. Sin embargo, hay otras en las que su giro principal es alrededor de la producción, en ellas la información debe identificarse para analizar y perfeccionar su producción (porcentajes de desecho, líneas de producción, distribución de materias, suministro, inventarios y almacenes, procesos internos, publicidad y mercadotecnia, preferencias del cliente, etc.). Para organizaciones públicas, la información debe brindarles el apoyo para cumplir con las necesidades de la sociedad en tiempo y forma.

Desde que se almacenan los datos debe entenderse que tendrían un fin utilitario en algún momento, caso contrario, cualquier dato de control sería desechado instantáneamente. Lo que si surge de súbito es la imprescindible necesidad de dar una respuesta rápida a los requerimientos de información para la toma de decisiones para ayudar a mejorar de alguna manera los procesos internos de negocio.

Para tener automatizada a la organización es necesaria una gran infraestructura en tecnología que soporte sistemas de información tales como Sistemas de Procesamiento de Datos, Sistemas de Manufactura, Administración de Recursos Empresariales (ERP), Sistemas de Información Ejecutiva, Sistemas de Manejo de Relación con Clientes (CRM), entre otros. Por ende, los datos que manejan dichos sistemas siempre serán almacenados en bases de datos y esos datos serán el soporte total a las decisiones de la organización.

Pero el problema de tener esta información acumulada no es su almacenamiento, sino en cómo utilizar esos datos guardados por años para aprovecharlos como un recurso estratégico en la organización. Por eso para descubrir y obtener únicamente la información más importante, es necesario analizar esos datos e interpretarlos.

1.2.1.2 – Toma de Decisiones

La toma de decisiones (TD) es una actividad intelectual que una persona realiza para escoger un curso de acción o elegir un objeto determinado de entre varias opciones, con el objeto de satisfacer una necesidad específica. Es decir, la TD es una expresión de la voluntad del individuo. El ser humano es un tomador natural de decisiones, su forma de vida, éxitos y fracasos están en función en buena medida de sus decisiones. De igual forma las empresas, como organizaciones compuestas por personal cuidan el proceso de toma de decisiones e invierten en los recursos y actividades que les provean los elementos necesarios para realizar una correcta elección, pues su curso y destino dependen de ello.

Dentro del marco organizacional de una empresa, cada uno de los miembros debe tomar decisiones de acuerdo con su labor, responsabilidad y jerarquía. Sin embargo, la cantidad, frecuencia y relevancia de las decisiones será mayor entre más autoridad ostente. Por lo tanto, el tomador de decisiones debe reunir los elementos intelectuales, logísticos e informativos necesarios para efectuar atinadamente su función. Pues el tomador de decisiones es el responsable de los resultados que se obtengan al poner en práctica su voluntad.

1.2.1.2.1 - Proceso de Toma de decisiones

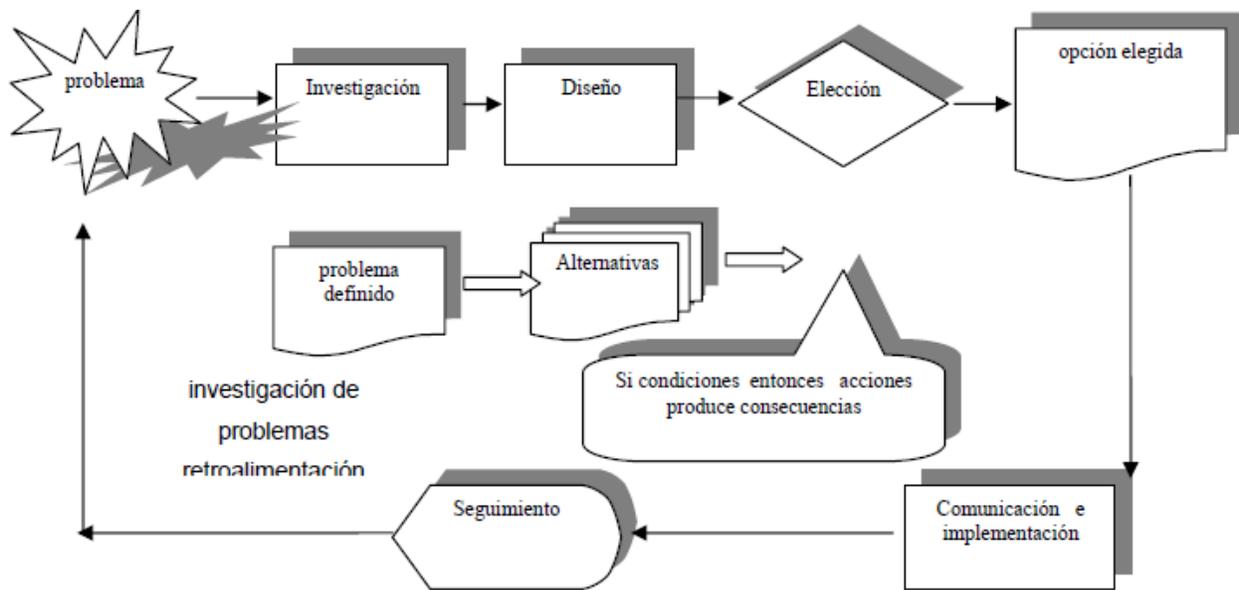


Ilustración 1: Proceso de Toma de decisiones

La Investigación es una tarea de discernimiento e interpretación compuesta por: Identificación de problemas, el rol del tomador de decisiones y la formulación de problemas, de acuerdo con la siguiente descripción.

Identificación de problemas. Busca alguna diferencia entre la situación existente y un estado deseado. Es decir, compara el modelo del estado esperado con el existente, precisa y evalúa las diferencias para determinar si existe un problema.

El rol del tomador de decisiones es la función que intenta no solamente ver de la aparición de las diferencias entre la expectativa y la realidad, sino también prevenir a que esto no ocurra; asumiendo entonces un papel pro y reactivo de acuerdo con las circunstancias¹.

¹ MacGrimmon, K.R. y R.N.Taylor. (1990). *Decision Making & Problem Solving*. capítulo 22.

La formulación de problemas. Para resolver un problema es indispensable identificar su origen, desarrollo y resultados que se han producido o están por suceder. Esta definición debe ser clara, procurando reducir la complejidad conforme a las siguientes estrategias:

- Precisión de los límites. Identifica claramente los elementos implicados en el problema.
- Examen de los argumentos. Los cuales pueden haber precipitado el problema.
- Descomposición del problema. En varios problemas más pequeños y específicos.
- Concentración. En los elementos controlables.

El Diseño es la abstracción, planteamiento de hipótesis, invención, análisis y desarrollo de cursos de acción. Para ello, el responsable debe comprender el problema, generar opciones, considerar su repercusión y estimar la factibilidad de ejecución con base a tres elementos: Condiciones, Acciones y Consecuencias, los cuales se presentan a continuación.

Condiciones. Describen la situación conforme a los valores que toman ciertos atributos, como los números “rojos” en las finanzas de una empresa, constituye un modelo del problema en sí.

Acciones. Es la secuencia de actividades a realizar bajo un programa y recursos determinados, que representa la respuesta de solución al problema.

Consecuencias. Estiman la situación que ocurrirá al cumplir las acciones establecidas, describiendo los valores de los atributos que caracterizan al problema.

La Elección es la toma de decisión que el responsable realiza con el afán de resolver el problema de acuerdo con los criterios considerados en su definición, además de los recursos disponibles e intereses organizacionales en vigor. El desarrollo de esta función clave, es matizado por diversos factores como la magnitud del problema, urgencia en resolverlo, consecuencias, los elementos de certidumbre al alcance del tomador, lo extraordinario que resulta ser el problema, así como los lineamientos establecidos por la propia empresa para normar el proceso. Para efectos de estudio, resulta conveniente describir los criterios de: Conocimiento de los resultados, grado de programación y el Grado de exigencia:

Conocimiento de los resultados. Se define la consecuencia de lo que ocurrirá al escoger una alternativa en función al grado de conocimiento:

- Certeza. Representada por el conocimiento completo y exacto del resultado de cada opción, donde se establece una consecuencia por alternativa.
- Riesgo. Aparece cuando existe la posibilidad de que ocurran varios resultados para cada curso de acción con una probabilidad asociada a ellos.
- Incertidumbre. Se presenta cuando se aprecian múltiples consecuencias para cada alternativa pero se ignora la probabilidad de que ocurran.
- Manejo de certidumbre. Requiere del uso de conocimiento e información especializada, modelos estadísticos y de la investigación de operaciones entre otros.

Grado de programación. Conforme a la naturaleza del problema y a los requerimientos, se pueden emplear dos modelos para estructurar el método de TD:

- Decisiones programadas. Son aquellas que resultan ser susceptibles de expresarse de una manera clara, sencilla y completa, mediante un conjunto de reglas, pudiendo documentarse a través de manuales, normas y políticas. Este modelo se aplica en condiciones de certeza.
- Decisiones no programadas. Su definición responde a situaciones particulares y extraordinarias, resulta complicado establecer un modelo que sirva como referencia tanto para la decisión en turno como para las posteriores. Normalmente, ocurren en respuesta a una crisis, cambios en las condiciones de la organización y de su mercado de trabajo.

Grado de exigencia. Hay dos modelos de toma de decisiones, cuyo alcance se pretende lograr con la decisión, a saber: prescriptivo o normativo, y el descriptivo.

- El modelo prescriptivo o normativo de toma de decisiones. Es aquel modelo que instruye en como tomar una clase de decisión, basado en el criterio de la maximización u optimización de la utilidad o valor esperado que se expresa cuantitativamente viene a ser

la función objetivo para una decisión procurando la utilidad máxima, rendimiento o menor costo. Observa los supuestos de conocer todas las alternativas y sus consecuencias, se busca maximizar el beneficio o utilidad y existe un marco de referencia completo de conocimiento y razonamiento.

- El modelo descriptivo de toma de decisiones. Precisa la manera como se toman actualmente las decisiones, procurando la satisfacción, donde el decisor no está completamente informado sobre las alternativas, ni aplica una racionalidad plena en su búsqueda pues simplifica los factores considerados y limita la exploración de opciones, por lo que acepta la primera que satisfaga todas las restricciones del problema, en lugar de proseguir hasta encontrar el camino óptimo. El modelo está basado en la heurística, asumiendo que el decisor no conoce todas las alternativas ni todos los resultados, hace una exploración limitada para descubrir unas pocas alternativas exploratorias y escoge una opción que cumpla con el nivel mínimo de satisfacción.

La Comunicación e implementación de la decisión una vez tomada la opción es necesario proceder a expresarla a los involucrados (personal, superiores, agentes, etc.), además de precisar el plan para su ejecución, organizar los recursos necesarios y proceder a la dirección de su puesta en marcha para que se realice conforme a los tiempos y formas estipuladas. Seguimiento y retroalimentación insta a supervisar la ejecución de las actividades para detectar y corregir desviaciones del curso y resultados planeados, ejerciendo la retroalimentación constante que inspire a modificar las acciones, los recursos y procedimientos participantes, en aras de llevar alcanzar su objetivo en la solución del problema planteado.

1.2.1.3 – Sistemas de Soporte a las Decisiones

El sistema de soporte a las decisiones (SSD) es un concepto que define un ambiente de trabajo compuesto por el usuario, procedimientos para el tratamiento de información y el equipo de cómputo, orientado a proveer información que apoye las operaciones, la administración y la función de toma de decisiones en una organización.

El SSD es una federación de subsistemas funcionales (producción, contabilidad, recursos humanos, etc.) y de actividades (como la planeación, el control y la estadística) estrechamente integrados; en donde el personal participa en la alimentación de datos, proceso de información y explotación conforme a los requerimientos que exige el desempeño de su labor. Por ejemplo, los empleados suministran el ingreso y control de las transacciones, los supervisores obtienen informes detallados de los resultados de las operaciones, mientras que la gerencia realiza consultas específicas y resúmenes de resultados útiles para el análisis y toma de decisiones.

El SSD se orienta a cubrir los requerimientos de los niveles jerárquicos de una organización, por medio de diversas versiones que integran en conjunto un SSD con los siguientes rasgos:

- Sistema de proceso de transacciones (SPT). Soporta las operaciones cotidianas de la empresa, proporciona información detallada, reportes de estado, imagen de documentos y consulta, procurando la precisión, rapidez de respuesta y facilidad de empleo.
- Sistema de administración de la información (SAI). Brinda la información requerida por los mandos medios para la planeación y control, por medio de reportes agregados, de excepción de resultados, consultas, enfatizando la operación flexible y sencilla.
- Sistema de información gerencial (SIG). Provee información y herramientas para analizar situaciones complejas, tomar decisiones no estructuradas y obtener parámetros de control generales de la empresa. Dirigido a los mandos medios y superiores de la empresa.

1.2.1.4 – Inteligencia de Negocios

La Inteligencia de Negocios (BI) es una combinación de tecnologías de colección de datos y manejo de información, que implementa soluciones orientadas al usuario final para apoyar la toma de decisiones, aprovechando la información estratégica disponible en cualquier parte de la organización. Se basa en la integración de datos históricos y actuales para determinar tendencias futuras y proponer escenarios posibles.

Inteligencia de Negocios es un concepto que se asocia 100% con los niveles directivos, surge de la necesidad de contar con información para dirigir el rumbo de la empresa por los altos mandos, sin embargo, con el tiempo se ha ido ampliando el alcance de este término hasta llegar prácticamente a toda la empresa.

A pesar de relacionar completamente el término con conceptos 100% computacionales, sobre todo las herramientas utilizadas para lograr implementar un desarrollo de este tipo, la verdad es que el concepto no se construye basándose en herramientas computacionales, sino de la formulación de estrategias efectivas de negocios que respondieran a los nuevos tiempos y sus demandas. El énfasis es en los requerimientos y de ahí se desprenden las aplicaciones, es decir, los hombres de negocio dictan las necesidades y la gente técnica investiga y adapta la tecnología para resolver favorablemente esos requerimientos con todos los medios a su alcance. BI se plantea una sinergia entre los Tomadores de Decisiones y las herramientas que emplean, la tecnología está claramente vinculada al management de las empresas, teniendo como resultado obtener ventajas competitivas, producto de decisiones mejor informadas. En función de esta se puede entender a BI como una combinación de tecnología y desarrollo de negocios.

BI no es una metodología, software, sistema o herramienta específica, es más bien un conjunto de tecnologías que van desde arquitecturas para almacenar datos, metodologías, técnicas para analizar información y software entre otros, con un fin común para el apoyo a la toma de decisiones.

El fin último es proveer de información al usuario final para apoyarlo en la toma de decisiones, y esta información puede provenir tanto de los almacenes operacionales como de arquitecturas diseñadas específicamente para el análisis como data mart y data warehouse (se verán en detalle más adelante). El usuario puede necesitar información de cualquier fuente primitiva o derivada para apoyarse en su labor, para lo cual BI utiliza o construye fuentes de datos o de información interna o externa, que son la principal materia prima de esta Tecnología.

Un sistema que exclusivamente brinde información no representa lo que se busca con BI, una segunda característica consiste en organizar y presentar los datos relevantes para que puedan verdaderamente apoyar una Toma de Decisiones. Esto implica tecnologías, técnicas de análisis y todo aquello que sea necesario para obtener de los datos, solo aquella información relevante y útil a la labor del usuario. Recordando el origen de BI, surge en la toma de decisiones para obtener ventajas competitivas producto de decisiones mejor informadas. Si su origen cae en el desarrollo de negocios, es lógico entender que BI sea un apoyo para tal efecto. BI abarca cualquier forma de organizar información, siempre y cuando sostenga la Toma de Decisiones.

Un factor que incidió en la tecnología BI para explotar información fue que el usuario final no poseía conocimientos técnicos que le permitieran tener un acceso sencillo y directo a los datos operacionales, pues esa área está reservada para informáticos. Por tal motivo, el usuario final no tenía de primera mano la información que necesitaba y las consultas no definidas, que son las que tradicionalmente realiza un ejecutivo, eran realizadas por terceras personas (secretarios, asistentes técnicos o gente de sistemas) con la dependencia consecuente. BI incluye herramientas de explotación de información orientadas a usuario final, para eliminar la dependencia de terceras personas. Se pretende brindar las facilidades necesarias para que, con la tecnología, el usuario actúe solo. Las herramientas de BI son sencillas, intuitivas y fáciles de entender y usar; pueden tener diversos fines, como son: Informar, reportar, permitir análisis, identificar tendencias, proyectar, etc. Cualquiera que sea su función final, el común es el mismo: orientación a usuario final.

Para la colección de datos usa o construye almacenes de datos y los maneja con técnicas de análisis y herramientas orientadas al usuario final. Los almacenes de datos son las fuentes

operacionales (bases de datos, archivos de texto, hojas de cálculo, administradores de archivos, etc.), bases de datos operacionales, bases de datos externas, data warehouse y data marts. Las técnicas de análisis principales son los Sistemas de Información Ejecutiva (SEI), Sistemas de Soporte de Decisiones (SSD), Data Mining y Herramientas de Reportes, estas últimas a veces forman parte de las anteriores.

El siguiente gráfico detalla los distintos componentes de la Inteligencia de Negocios y como se integran:

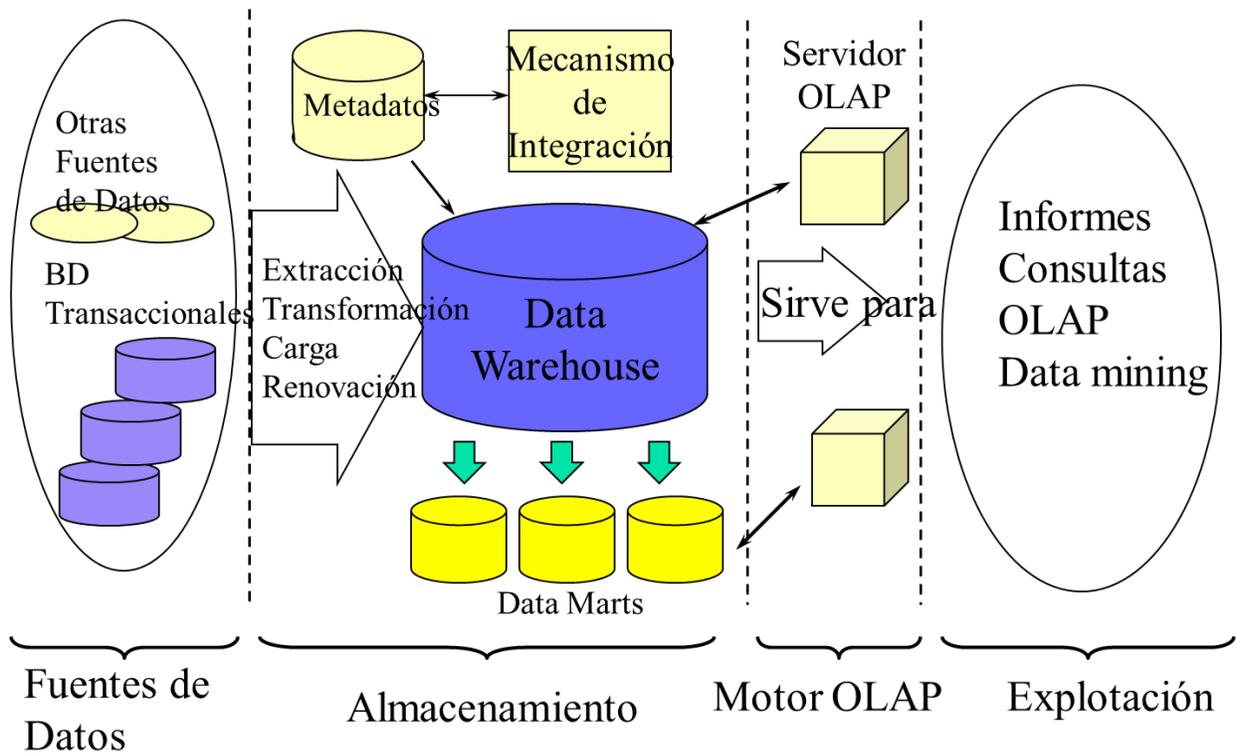


Ilustración 2: Inteligencia de Negocios

1.2.1.5 – Sistemas de Información Ejecutiva

No puede haber decisiones sin información. La toma de decisiones que se lleva a cabo dentro de las organizaciones debe ser rápida, oportuna, fundamentada en datos concretos, que permita tomar decisiones eficientes, efectivas y con un bajo costo para la empresa; pues de ello dependerá el éxito fracaso de la organización.

Un Sistema de información Ejecutiva (SIE-EIS:Executive Information Systems) ayuda a los ejecutivos a contar con información oportuna y precisa para tomar decisiones. Un SIE es un sistema que proporciona al ejecutivo información sobre el desempeño global de la organización. La información se puede recuperar fácilmente y puede presentarse con distintos niveles de detalle².

El SIE es una técnica de más alto nivel dentro de las herramientas de BI. Tiene como objetivo primordial proveer de toda la información necesaria a la gente que toma decisiones, de modo fácil y prácticamente con una mínima interacción con el sistema. En términos formales, "un SIE es un sistema de información que permite a los ejecutivos acceso rápido y efectivo a información compartida, crítica para el negocio, utilizando interfaces gráficas". Las interfaces que son utilizadas en estos sistemas deben ser más sofisticadas que los sistemas transaccionales y deben incluir, en el menor número de páginas posible, la mayor cantidad de información que el usuario necesita para monitorear su empresa.

La diferencia de los SIE con otros sistemas no solo es la vistosidad y facilidad de uso. Aparte del front-end, los SIE interpretan y manipulan de forma diferente la información, pues trabajan con formatos de datos no típicos, tales como data warehouse o data mart. Prácticamente todos los SIE obtienen sus datos de matrices multidimensionales denominadas "cubos" y las herramientas en las que se desarrollan estos sistemas tienen tecnología que permite realizar consultas amplias y complejas de diversas fuentes de datos en tiempos mínimos.

² McLeod Jr, Raymond. (2000). *Sistemas de información genencial*, 7^a ed. Prentice Hall Hispanoamericana. P. 444

Estos sistemas proporcionan medios sumamente fáciles de usar para consulta y análisis de la información. Generalmente se diseñan para el usuario que necesita conseguir los datos rápidamente, pero quiere utilizar el menor tiempo posible para comprender el uso de la herramienta. Tienen como meta obtener información correcta para las personas adecuadas en el momento conveniente, de tal forma que tomen decisiones que pueden valer un incremento significativo de sus ganancias.

Este tipo de sistemas se enfocan primordialmente a proporcionar información de la situación actual de la compañía y dejan en un plano secundario la visualización o proyección de esta información en escenarios futuros. Se construyen generalmente mediante la integración de software diseñado para operar conjuntamente con la infraestructura y las aplicaciones de información existentes en la organización.

Beneficios de un Sistema de Información Ejecutiva:

- Información a tiempo. Los ejecutivos de la organización cuentan con información importante, precisa y relevante sobre la operación de la empresa. Con esta información, el ejecutivo y administrador del negocio pueden tomar decisiones más acertadas y precisas.
- Sensibilidad al medio. Se obtiene información de mayor calidad, independientemente de cuál sea el origen o el medio. Lo que pretende es obtener información competitiva.
- Efectividad de ejecutivos. Mejora la comunicación entre ejecutivos de la organización, aumentando el desempeño y ahorrando tiempo.
- Cumplimiento de los objetivos estratégicos. Con información precisa, se pueden hacer mejores planes para una mejor toma de decisiones. Y cuando se presente un problema, también se pueden tener alternativas de solución óptimas para el cumplimiento de los objetivos.
- Economía en el uso de recursos. Se ahorran costos en papeleo y se obtienen una mejor respuesta al cambio en las necesidades de la sociedad.

- Aumento de flexibilidad. El uso de un mismo sistema de información para compartir datos, permite a las organizaciones lograr la ventaja competitiva, sin alterar el funcionamiento de sus departamentos o sucursales ni su interacción con otras organizaciones.
- Ampliación de servicios complementarios. Compartir un sistema común puede originar propuestas de servicios adicionales como por ejemplo, servicios de facturación electrónica, promociones, ofertas y descuentos.
- Reducción y/o eliminación de los costos administrativos. Poseer mayor información de manera electrónica aumenta las posibilidades de reducir el movimiento de papel típico en estas operaciones.
- Reducción de intermediarios. La conexión a través de una red de comunicaciones entre los productores de la información (personal operativo) con el consumidor final (estratega), podría eliminar la necesidad de representantes e intermediarios, o en su caso provocar la redefinición de sus funciones.

1.2.1.6 – Data Warehouse

Un data warehouse (DW) es una base de datos con una estructura orientada al negocio, integrada, variable en el tiempo y no volátil.

- Orientada al negocio. Organiza y presenta los datos desde la perspectiva de los conceptos que maneja la organización. Los datos tienen el nivel de detalle y la estructura que necesitan los que toman decisiones.

- Integrada. Se construye a partir de fuentes de datos heterogéneas (bases de datos relacionales, archivos planos, hojas de cálculo, documentos impresos, entre otros). Se genera una limpieza e integración de los mismos (se unifican denominaciones, codificaciones, formatos).
- Variable en el tiempo. El horizonte temporal del DW es más amplio que el de los sistemas transaccionales, contiene datos con mayor vida útil y se manejan datos históricos. La fecha del dato es fundamental para generar una marcación temporal.
- No volátil. En el DW los datos no se modifican, es decir, no hay aplicaciones de usuario que modifiquen datos. Los DW se renuevan y los datos permanecen intactos entre renovaciones. Sólo existen las operaciones de carga y acceso.

El data warehouse convierte entonces los datos operacionales de una organización en una herramienta competitiva, haciéndolos disponibles a los empleados que lo necesiten para el análisis y la toma de decisiones.

El objetivo del DW es el de satisfacer los requerimientos de información interna de la empresa para una mejor gestión. El contenido de los datos, la organización y estructura son dirigidos a satisfacer las necesidades de información de los analistas y usuarios tomadores de decisiones.

El DW puede verse como una bodega donde están almacenados todos los datos necesarios para realizar las funciones de gestión de la empresa, de manera que puedan utilizarse fácilmente según se necesiten.

Los sistemas transaccionales son dinámicos, constantemente se encuentran actualizando datos. Analizar esta información puede presentar resultados distintos en cuestión de minutos, por lo que se deben extraer y almacenar fotografías de datos, para estos efectos, con la implicancia de un consumo adicional de recursos de cómputo. Llevar a cabo un análisis complejo sobre un sistema transaccional, puede resultar en la degradación del sistema, con el consiguiente impacto en la operación del negocio.

Los DW generan bases de datos tangibles con una perspectiva histórica, utilizando datos de múltiples fuentes que se fusionan en forma congruente. Estos datos se mantienen actualizados, pero no cambian al ritmo de los sistemas transaccionales. Muchos data warehouses se diseñan para contener un nivel de detalle hasta el nivel de transacción, con la intención de hacer disponible todo tipo de datos y características, para reportar y analizar. Así un data warehouse resulta ser un recipiente de datos transaccionales para proporcionar consultas operativas, y la información para poder llevar a cabo análisis multidimensional.

El DW está basado en un procesamiento distinto al utilizado por los sistemas operacionales, es decir, este se basa en OLAP -Procesos de Análisis en Línea- (OnLine Analysis Process), usado en el análisis de negocios y otras aplicaciones que requieren una visión flexible del negocio.

Para ampliar los conceptos anteriores, en la siguiente tabla se exponen las principales diferencias entre los sistemas Transaccionales (OLTP) y los basados en data warehouse.

Transaccionales	Basados en Data Warehouse
Admiten el acceso simultáneo de muchos usuarios -miles- que agregan y modifican datos.	Admiten el acceso simultáneo de muchos usuarios -cientos- que consultan y no modifican datos.
Representan el estado, en cambio constante, de una organización, pero no guardan su historial.	Guardan el historial de una organización.
Contienen grandes cantidades de datos, incluidos los datos extensivos utilizados para comprobar transacciones.	Contienen grandes cantidades de datos, consolidados y transformados. También de detalle pero solo los necesarios para el análisis.
Tienen estructuras de base de datos complejas.	Tienen estructuras de Base de datos simples.
Se ajustan para dar respuesta a la actividad transaccional.	Se ajustan para dar respuesta a la actividad de consultas.
Proporcionan la infraestructura tecnológica necesaria para admitir las operaciones diarias de la empresa.	Proporcionan la infraestructura tecnológica necesaria para admitir análisis de los datos de la empresa.

Los analistas carecen de la experiencia técnica necesaria para crear consultas "ad hoc" contra la compleja estructura de datos.	Pueden combinar datos de orígenes heterogéneos en una única estructura homogénea y simple, facilitando la creación de informes y consultas.
Las consultas analíticas que resumen grandes volúmenes de datos afectan negativamente a la capacidad del sistema para responder a las transacciones en línea.	Organizan los datos en estructuras simplificadas buscando la eficiencia de las consultas analíticas más que del proceso de transacciones.
El rendimiento del sistema cuando está respondiendo a consultas analíticas complejas puede ser lento o impredecible, lo que causa un servicio poco eficiente a los usuarios del proceso analítico en línea.	Contienen datos transformados que son válidos, coherentes, consolidados y con el formato adecuado para realizar el análisis sin interferir en la operatoria transaccional diaria.
Los datos que se modifican con frecuencia interfieren en la coherencia de la información analítica.	Proporcionan datos estables que representan el historial de la empresa. Se actualizan periódicamente con datos adicionales, no con transacciones frecuentes.
La seguridad se complica cuando se combina el análisis en línea con el proceso de transacciones en línea.	Simplifican los requisitos de seguridad.

Tabla 1: Diferencias entre Sistemas Transaccionales y Basados en Data Warehouse

Trabajar con bases de datos dimensional es solo un paso que permitirá la implementación de la solución de data warehouse. Sin embargo, una implementación de este tipo de tecnología debe complementarse con los siguientes aspectos:

- Integrar datos de diversas fuentes.
- Cuidar la calidad de los datos que van a ser consolidados hacia el data warehouse.
- Crear los mecanismos de sincronización de las fuentes de datos con el data warehouse para asegurar una actualización constante del mismo cuando se creen nuevos datos dentro de dichas fuentes.
- Cuidar los aspectos de desempeño relacionados con la infraestructura tecnológica como servidores y plataformas de sistemas manejadores de bases de datos y OLAP.

- Dar mantenimiento a la estructura de las bases de datos dimensionales y relacionales para implementar nuevos requerimientos de información.

1.2.1.6.1 - Características de un DW

Un data warehouse posee las siguientes características:

- Integra información. Es una herramienta que va a permitir consolidar información de muy diversas fuentes con la ventaja de guardar datos históricos que ya no son útiles en los sistemas operacionales pero que sin embargo sirven para análisis de escenarios, determinación de pronósticos, comparación de comportamientos organizacionales, etc.
- Está dirigido al usuario. Generalmente, estas aplicaciones están dirigidas al personal que conforma la alta gerencia o dueños del negocio y a todas aquellas personas encargadas de tomar decisiones diariamente. La finalidad de estas herramientas es apoyar en la toma de decisiones para que sea un proceso más rápido y con mayor precisión.
- Evoluciona con el tiempo. Son herramientas que van a tener que cambiar y ser adaptadas a requerimientos cada vez más precisos por parte de sus usuarios finales, ya que van a demandar información más exacta y estratégica, de acuerdo al contexto de la organización y a las condiciones de mercado. Por lo que el personal de TI tendrá que realizar ajustes, buscar nuevas fuentes de información, cambiar las reglas de extracción, carga y cálculo de la información, etc.
- Está orientado a la toma de decisiones. Un data warehouse se construye para consolidar información de diversas fuentes en una sola y con ello apoyar en la mejor toma de decisiones y así obtener una ventaja competitiva.
- Su uso está enfocado a la alta gerencia. Un data warehouse no está orientado a los procesos operativos de la organización. Sino más bien a los ejecutivos o dueños del

negocio que son quienes toman las decisiones y guían el curso de la empresa de acuerdo a la información almacenada en el data warehouse.

- Es no volátil. En el DW los datos no se modifican, es decir, no hay aplicaciones de usuario que modifiquen datos. Los DW se renuevan y los datos permanecen intactos entre renovaciones. Sólo existen las operaciones de carga y acceso. El agregado de información y eventualmente la corrección de datos se realiza por un proceso que realiza el ETL.

1.2.1.6.2 - Construcción e Implementación de un Data Warehouse

Una buena estrategia para el desarrollo de un data warehouse en grandes empresas, es construirlo poco a poco. En algunas organizaciones puede consistir en un proceso a largo plazo que involucre las siguientes actividades:

1. Determinación de los requerimientos de información del negocio. Realiza las actividades definidas en la etapa de visualización y el diseño conceptual de la planeación.
2. Determinar las fuentes de información existentes como sistemas operacionales, intranets, Internet, hojas de cálculo, etc. Es recomendable que se busquen fuentes estructuradas de información aunque es posible también consolidar fuentes no estructuradas como los documentos de Word y libros de Excel.
3. A partir de los sistemas operacionales, desarrollar data marts para cada uno de ellos. Estos mercados de datos tendrán como objetivo consolidar la información más importante para la toma de decisiones dentro de una sola área. Apoyarán en las actividades de reportes, análisis de información, análisis de escenarios, etc. Más adelante se estudiará con mayor detalle el concepto de data mart.

4. Consolidar la información de cada uno de los data marts construidos hacia el data warehouse. A su vez, estos mercados de datos son independientes unos de otros, servirán para consolidar la información hacia un data warehouse que contendrá las métricas necesarias para su análisis. Cada vez que se construya un data mart, se podrá enviar su información hacia un data warehouse el cual contendrá toda la información de la organización, pero deberá estar consolidada de tal forma que apoye en el análisis y proceso de la toma de decisiones de la alta gerencia.
5. Formulación de los procesos de análisis de datos en línea (OLAP). Construir herramientas de software necesarias para explotar la información del data warehouse. Pueden ser portales o páginas web, algunas herramientas de reportes o construir un Sistema de Información Ejecutiva que permita la explotación de esta información.

Estos cinco pasos son, a grande rasgos, las actividades que se pueden llevar a cabo para la construcción de un data warehouse. Se hace mayor énfasis en el uso de sistemas operacionales para la construcción de los data marts debido a que generalmente se puede tener un mayor control y una mayor certidumbre de la información almacenada en el data warehouse, cuando esta proviene de sistemas de información que contienen la regla del negocio específica para las áreas que los usen, así como un conjunto de restricciones y de elementos de seguridad que van a garantizar que la información procesada en la base de datos transaccional tenga la menor cantidad de errores posibles. A diferencia de las hojas de cálculo y de otras fuentes de información en las que es menos probable que se tengan elementos de seguridad que garanticen la integridad de la información que se almacena en los data marts y en el data warehouse.

Esta estrategia de construir un data warehouse a partir del desarrollo de data marts, tiene algunas ventajas. En primer lugar, se obtiene el mejor desempeño en la consolidación de la información en el DW a partir de información que ya se encuentra depurada en un data mart. Esto disminuirá los tiempos de consulta y explotación de la información. Y en segundo lugar, se tiene la ventaja de desarrollar un sistema de DW modular que permitirá detectar rápidamente el origen de alguna información incorrecta, ya que no es lo mismo consolidar y buscar información de todas las

fuentes de datos organizacionales directamente en un DW, que buscar el origen de dato erróneo en un data mart, el cual es muy particular y específico de un área.

En la implementación también se debe tener en cuenta que en muchos casos se debe recoger datos de fuentes que estén geográficamente dispersas para después integrarlos y finalmente entregarlos a usuarios que también estén separados físicamente en delegaciones o sedes. Por lo tanto, la infraestructura de comunicaciones se vuelve crítica en este tipo de soluciones.

La forma en la cual se estructure el almacenamiento de datos DW, genera una clasificación respecto a la forma de implementar una arquitectura DW. La estructura adoptada para el data warehouse se debe realizar de la manera que mejor satisfaga las necesidades empresariales, siendo entonces dicha elección factor clave en la efectividad del DW. Las implementaciones más utilizadas son:

- EL DW central: es una implementación de un solo nivel con un solo almacén para soportar los requerimientos de información de toda la empresa. En el DW central todos los usuarios de la organización acceden a la misma base de datos.

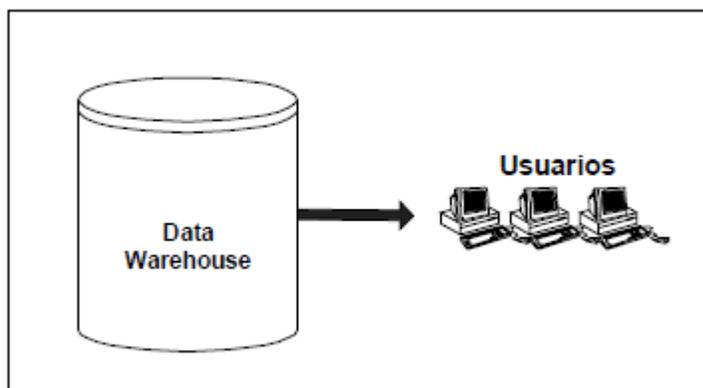


Ilustración 3: Data Warehouse Central

- El DW distribuido: es también una estructura de un nivel, pero particiona el almacén para distribuirlo a nivel departamental. En el DW distribuido, cada departamento, área o línea de negocio dispone de una base de datos propia con la información que solo le compete a los usuarios pertenecientes a estas áreas.

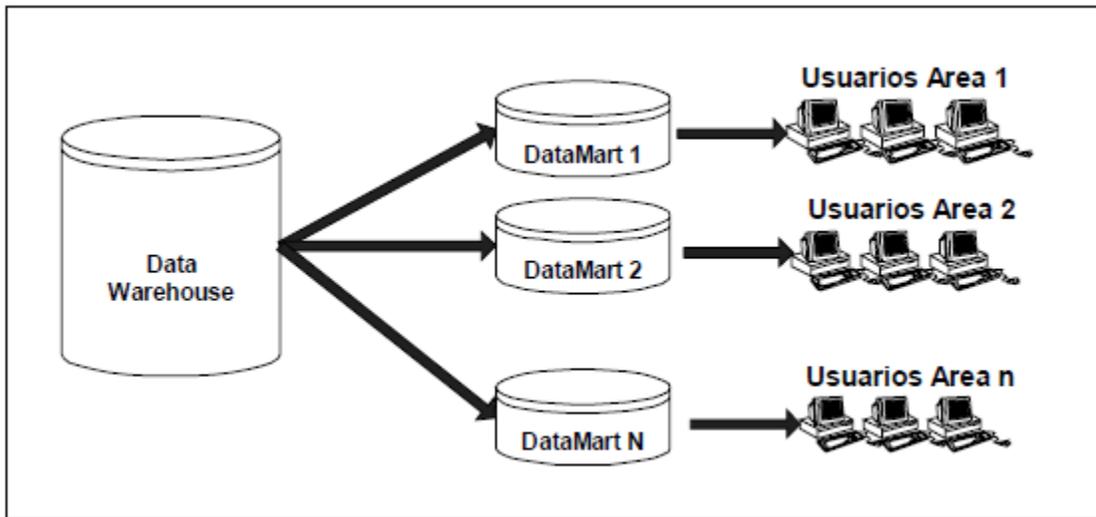


Ilustración 4: Data Warehouse Distribuido

- El DW de dos niveles: combina ideas de los dos anteriores, se implementa el almacén empresarial como los departamentales. En el DW de dos niveles se dispone de una base de datos, generalmente de detalle o de información común a todos los usuarios y además cada departamento, área o línea de negocio dispone de su propia base de datos.

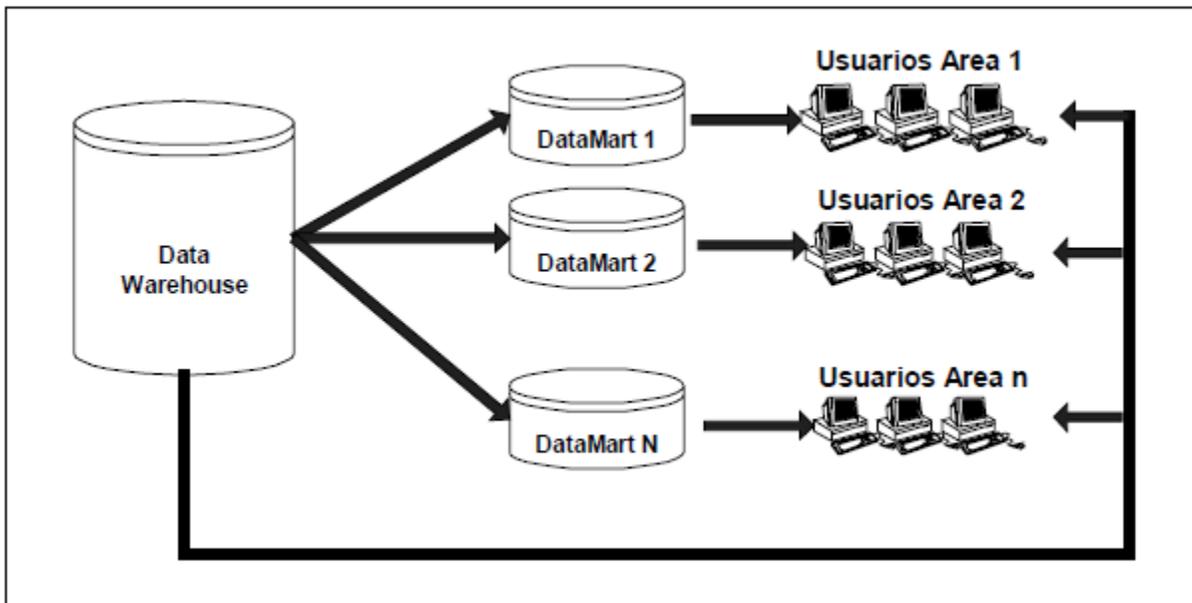


Ilustración 5: Data Warehouse de 2 Niveles

1.2.1.6.3 - Personal involucrado

Existe personal interno y externo a la organización, que invariablemente deberá participar en el diseño y desarrollo de un data warehouse ya que sin su apoyo, sería muy difícil identificar claramente los requerimientos de la información.

Por parte de la organización, el personal que participe en la implementación de un data warehouse debe ser aquel que esté directamente involucrado con la operación diaria y manejo de la información. Esto permitirá modelar el proceso de negocio e identificar el flujo de la información, facilitando la determinación de los indicadores necesarios para la toma de decisiones.

También deben estar involucrados los usuarios ejecutivos o de la alta gerencia quienes serán finalmente los usuarios de la herramienta y serán quienes utilizarán su información para tomar decisiones. Así que solamente ellos podrán determinar cuáles serán las métricas del negocio de alto nivel que necesitarán visualizar para su control.

Otro tipo de personal involucrado es el técnico responsable de implementar la solución de data warehouse. Generalmente es personal contratado por consultoría o por outsourcing. Pero será necesario que en el desarrollo de todas las actividades de construcción, esté involucrad también el personal técnico interno de la organización, con el fin de garantizar que el conocimiento del desarrollo y mantenimiento de la nueva herramienta se quede dentro de la empresa y evitar así la dependencia hacia el consultor.

En resumen el personal involucrado deberá ser el siguiente³:

³ Gill, H. S. (1996). *Data warehousing La integración de la información para la mejor toma de decisiones*. Prentice Hall. P. 24

- Los ejecutivos e inversionistas invierten en la tecnología del data warehouse y pagan por el desarrollo de la solución. Son responsables de aprobar el presupuesto y valorar si la retribución del data warehouse conviene a su inversión. Les interesa la permanencia de su inversión y su retribución continua.
- Los arquitectos de sistemas son responsables de establecer las bases del Data warehouse. Son los responsables de interpretar los requerimientos del inversionista y diseñar la arquitectura del data warehouse que los cumpla. Les interesa la duración, la modulación, la facilidad de implementación y el grado de complejidad con sus tecnologías y sistemas existentes.
- Los constructores son los profesionales de la tecnología de información que realmente construirán el Data warehouse. Son responsables del tiempo, la calidad, el desempeño y la satisfacción de las personas que realmente emplearán la herramienta cada día.
- Los usuarios son los profesionales de la organización y su personal de apoyo que accederán y utilizarán el data warehouse como una herramienta de soporte de decisiones en su trabajo cotidiano. Son personas cuyo trabajo se basa en la información, y los datos que adquieren les ayudará a formular recomendaciones empresariales. Son especialistas en analizar la información de diversas maneras para tratar de destacar hechos acerca de clientes, mercado y productos, y utilizarla como una ventaja competitiva.

1.2.1.6.4 - Madurez de un Data Warehouse en la Organización

Conforme las organizaciones comienzan a utilizar un data warehouse para realizar las actividades diarias y cuando los usuarios se dan cuenta de la importancia que puede llegar a tener una herramienta de este tipo, el sistema de data warehouse se convierte en un sistema de misión crítica. Generalmente este tipo de aplicaciones comienzan a utilizarse con ciertas reservas y

dudas sobre su funcionamiento. Incluso en algunas organizaciones se utilizan en paralelo los mismos reportes con los que operaban antes de que el DW arrancara en producción.

Sin embargo, cuando los usuarios se dan cuenta de que pueden tener en unos cuantos minutos la misma información que anteriormente recibían después de un largo proceso de integración que hacían los responsables de diferentes áreas, entonces es cuando ya no podrán dejar de utilizar el data warehouse para tomar decisiones.

A partir de ese momento el data warehouse se convertirá en un sistema de misión crítica, ya que los usuarios tendrán la confianza de operarlo e incluso la falla del sistema podría ocasionar un mal funcionamiento del negocio.

La aceptación de la nueva herramienta conlleva un proceso de evolución y de madurez tanto del personal como de la misma tecnología. A pesar de las bondades tecnológicas que en la actualidad existen para analizar la información de un data warehouse, no deja de ser algo nuevo para los usuarios que finalmente esperarían contar con una herramienta que tomara las decisiones casi por ellos con solo hacer clic sobre un botón.

El personal que utilice el Data warehouse debe estar realmente consciente de las necesidades y de la visión del negocio. Debe tener claro hacia dónde se dirige la organización y su estrategia para alcanzar sus objetivos. De alguna manera, los usuarios del DW deben tener un cierto perfil metodológico para analizar la información. Finalmente, un DW contiene grandes volúmenes de información y si no se sabe por dónde comenzar a analizar, es muy fácil perderse entre un océano de datos y más datos.

Conforme la organización comienza a usar cada vez más la información del data warehouse para las actividades diarias, su disponibilidad comienza a ser cada vez más importante.

Los requerimientos para un sistema de misión crítica incluyen lo siguiente⁴:

⁴ Gill, H. S. (1996). *Data warehousing La integración de la información para la mejor toma de decisiones*. Prentice Hall. P. 20

- Disponibilidad
- Consistencia y precisión
- Fuerza
- Estandarización
- Actividades basadas en requerimientos
- Compatibilidad con la tecnología existente y la infraestructura de la base de datos
- Uso diario
- Uso amigable
- Alto desempeño
- Verificable
- Seguro

1.2.1.6.5 - Proceso ETL

El proceso ETL (siglas de Extracción, Transformación y Carga en inglés) es un proceso que consume el setenta por ciento del tiempo y esfuerzo de la construcción de un data warehouse; y como dicen las iniciales es un proceso que se compone de tres actividades:

- Extracción. Consiste en extraer los datos desde las diversas fuentes para ser usados en el data warehouse. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación. En esta etapa se deben analizar los datos extraídos,

resultando un chequeo que verifica si los datos cumplen la pauta o estructura que se esperaba. De no ser así los datos son rechazados.

- **Transformación.** Los datos se depuran, completan y transforman. Se aplican una serie de reglas de negocio sobre los datos extraídos para convertirlos en datos que serán cargados.
- **Carga.** Los datos se ordenan, se consolidan, se calculan los datos derivados, se verifica la integridad y se incorporan al data warehouse.

Este es un proceso repetitivo por lo cual es muy complejo y consume muchos recursos; pero sin él sería imposible nutrir el data warehouse con los datos provenientes de diferentes fuentes y con distintos formatos.

1.2.1.6.6 - Ventajas del uso de un Data Warehouse

En la siguiente tabla se analizarán algunas de las ventajas del uso de data warehouse.

Ventajas
Integrador de sistemas
Información consolidada, limpia, consistente, histórica, válida
Enfoque al análisis complejo y eficiente
Automatizable
Normalizador

Tabla 2: Ventajas de un Data Warehouse

1.2.1.7 – Data Marts

El acceso a los datos de toda la empresa a veces no es conveniente (o necesario) para determinados usuarios que solo necesitan un subconjunto de estos datos, en estos casos se utilizan los data marts. El concepto data mart es una especialización del data warehouse, y está enfocado a un departamento o área específica, como por ejemplo los departamentos de Finanzas o Marketing, permitiendo así un mejor control de la información que se está abarcando. El data mart se puede restringir a un tipo particular de datos, a determinada función de negocio, a una unidad de negocio específica o a un área geográfica⁵.

Los principales beneficios de utilizar data marts son:

- Acelerar las consultas reduciendo el volumen de datos a recorrer
- Estructurar los datos para su adecuado acceso por una herramienta
- Dividir los datos para imponer estrategias de control de acceso
- Segmentar los datos en diferentes plataformas hardware
- Permite el acceso a los datos por medio de un gran número de herramientas del mercado, logrando independencia de estas.

⁵ Kroenke, D. (2003). *Procesamiento de bases de datos*. Prentice Hall. P. 541

Tanto los data marts como el data warehouse pueden ser explotados utilizando las herramientas OLAP y pueden ser utilizados para la construcción de Sistemas de Información Ejecutiva.

1.2.1.7.1 - Construcción de un Data Mart

Para construir un data mart, es necesario hacer énfasis en tres aspectos muy importantes: el diseño, la explotación y su mantenimiento.

La fase de diseño contempla realizar un análisis detallado del proceso de negocio del área que se desea modelar y de las posibles fuentes de información disponibles. De un buen diseño, depende que el data mart cumpla con las expectativas del usuario final. Por lo tanto, un modelo de un proceso siempre deberá ir validado y autorizado por el área responsable con el fin de que el constructor esté seguro de que lo que se va a desarrollar, realmente aporte valor a toda la organización.

La explotación de la información implica el uso de herramientas que permitan utilizar la información almacenada en el data mart. Pueden utilizarse herramientas de análisis multidimensional que permitirán realizar consultas desde diferentes enfoques y de acuerdo a las necesidades muy particulares de diferentes usuarios. También pueden utilizarse hojas de cálculo e incluso desarrollar aplicaciones de carácter ejecutivo que permitan consultar la información en tableros de control. Ya existen en el mercado algunos productos de software que proporcionan herramientas para generar tableros de control y reportes ejecutivos de una forma rápida.

El mantenimiento del data mart es una de las fases en la que también deberá ponerse mucho cuidado. La actualización de la información almacenada es un aspecto muy importante para las organizaciones. Por este motivo el data mart debe ser lo suficientemente flexible, de tal forma que se pueda agregar nueva información o quitar información obsoleta. También es importante considerar la automatización de archivos de carga de información, con el fin de que esta tarea se

repita de manera periódica obteniendo la información de los sistemas operacionales y consolidando y calculando información.

Por otro lado, aunque no menos importante se encuentra la capacitación en el uso y explotación de los data marts. Si bien es cierto que este tipo de herramientas ayudan a agilizar los procesos de toma de decisiones de los usuarios, también es cierto que la capacitación es indispensable para un mayor aprovechamiento de la tecnología. Lo importante es que el usuario no sea dependiente de las áreas técnicas para generar consultas. Es necesario dotar a los usuarios de aquellas herramientas de reporte que les permitan realizar consultas y que el data mart responda en tiempos adecuados.

Un data mart tiene componentes similares a los del data warehouse. La principal diferencia está en el enfoque del usuario final. Como en el caso de un data warehouse, cada componente o bloque contempla el desarrollo de un conjunto de procesos para:

- Diseñar el modelo de datos con base en análisis de requerimientos.
- Definir y crear la estructura de los metadatos que almacenarán la información en el data mart.
- Filtrar la información de las fuentes de datos.
- Ajustar la información al formato requerido para el data mart.
- Integrar la información en un bloque de datos para ser importados al data mart.
- Consolidar y agregar la información en el data mart.
- Ejecutar procesos de cálculo una vez consolidada la información en el data mart.
- Explotar la información almacenada en el data mart.

1.2.1.7.2 - Requerimientos para la Construcción de un Data Mart

Una forma de apreciar los beneficios es analizar las necesidades empresariales con las que deberá cumplir el desarrollo de un data mart. Las necesidades son diferentes para cada tipo de usuario: El inversionista/dueño de la empresa, el arquitecto o constructor de la solución y el usuario final⁶.

El inversionista o dueño de la empresa participa en la definición de la visión de lo que será el data warehouse o data mart. Participa en la planeación y coordinación de todo el proyecto. Por su capacidad de toma de decisiones y de control sobre todas las actividades de la organización, este usuario tiene la facultad de hacer que las cosas se lleven a cabo. Por lo tanto, él autoriza la construcción de la solución así como los costos del desarrollo.

El arquitecto o constructor de la solución se encarga de definir el modelo de datos así como de estimar las necesidades tecnológicas para el desarrollo y operación del data warehouse o data mart. Si es personal interno a la organización, este usuario se responsabiliza también por el mantenimiento de la aplicación, así como de la automatización de los procesos de carga y consolidación de información en la solución de data warehousing. Si es personal externo a la organización, este usuario se responsabiliza por dejar el conocimiento necesario al personal técnico interno que se encargará de mantener y automatizar las tareas a fin de que la solución se ejecute con el menor número de errores y con la mínima intervención humana.

El usuario final es la persona que utilizará diariamente la solución de data warehousing. Su principal interés es que la herramienta le proporcione la información necesaria para conocer el estado del negocio y tomar así mejores decisiones aumentando la productividad y la eficiencia de su trabajo. Estos usuarios necesitan contar con la información y visualizarla de diferentes formas, utilizando herramientas de análisis multidimensional y análisis de datos.

⁶ Gill, H. S. (1996). *Data warehousing La integración de la información para la mejor toma de decisiones*. Prentice Hall. P. 22

1.2.1.7.3 - Data Mart Distribuido

Muchas organizaciones, sobre todo los grandes corporativos, están divididas en organizaciones más pequeñas que se especializan en realizar cierta función.

Una de las ventajas de utilizar data marts para construir un data warehouse, es que los data marts pueden estar dispersos geográficamente en diversas localidades. Esto permite que áreas o departamentos exploten la información de sus propios data marts, a la vez que la información almacenada en ellos se puede consolidar hacia el data warehouse empresarial. Esta capacidad de los data marts permite también compartir información entre los mismos.

Beneficios de distribuir data marts en áreas geográficas:

- Permiten que las áreas usuarias por departamento tengan a disposición un mercado de datos de todas las operaciones consolidadas en su área.
- Cuando un usuario del data warehouse necesita consultar una información o algún dato con mayor detalle, puede enlazarse desde el data warehouse hacia el data mart directamente.
- Es más sencillo el mantenimiento de un data mart que el de todo un data warehouse.
- La consolidación de información hacia un data mart es mucho más rápida que intentar consolidar la información de toda una organización hacia un data warehouse.

Los data marts distribuidos son diseñados por el arquitecto o constructor de la solución de data warehousing en respuesta a las necesidades del negocio. El constructor también diseña la manera en la que se van a comunicar los data marts entre sí y la forma en la que se consolidará la información hacia el data warehouse de la organización.

1.2.1.8 – Modelo Dimensional

Como cada proyecto de software, la construcción de una solución de data warehousing inicia por el modelo de datos. Los modelos de datos que se usan para representar los datos almacenados en un data warehouse no son necesariamente los mismos que los usados para representar una base de datos relacional (orientado a la implementación de los procesos transaccionales y de tareas operativas). Los data warehouse y las herramientas OLAP están basadas en un modelo dimensional, orientado a las características del negocio y tareas de análisis. Este modelo de datos es lo que hace que un data warehouse sea una base de datos orientada al negocio.

El modelo dimensional visualiza los datos en forma de un cubo de datos⁷. Un cubo de datos permite modelar y visualizar datos en múltiple dimensiones. Son matrices multidimensionales que sirven para representar el data warehouse y su consulta desde múltiples puntos de análisis.

Este concepto de modelos dimensionales, convierte el análisis de información visto en dos dimensiones (renglones y columnas) en múltiples dimensiones representadas en un cubo de información. Las fases o caras del cubo representan las dimensiones. Las dimensiones son las variables del negocio, y sirven para agrupar, mostrar y filtrar. Por ejemplo, supongamos que una empresa necesita crear un data warehouse para almacenar sus registros de ventas con respecto al tiempo, producto y sucursal. Las dimensiones le permitirán al usuario poder obtener información semejante a las ventas por mes, clasificándola por tipo de producto y por sucursal. Cada dimensión tiene una tabla asociada, a la cual se la llama tabla de dimensión. Por ejemplo, una tabla de dimensión para producto puede contener los atributos nombre, marca y tipo.

⁷ Han, J. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.

Un modelo dimensional se organiza con base en un tema central como las ventas de una empresa. El tema central de un cubo es representado en el modelo dimensional por medio de una tabla de hechos. Las tablas de hechos contienen métricas o parámetros numéricos. Los hechos representan las cantidades que se obtienen mediante las relaciones entre dimensiones. Ejemplos de tablas de hechos para un data warehouse de ventas pueden ser unidades vendidas e importe de ventas.

Aunque usualmente representamos un cubo como una estructura tridimensional, en un modelo dimensional los cubos de datos pueden tener n dimensiones. Cada dimensión o cara del cubo es identificada por una etiqueta que representa la percepción que el usuario tiene de los datos que esa cara representa.

Para comprender mejor los cubos de datos y los modelos dimensionales, apliquemos un ejemplo de un cubo de datos de tres dimensiones (representación 3-D). El siguiente cubo contiene información referente a las ventas de sucursales en periodos de fechas determinadas.

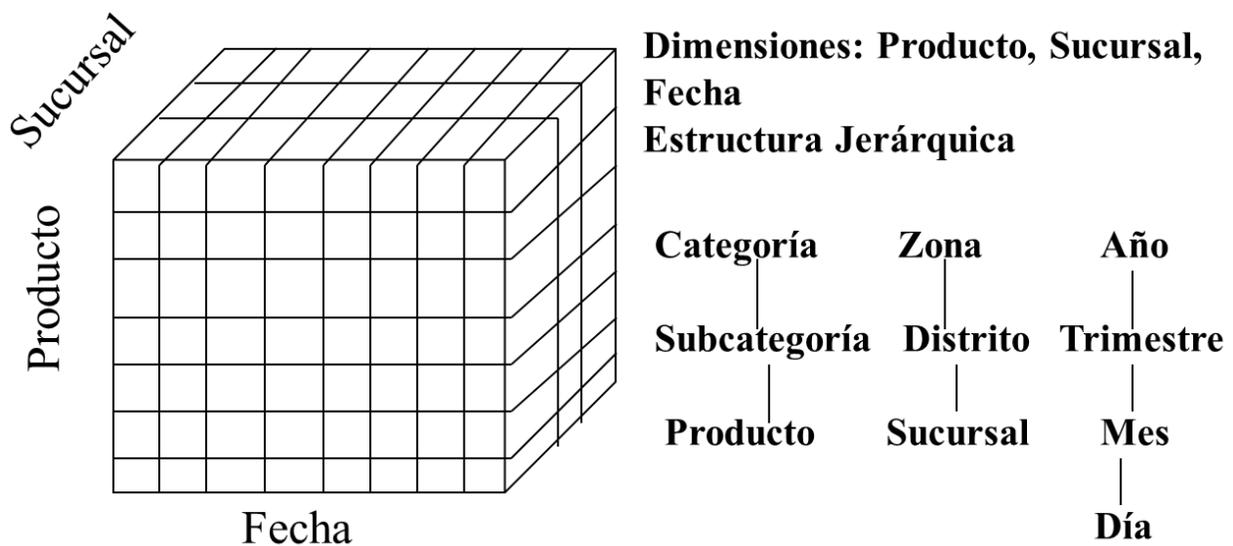


Ilustración 6: Cubo de tres Dimensiones

Cada dimensión es una arista del cubo. En las celdas del cubo están las medidas o hechos (importe o cantidad de ventas). Como podemos ver en el gráfico cada dimensión posee una jerarquía. Las jerarquías definidas para las dimensiones permiten bajar el nivel de complejidad de los datos y ayudar al usuario a ver los datos en el nivel de detalle que necesita. Por ejemplo la dimensión “sucursal” posee una jerarquía que comienza desde la “zona”, dividiéndola por “distrito” y el mismo por “sucursal”. Esto es aconsejable siempre y cuando el usuario necesite agrupar las ventas por un nivel superior en dicha dimensión. Por ejemplo, en este caso para el ejecutivo es factible poder saber cuáles productos se vendieron menos en el año “x”, en la zona “z”; ya que esa información le posibilita realizar una estrategia para relanzar cada producto según la zona.

Trabajar con cubos de datos para construir un data mart o un data warehouse tiene diferentes ventajas:

- Permitir que los usuarios generen sus consultas de diferentes formas y vistas ya que no se requieren conocimientos de programación.
- Compartir y ser la única fuente de información en toda la organización.
- Combinar información de múltiples fuentes y con diversas estructuras de datos.

Con estos cubos de información se pueden construir sistemas de información ejecutiva y Sistemas de ayuda a la toma de decisiones. La responsabilidad del análisis de la información queda en manos del dueño de los datos con el objeto de que extraiga información útil para realizar predicciones y determinar sus propios indicadores, liberándole así el tiempo necesario para definir qué información desea consultar y no cómo la puede obtener.

Un modelo dimensional está organizado alrededor de un tema central como las ventas de una empresa.

1.2.1.8.1 - Modelo de Estrella

El modelo de estrella es el modelo más simple para desarrollar un data warehouse. Es llamado así porque el diagrama asemeja a una estrella. En el centro de la estrella se tiene una tabla llamada tabla de hechos y alrededor se tienen otras tablas a las que se les llama tablas de dimensión.

El modelo de estrella es el más sencillo para diseñar un data warehouse o un data mart. Existe solo una relación entre la tabla de hechos contra cada tabla de dimensión. Esto genera un mejor rendimiento en las consultas.

En el modelo estrella la tabla de hechos está en tercera forma normal (no tiene filas repetidas), mientras las tablas de dimensiones están en segunda forma normal (todos los productos de una misma familia llevan como atributo el nombre de la familia)

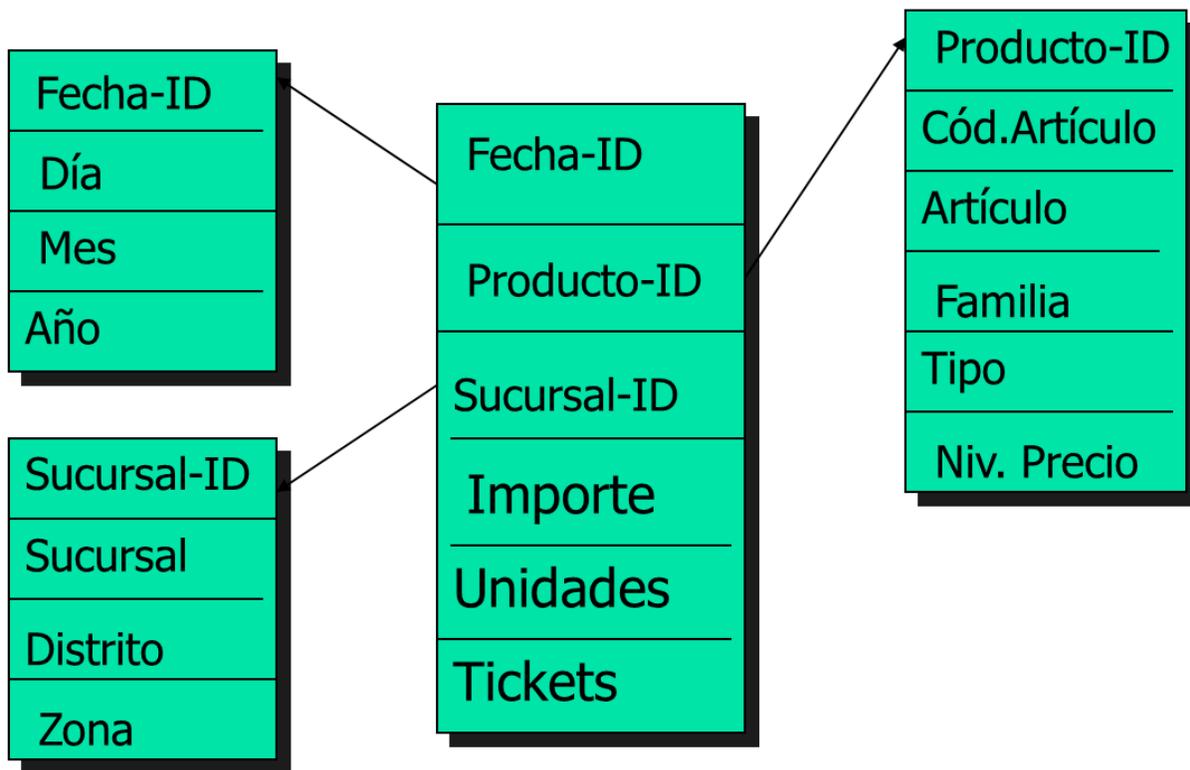


Ilustración 7: Modelo Estrella

1.2.1.8.2 - Modelo de Copo de Nieve

La diferencia entre el modelo de copo de nieve y el de estrella es que todas las tablas, tanto la de hechos como las de dimensiones, están en tercera forma normal. Este tipo de esquema, a cambio del de estrella, se caracteriza por tener tablas dimensionales relacionadas con otras tablas dimensionales además de vincularse a la tabla de hechos.

El modelo de copo de nieve se utiliza cuando hay tablas de dimensiones con grandes cantidades de filas.

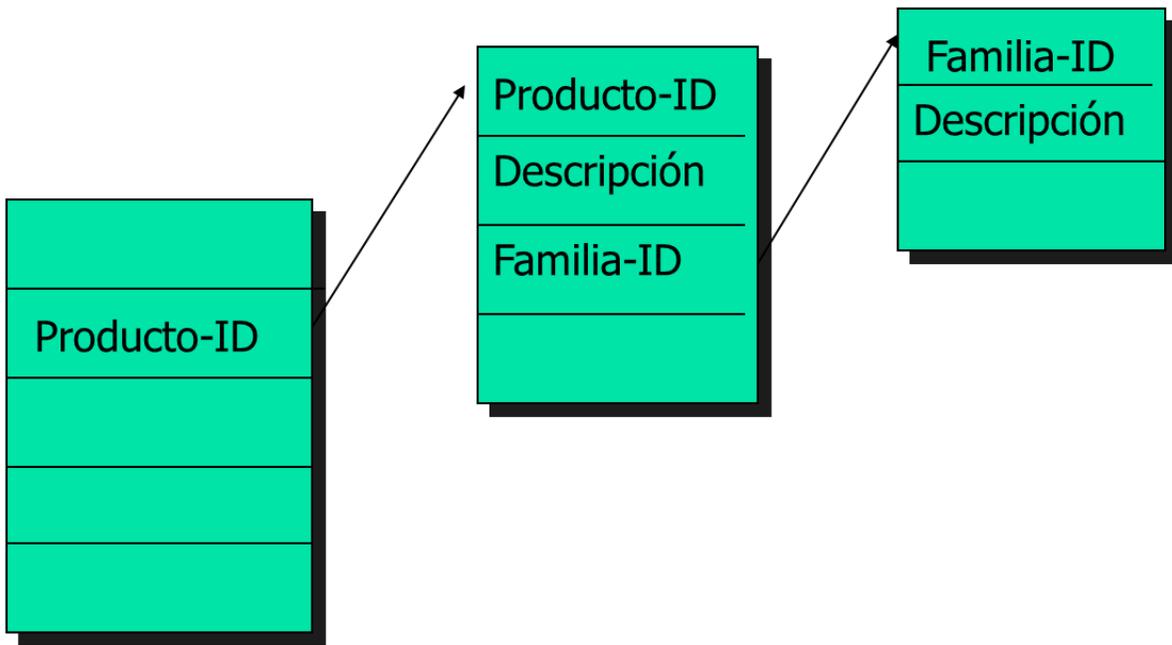


Ilustración 8: Modelo Copo de Nieve

1.2.1.9 – Proceso OLAP

Es el proceso de almacenar y administrar datos sobre la base de las variables del negocio (dimensiones), para permitir a los profesionales de negocios visualizarlos y analizarlos para entender cuál es su significado.

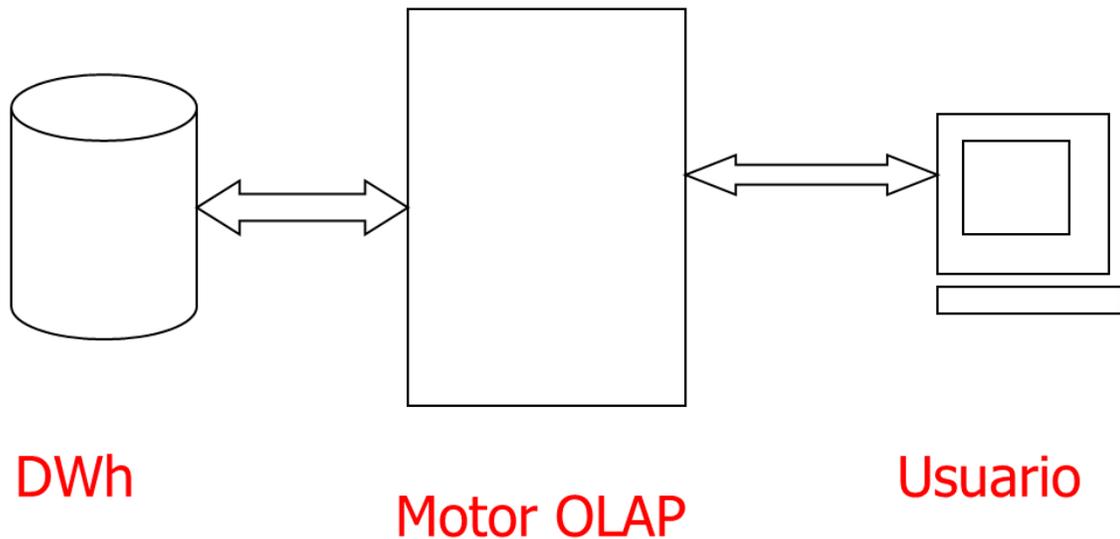


Ilustración 9: Proceso OLAP

La tecnología de Procesamiento Analítico en Línea –OLAP- (Online Analytical Processing) permite un uso más eficaz de los data warehouses para el análisis de datos en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas utilizada generalmente para sistemas de ayuda para la toma de decisiones. Primero y más importante, el OLAP presenta los datos a los usuarios a través de un modelo de datos intuitivo y natural. Con este estilo de navegación, los usuarios finales pueden ver y entender más efectivamente la información de sus bases de datos, permitiendo así a las organizaciones reconocer mejor el valor de sus datos.

En segundo lugar, el OLAP acelera la entrega de información a los usuarios finales que ven estas estructuras de datos como cubos denominadas multidimensionales debido a que la información es vista en varias dimensiones. Esta entrega es optimizada ya que se prepararan algunos valores calculados en los datos por adelantado, en vez de realizar el cálculo al momento de la solicitud.

La combinación de navegación fácil y rápida le permite a los usuarios ver y analizar información más rápida y eficientemente que lo que es posible con tecnologías de bases de datos relacionales solamente. El resultado final: se pasa más tiempo analizando los datos y menos tiempo analizando las bases de datos.

A pesar del proceso de almacenamiento de datos de preparar información para el consumo del usuario final se debe facilitar la búsqueda de la información. Estos retos son enfrentados con herramientas avanzadas de peticiones (queries), las cuáles ocultan al usuario final la complejidad de las base de datos. Esta es la función de las herramientas OLAP.

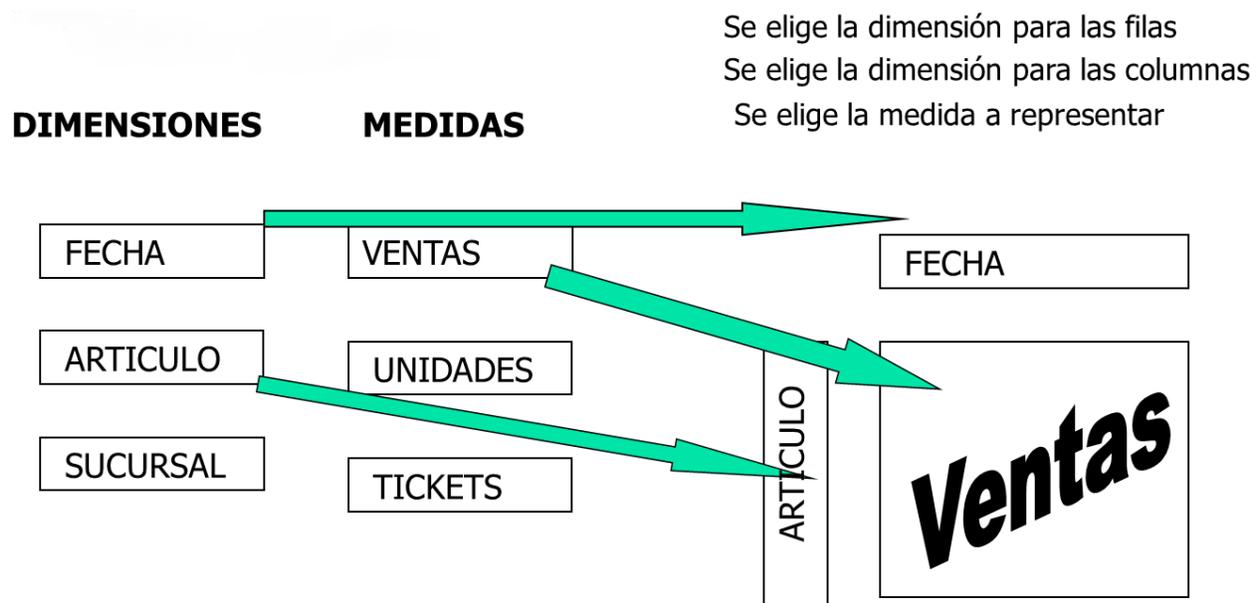


Ilustración 10: Representación Final del Proceso OLAP ante la Vista del Usuario

Las aplicaciones OLAP deberían proporcionar análisis rápidos de información multidimensional compartida. Las características principales del OLAP son:

- **Rápido:** proporciona la información al usuario a una velocidad constante. La mayoría de las peticiones se deben de responder al usuario en cinco segundos o menos.

- **Análisis:** realiza análisis estadísticos y numéricos básicos de los datos, predefinidos por el desarrollador de la aplicación o definido específicamente por el usuario.
- **Compartida:** implementa los requerimientos de seguridad necesarios para compartir datos potencialmente confidenciales a través de una gran población de usuarios.
- **Multidimensional:** llena la característica esencial del OLAP, que es ver la información en determinadas vistas o dimensiones.
- **Información:** acceden a todos los datos y a la información necesaria y relevante para la aplicación, donde sea que ésta resida y no esté limitada por el volumen.

El OLAP es un componente clave en el proceso de almacenamiento de datos (data warehousing) y los servicios OLAP proporcionan la funcionalidad esencial para una gran variedad de aplicaciones que van desde reportes corporativos hasta soporte avanzado de decisiones.

La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y unas posibilidades de navegación, seleccionando la información a obtener.

Normalmente este tipo de selecciones se ve reflejada en la visualización de la estructura dimensional, en unos campos de selección que permitan elegir el nivel de agregación (jerarquía) de la dimensión, y/o la elección de un dato en concreto, la visualización de los atributos del sujeto, frente a una(s) dimensiones en modo tabla, pudiendo con ello realizar, entre otras las siguientes acciones:

- **Rotar (Swap):** alterar las filas por columnas (permutar dos dimensiones de análisis)
- **Bajar (Down):** bajar el nivel de visualización en las filas a una jerarquía inferior
- **Detallar (Drill down):** informar para una fila en concreto de datos a un nivel inferior
- **Expandir (Expand):** ídem anterior sin perder la información a nivel superior para éste y el resto de los valores

Los cubos, las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Al describir y representar la información en esta forma, los usuarios pueden navegar intuitivamente en un conjunto complejo de datos. Sin embargo, el solo describir el modelo de datos en una forma más intuitiva, hace muy poco para ayudar a entregar la información al usuario más rápidamente.

Un principio clave del OLAP es que los usuarios deberían de ver tiempos de respuesta consistentes para cada vista de datos que requieran. Dado que la información se colecta en el nivel de detalle solamente, el resumen de la información es usualmente calculado por adelantado. Estos valores pre calculados, son la base de las ganancias de desempeño del OLAP.

En los primeros días de la tecnología OLAP, la mayoría de las compañías asumía que la única solución para una aplicación OLAP era un modelo de almacenamiento no relacional. Después, otras compañías descubrieron que a través del uso de estructuras de base de datos (esquemas de estrella y de copo de nieve), índices y el almacenamiento de agregados, se podrían utilizar sistemas de administración de bases de datos relacionales (RDBMS) para el OLAP.

Estos vendedores llamaron a esta tecnología OLAP relacional (ROLAP). Las primeras compañías adoptaron entonces el término OLAP multidimensional (MOLAP), estos conceptos, MOLAP y ROLAP, se explican con más detalle en los siguientes párrafos. Las implementaciones MOLAP normalmente se desempeñan mejor que la tecnología ROLAP, pero tienen problemas de escalabilidad. Por otro lado, las implementaciones ROLAP son más escalables y son frecuentemente atractivas a los clientes debido a que aprovechan las inversiones en tecnologías de bases de datos relacionales preexistentes. Un desarrollo reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

La arquitectura MOLAP usa unas bases de datos dimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos

multidimensionalmente. Por el contrario, la arquitectura ROLAP cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales.

La arquitectura ROLAP, accede a los datos almacenados en un data warehouse para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales.

Cuando se comparan las dos arquitecturas, se pueden realizar las siguientes observaciones:

- El ROLAP delega la negociación entre tiempo de respuesta y el proceso por lotes al diseño del sistema. Mientras, el MOLAP, suele requerir que sus bases de datos se pre compilen para conseguir un rendimiento aceptable en las consultas, incrementando, por tanto los requerimientos por lotes.
- Los sistemas con alta volatilidad de los datos (aquellos en los que cambian las reglas de agregación y consolidación), requieren una arquitectura que pueda realizar esta consolidación ad-hoc. Los sistemas ROLAP soportan bien esta consolidación dinámica, mientras que los MOLAP están más orientados hacia consolidaciones por lotes.
- Los ROLAP pueden crecer hasta un gran número de dimensiones, mientras que los MOLAP generalmente son adecuados para diez o menos dimensiones.
- Los ROLAP soportan análisis OLAP contra grandes volúmenes de datos elementales, mientras que los MOLAP se comportan razonablemente en volúmenes más reducidos (menos de 5 Gb)

Por ello, y resumiendo, el ROLAP es una arquitectura flexible y general, que crece para dar soporte a amplios requerimientos OLAP. El MOLAP es una solución particular, adecuada para soluciones departamentales con unos volúmenes de información y número de dimensiones más modestos.

1.2.1.10 – Data Mining

Data mining, la extracción de información oculta y predecible de grandes bases de datos, es una tecnología para ayudar a las compañías a descubrir información relevante en sus bases de información (data warehouses). Las herramientas de data mining predicen futuras tendencias y comportamientos. Los análisis prospectivos automatizados ofrecidos por la automatización del data mining van más allá de los eventos pasados provistos por las herramientas usuales de sistemas de soporte de decisión. Es un proceso automático que permite extraer esquemas interesantes, no triviales y previamente desconocidos, de los datos y descubrir relaciones entre variables.

Las herramientas de data mining pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas. Estas herramientas exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar.

Muchas compañías ya colectan y refinan cantidades masivas de datos. Las técnicas de data mining pueden ser implementadas rápidamente en plataformas ya existentes de software y hardware para acrecentar el valor de las fuentes de información existentes y pueden ser integradas con nuevos productos y sistemas pues son traídas en línea (on-line). Una vez que las herramientas de data mining fueron implementadas en computadoras cliente servidor de alta performance o de procesamiento paralelo, pueden analizar bases de datos masivas para brindar respuesta a preguntas tales como, "¿Cuáles clientes tienen más probabilidad de responder al próximo mailing promocional, y por qué? y presentar los resultados en formas de tablas, con gráficos, reportes, texto o hipertexto⁸.

Existen dos tipos de modelos de data mining:

⁸ Groth, Robert. (1998). *Data Mining: A Hands On Approach for Business Professionals*. Prentice Hall.

- **Los modelos predictivos:** Utiliza variables de entrada (atributos descriptivos) y una variable objetivo (atributo objetivo). Se realizan entrenamientos con datos históricos que tienen por objeto descubrir las relaciones entre las variables de entrada y la variable objetivo. Luego en producción utiliza ese conocimiento para predecir el valor de la variable objetivo.
- **Los modelos descriptivos:** Estos en cambio no realizan predicciones, si no que analizan otros aspectos de los datos como la asociación y la segmentación.

El nombre de data mining deriva de las similitudes entre buscar información de negocios en grandes bases de datos, encontrar información de la venta de un producto entre grandes montos de Gigabytes almacenados y minar una montaña para encontrar una veta de metales valiosos. Ambos procesos requieren examinar una inmensa cantidad de material, o investigar inteligentemente hasta encontrar exactamente donde residen los valores. Dadas bases de datos de suficiente tamaño y calidad, la tecnología de data mining puede generar nueva información al proveer las siguientes capacidades:

- Predicción automatizada de tendencias y comportamientos. Data mining automatiza el proceso de encontrar información predecible en grandes bases de datos. Preguntas que tradicionalmente requerían un intenso análisis manual, ahora pueden ser contestadas directa y rápidamente desde los datos. Un típico ejemplo de problema predecible es el marketing apuntado a objetivos (targeted marketing). Data mining usa datos en mailing promocionales anteriores para identificar posibles objetivos para maximizar los resultados de la inversión en futuros mailing. Otros problemas predecibles incluyen pronósticos de problemas financieros futuros y otras formas de incumplimiento, e identificar segmentos de población que respondan a determinadas características.
- Descubrimiento automatizado de modelos previamente desconocidos. Las herramientas de data mining barren las bases de datos e identifican modelos previamente escondidos en un sólo paso. Otros problemas de descubrimiento de modelos incluye detectar

transacciones fraudulentas de tarjetas de créditos e identificar datos anormales que pueden representar errores de tipeado en la carga de datos.

Las técnicas más comúnmente usadas en data mining son:

- **Redes neuronales artificiales:** modelos predecibles no-lineales que aprenden a través del entrenamiento y semejan la estructura de una red neuronal biológica.
- **Árboles de decisión:** estructuras de forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Métodos específicos de árboles de decisión incluyen Árboles de Clasificación y Regresión (CART: Classification And Regression Tree) y Detección de Interacción Automática de Chi Cuadrado (CHAI: Chi Square Automatic Interaction Detection).
- **Algoritmos genéticos:** técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño basado en los conceptos de evolución.
- **Método del vecino más cercano:** una técnica que clasifica cada registro en un conjunto de datos basado en una combinación de las clases del/de los k registro (s) más similar/es a él en un conjunto de datos históricos (donde $k > 1$). Algunas veces se llama la técnica del k-vecinos.
- **Regla de inducción:** la extracción de reglas if-then de datos basados en significado estadístico.

Para aplicar mejor estas técnicas, deben estar totalmente integradas con el data warehouse así como con herramientas flexibles e interactivas para el análisis de negocios (herramientas OLAP). Varias herramientas de data mining actualmente operan fuera del warehouse, requiriendo pasos extra para extraer, importar y analizar los datos. Además, cuando nuevos conceptos requieren implementación operacional, la integración con el warehouse simplifica la aplicación de los resultados desde data mining.

El punto de inicio ideal es un data warehouse. Este data warehouse puede ser implementado en una variedad de sistemas de bases relacionales y debe ser optimizado para un acceso a los datos flexible y rápido. Un server OLAP permite que el usuario analice los datos de acuerdo a como quiera mirar el negocio, resumido por línea de producto, u otras perspectivas claves para su negocio.

El server de data mining debe estar integrado con el data warehouse y el server OLAP para insertar el análisis de negocios directamente en esta infraestructura. A medida que el data warehouse crece, la organización puede aplicar extraer la información oculta y aplicarla en futuras decisiones.

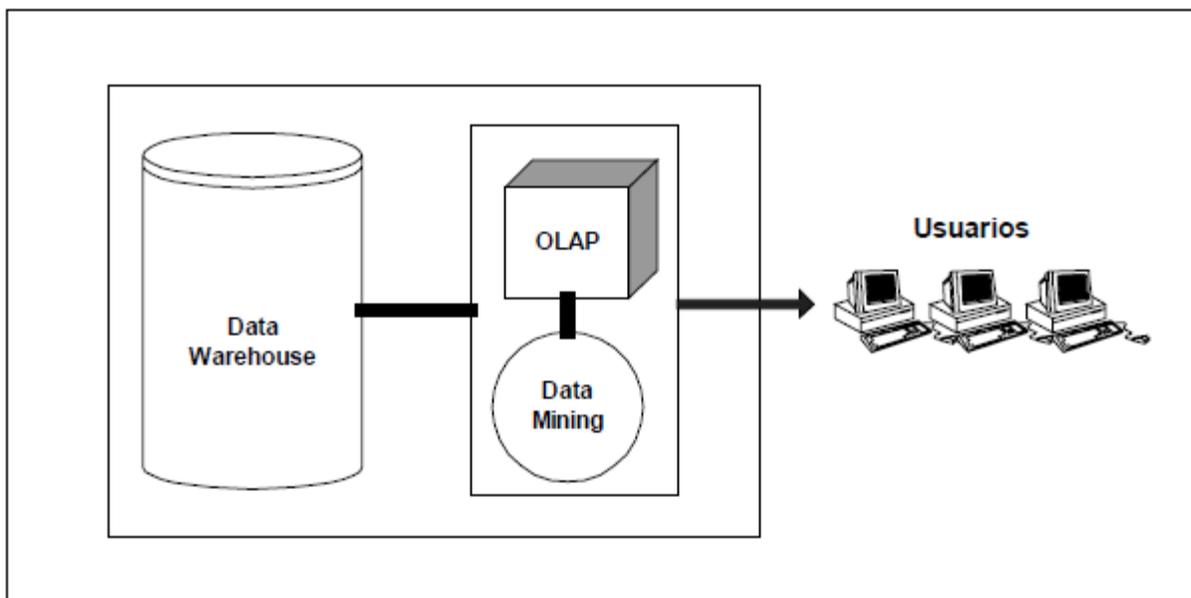


Ilustración 11: Infraestructura de un Análisis de Negocios con Data Mining

1.3. – Justificación del Estudio

En la actualidad la Inteligencia de Negocios es un tema en el que se han realizado numerosos proyectos e investigaciones, pero a pesar de que es un tema estudiado, no se conocen muchos casos de Inteligencia de Negocios para administrar el flujo de información dentro del gobierno y así alcanzar sus objetivos propuestos. Hoy en día existen muchas herramientas capaces de integrar información proveniente de distintos sistemas. Pero las dependencias pueden tener datos dispersos tanto en sistemas tecnológicos avanzados como en programas elementales tipo Excel y dichas herramientas tendrán que tener la capacidad de extraer automáticamente la información relevante que requieren aquellos que tomen decisiones sobre la educación, el empleo, la salud o la pobreza, entre otros. La eficiencia y la eficacia de la gestión pública son de suma importancia en las preocupaciones de la sociedad, por eso este estudio intenta poner en evidencia que el Estado puede sacar provecho de la Inteligencia de Negocios para la toma de decisiones.

1.4. – Limitaciones de la Investigación

Debido a la dificultad para acceder a los datos de todos los organismos del Estado y la diversidad de objetivos para distintos organismos, es inadecuado pensar en un solo gigantesco data warehouse, por lo tanto, el caso práctico incluirá solo aquellos datos que posee la Sindicatura General de la Nación dispersos tanto en sus sistemas tecnológicos como en otros registros elaborados con programas elementales como Excel.

1.5. – Alcance de la Investigación

En esta investigación se pretende poner en evidencia que el Estado puede sacar provecho de la Inteligencia de Negocios para la toma de decisiones sobre cuestiones actuales que afectan a la sociedad.

El estudio solo profundizará la Inteligencia de Negocios y las herramientas de data warehouse para organismos del Estado. El mismo no pretende dar un modelo dimensional centralizado para todos los datos recolectados por el Estado, sino pretende contribuir a su construcción con un marco teórico y un caso práctico de la investigación.

Para el desarrollo de esta propuesta de investigación, se desarrollará un caso práctico con datos de entrada perteneciente a los generados por los sistemas y registros de la Sindicatura General de la Nación. Se usará como herramienta de recolección e integración de datos el software para data warehouses O3 con licencia para un servidor y un cliente.

1.6. – Objetivos de la Investigación

1.6.1 -- Objetivo General

Poner en evidencia que el Estado puede sacar provecho de la Inteligencia de Negocios para la toma de decisiones sobre cuestiones actuales que afectan a la sociedad encontrando los factores que maximicen la mejora de la gestión del Estado y prediciendo el comportamiento futuro con un alto porcentaje de certeza.

1.6.2 -- Objetivos Específicos

1. Realizar un diagnóstico de la situación actual y un levantamiento de los procesos involucrados.
2. Proponer técnicas para fijar objetivos claros y depurar los datos que sirvan como indicadores claves de desempeño útiles para la toma de decisiones de los organismos estatales.
3. Determinar los principales elementos de los requerimientos que el Estado busca en una modelo dimensional de Inteligencia de Negocios.
4. Implementar un caso práctico del Estado que permita reflejar la construcción de un modelo dimensional.
5. Evaluar y validar los resultados del caso práctico para corroborar la calidad del modelo y las ventajas que produce para la toma de decisiones.

1.7. – Hipótesis

Teniendo en cuenta que el dominio de datos en el Estado es mucho más complejo que en una empresa ya que las fuentes son mucho más diversas, y es necesario poder encapsular todos esos datos dispersos para obtener una medición eficaz, además los patrones son más sensibles a la sociedad, debemos llevar a cabo una estrategia aplicando un modelo dimensional correcto sin sacar del foco central el objetivo institucional; de esta forma las herramientas de data warehouse permiten el análisis y presentación de estudios multidisciplinarios en forma dinámica, lo que

conlleva a la rápida obtención de resultados a través de los cuales pueden identificarse patrones o tendencias de comportamiento en situaciones determinadas ayudando a resolver los temas sensibles de la población. Por ejemplo, si el objetivo institucional es enfrentar irregularidades en la educación, construimos un modelo dimensional teniendo en cuenta los datos y patrones relacionados con el objetivo, que seguramente deberán ser provenientes del Ministerio de Educación y otros organismos relacionados (Ministerio de Ciencia y Tecnología, Educar S.E., etc), y debemos dejar de lado cualquier información que no sea útil para alcanzar dicho objetivo. De esta forma, mantendremos el foco en un conjunto más acotado del universo de datos del Estado.

Las aplicaciones de Inteligencia de Negocios son capaces de integrar información proveniente de distintos sistemas. El gobierno puede tener datos pero gracias a esta herramienta podrán extraer automáticamente la información relevante que requieren aquellos que tomen decisiones para alcanzar sus objetivos maximizando los beneficios y minimizando los riesgos. Las estrategias maximizan la eficiencia en el manejo de la información y la eficacia de los procesos y aplicaciones; pero al mismo tiempo minimizan el riesgo de la rigidez institucional, motivada por la inexistencia de información accesible para generar los conocimientos necesarios, así como el riesgo de tener procesos y sistemas redundantes.

La información seleccionada y extraída de las bases de datos se presenta al usuario en su escritorio a través de pantallas dinámicas, con amplias capacidades para analizar y proyectarla de manera multidimensional, o bien, combinarla con distintos parámetros que arrojen mejores interpretaciones.

2-MATERIAL Y MÉTODOS

2.1. – Material

2.1.1 – Lugar y Tiempo de Trabajo de la Investigación

1. Lugar: La investigación se lleva a cabo en la Sindicatura General de la Nación -Ciudad Autónoma de Buenos Aires, Argentina- con datos recopilados de organismos pertenecientes a la administración pública de todas las provincias de la Nación Argentina.

2. Tiempo: Los datos recopilados pertenecen al periodo que comprende entre Enero del 2011 y Diciembre del 2014

2.1.2 – Descripción del Objeto de Estudio

La investigación incluirá solo aquellos datos que posee la Sindicatura General de la Nación dispersos tanto en sus sistemas tecnológicos como en otros registros elaborados con programas elementales como Excel, pertenecientes a auditorías sobre problemáticas de la sociedad realizadas en el transcurso de Enero del 2011 y Diciembre del 2014.

Se decidió estudiar sobre aquellos datos, ya que los mismos son accesibles y representan un grupo resumido de trabajos realizados para afrontar los distintos tópicos problemáticos que la Nación debe enfrentar año tras año.

2.1.3 – Descripción Población y Muestra

Población

Población Teórica: Todos aquellos datos pertenecientes a distintas fuentes de los organismos del Estado.

Población Estudiada: Datos pertenecientes a la actividad diaria de la Sindicatura General de la Nación sobre las auditorías realizadas por Organismos de Control Gubernamental en el transcurso de Enero del 2011 y Diciembre del 2014.

Muestra

Para el muestreo se utilizarán los datos provenientes del sistema Sisio Red Federal de la Sindicatura General de la Nación y el sistema de Entes. Ya que allí se pueden extraer los datos resultantes de las auditorías realizadas en los distintos puntos del país años tras años.

Se utilizará un método de muestreo no probabilístico, el Muestreo Casual o Incidental, ya que se selecciona directamente los individuos de la población.

Objetivo y Alcance de Sisio Red Federal

La Red Federal de Control Público se instituye como un emprendimiento organizacional que integra las Estructuras Estatales de Fiscalización y Auditoría, con el fin de evaluar el desempeño de los programas sociales ejecutados por las Jurisdicciones de carácter nacional. Se constituye además, como la garante de cobertura de control en el ámbito nacional, a través de relevamientos y acciones directas en el campo de ejecución de la Política Social.

Con motivo del incesante crecimiento del alcance de la Red Federal de Control Público, así como también del volumen de auditorías y programas a auditar, se hizo necesaria la existencia de un sistema de información que sirviese de apoyo, tanto al proceso de planificación de la Red como al

seguimiento de su ejecución. El sistema cubre desde la elaboración de las matrices de planificación de cada Órgano de Control Local hasta la incorporación de los informes de auditoría.

Objetivo y Alcance de Entes

Sistema de información de apoyo a la Sindicatura General de la Nación la cual cubre la carga y obtención de información en detalle de cada Ente perteneciente a la Nación Argentina y sus relaciones.

2.2. – Métodos

2.2.1 –Diseño de la Investigación

Para la obtención y análisis de los datos se utilizará un diseño Empíricos o de campo, ya que los datos de interés se recogen en forma directa de la realidad.

2.2.1.1 - Introducción

En el marco teórico presentado en el anterior capítulo, se comentaban varias hipótesis en pro de la utilización de inteligencia de negocios para el apoyo en la toma de decisiones dentro de una organización. El planteamiento expuesto justifica dichas hipótesis en base a considerar las ventajas generadas aplicando inteligencia de negocios para la toma de decisiones sobre

cuestiones actuales que afectan a la sociedad, aprovechando al máximo la información generada por el Estado.

Una revisión en profundidad de los estudios más relevantes publicados sobre este tema, muestra que aún no existe una evidencia empírica que permita validar las hipótesis teóricas. Entre los estudios realizados, existen distintos comportamientos según que indicador se considere. Además, se observa un distinto nivel de madurez en el desarrollo de inteligencia de negocios en organismos del Estado.

De acuerdo con la voluntad de consolidar una línea de investigación, se plantea este estudio, centrado en buscar una evidencia empírica de las hipótesis teóricas en el caso del sistema de auditorías Sisio Red Federal de la Sindicatura General de la Nación, considerando todas las auditorías generadas por la misma en todo el país entre el periodo que abarca Enero del 2011 a Diciembre del 2014. Las hipótesis planteadas en el primer capítulo expresan los objetivos concretos que se persiguen.

Este capítulo tiene por objetivo formalizar el diseño del estudio empírico que se ha realizado. En los apartados que siguen, se presentarán en detalle los parámetros que caracterizan la investigación, las captaciones de horas y cantidad de auditorías realizadas en base a la muestra de las bases de datos, y las consideraciones generales acerca de la metodología estadística utilizada para contrastar las Hipótesis de Investigación planteadas.

2.2.1.2 – Parámetros del Estudio

El objetivo básico de esta investigación, tal y como se citaba en el primer capítulo, es poner en evidencia que el Estado puede sacar provecho de la Inteligencia de Negocios para la toma de decisiones sobre cuestiones actuales que afectan a la sociedad encontrando los factores que

maximicen la mejora de la gestión del Estado y prediciendo el comportamiento futuro con un alto porcentaje de certeza.

Cabe añadir que el objetivo básico presentado se analiza considerando los datos recolectados por la Sindicatura General de la Nación.

Tomando en consideración dicho objetivo y el alcance citado, se deducen que los parámetros de diseño más relevantes son:

- La delimitación de la actividad del Estado.
- Estructura del Sisio Red Federal y Entes.
- El ámbito geográfico de las auditorías.
- El intervalo temporal de captación de datos.
- La muestra de organismos seleccionados
- La muestra de programas seleccionadas
- Las auditorías observadas, tipos de auditorías, estados de auditorías y la cantidad de horas observadas.

2.2.1.2.1 - La delimitación de la actividad del Estado

En el primer capítulo se justificaban las razones que indujeron a seleccionar para este estudio la inteligencia de negocios para los organismos del Estado.

La primera de las razones entonces expuestas, hacía referencia a la necesidad de centrar el estudio en un objetivo institucional en concreto, eligiéndose de organismos con actividades gubernamentales.

En pro de poder aplicar el modelo teórico basado en suponer una mayor eficiencia de la información aplicando inteligencia de negocios en el Estado, se eligió un organismo que posee como actividad central el control gubernamental en el Estado, la Sindicatura General de la Nación (SIGEN).

Dentro de la SIGEN se ha seleccionado como actividad de estudio la Red Federal, ya que ha sido la actividad con más información disponible y de posible alcance para objeto de estudio. La Red Federal sintetiza la visión del control como constructor de una estructura con mecanismos permanentes de prevención, detección y medición, útiles para minimizar los riesgos a los que se encuentra expuesta toda gestión y, de este modo, propender al cumplimiento de los principios consagrados por la ley y el fortalecimiento del sistema de información pública necesario para optimizar la formulación de las políticas públicas.

2.2.1.2.2 - Estructura del Sismo Red Federal y Entes

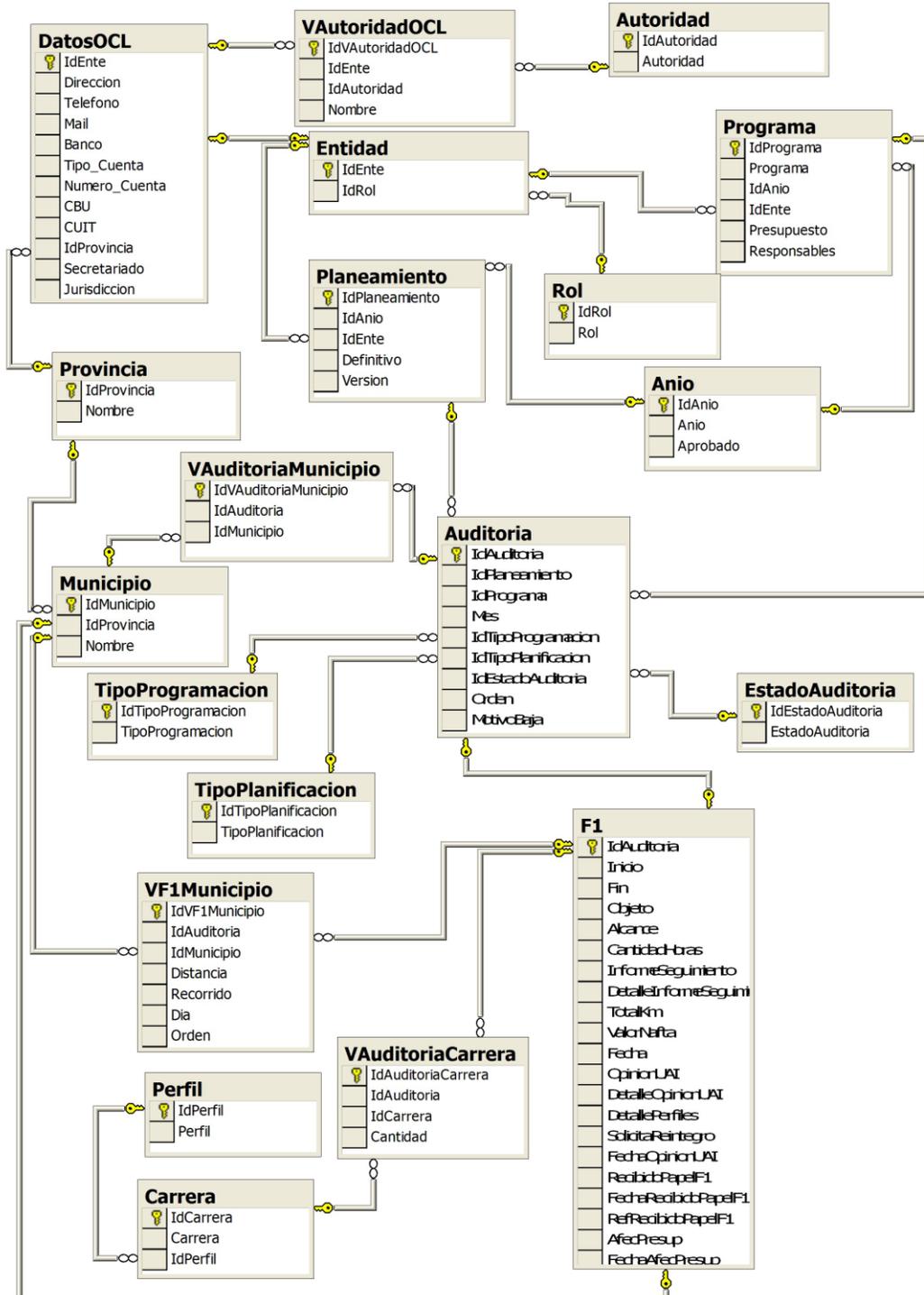
Sismo Red Federal

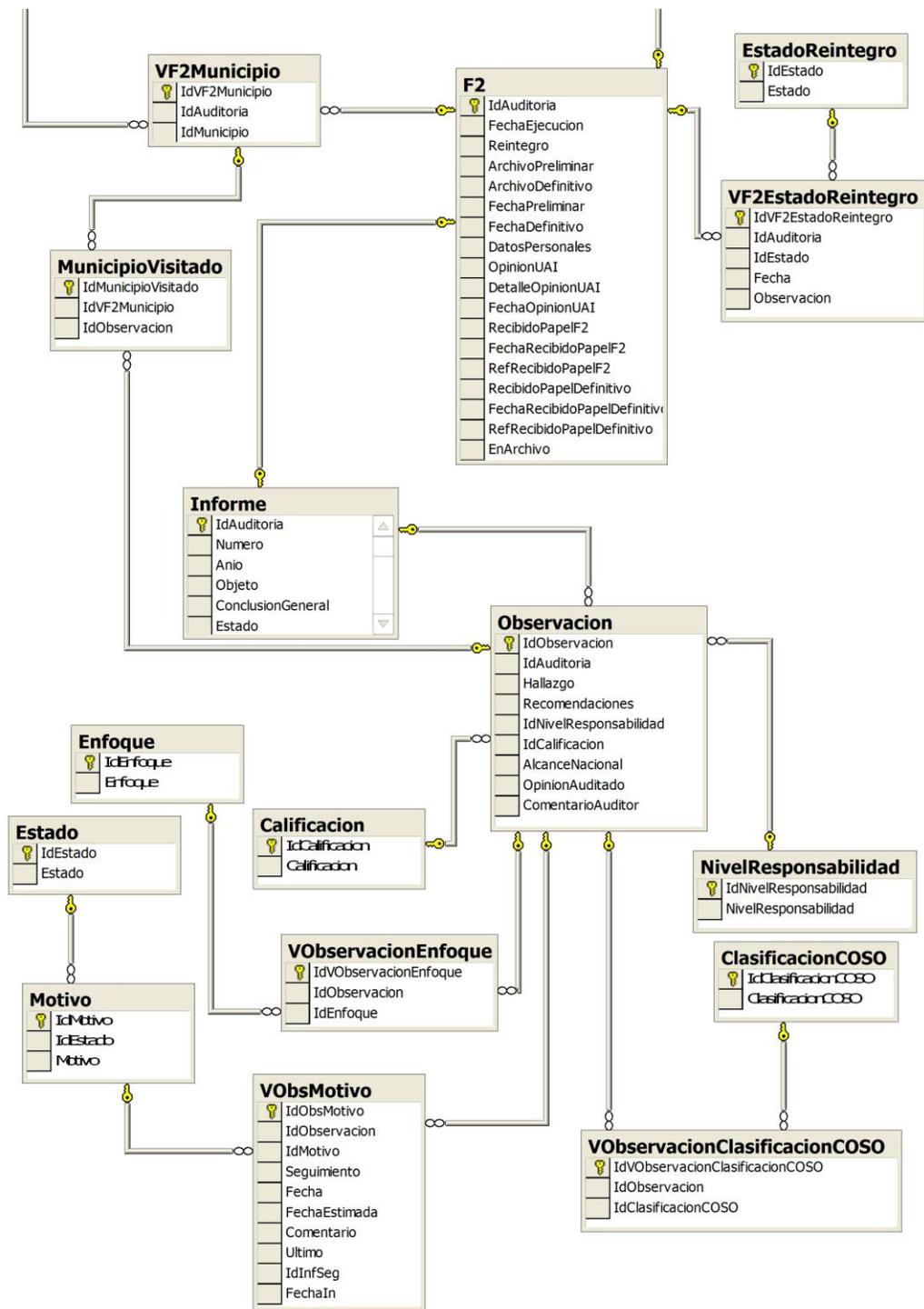
La Red Federal de Control Público se instituye como un emprendimiento organizacional que integra las Estructuras Estatales de Fiscalización y Auditoría, con el fin de evaluar el desempeño de los programas sociales ejecutados por las Jurisdicciones de carácter nacional. Se constituye además, como la garante de cobertura de control en el ámbito nacional, a través de relevamientos y acciones directas en el campo de ejecución de la Política Social.

Con motivo del incesante crecimiento del alcance de la Red Federal de Control Público, así como también del volumen de auditorías y programas a auditar, se hizo necesaria la existencia de un sistema de información que sirviese de apoyo, tanto al proceso de planificación de la Red como al seguimiento de su ejecución. El sistema cubre desde la elaboración de las matrices de planificación de cada Órgano de Control Local hasta la incorporación de los informes de auditoría.

A continuación se muestra la estructura de la base de datos relacional perteneciente al Sisio Red Federal:

SISIO Red Federal - 01/06/2011





2-1

Ilustración 12: Estructura de la base de datos relacional de Sismo Red Federal

Entes

Sistema de información de apoyo a la Sindicatura General de la Nación la cual cubre la carga y obtención de información en detalle de cada Ente perteneciente a la Nación Argentina y sus relaciones.

A continuación se muestra la estructura de la base de datos relacional perteneciente a Entes:

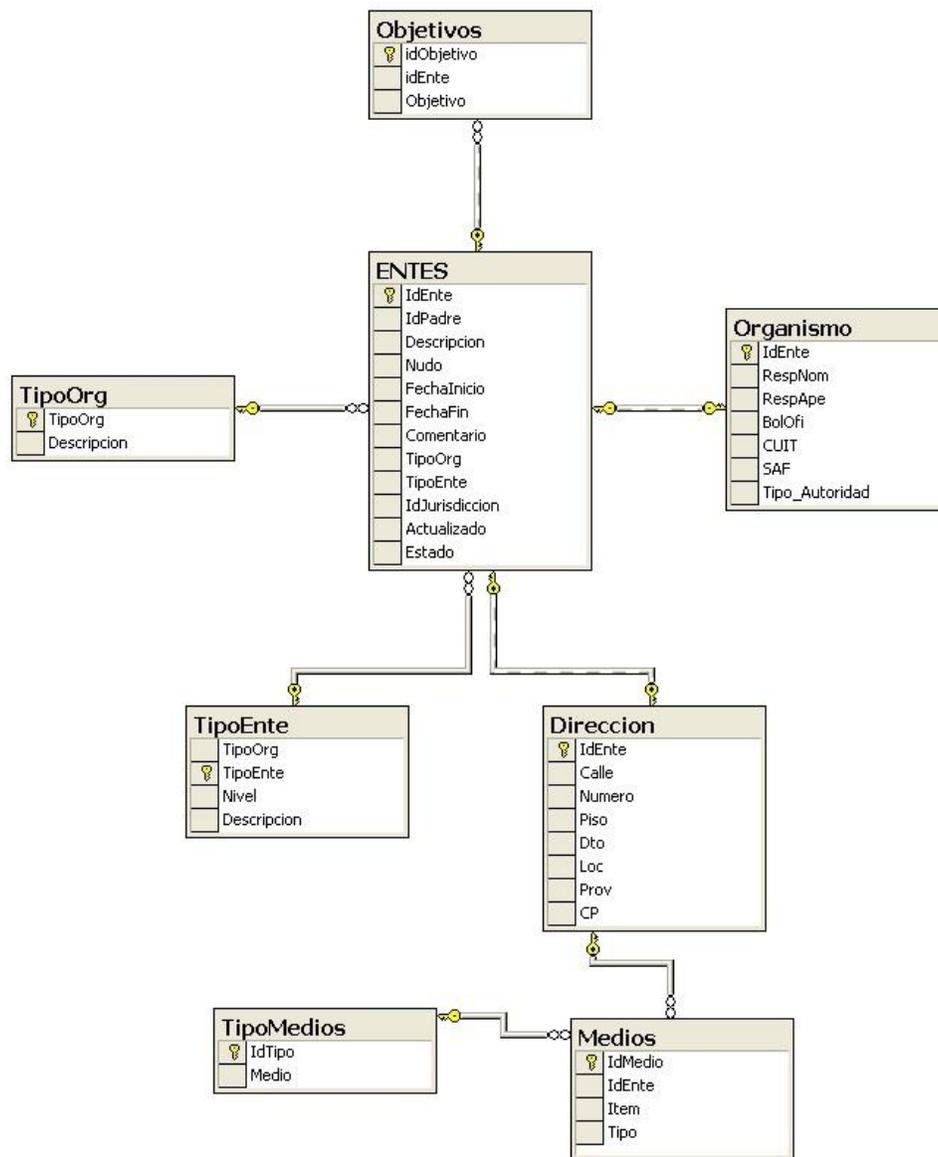


Ilustración 13: Estructura de la base de datos relacional de Entes

2.2.1.2.3 - El ámbito geográfico de las auditorías

El conjunto de auditorías que se han incluido en la muestra pertenecen exclusivamente a las provincias y ciudades incluidas en la Red Federal, y consecuentemente, las conclusiones del estudio son directamente aplicables a dichas auditorías. Cabe destacar que las auditorías de la Red Federal incluyen todas las provincias de la República Argentina y algunas de sus principales ciudades.

A continuación se detallan las regiones incluidas en las auditorías de la Red Federal:

Región de Auditorías	Provincia
BUENOS AIRES - AUDITORIAS PROVINCIALES	Buenos Aires
CABA - AUDITORIAS MUNICIPALES	Capital Federal
CATAMARCA - AUDITORIAS PROVINCIALES	Catamarca
CHACO - AUDITORIAS PROVINCIALES	Chaco
CHUBUT - AUDITORIAS PROVINCIALES	Chubut
COMODORO RIVADAVIA - AUDITORIAS MUNICIPALES	Chubut
CORDOBA - AUDITORIAS PROVINCIALES	Córdoba
CORRIENTES - AUDITORIAS PROVINCIALES	Corrientes
ENTRE RIOS - AUDITORIAS PROVINCIALES	Entre Ríos
FORMOSA - AUDITORIAS PROVINCIALES	Formosa
GENERAL ROCA - AUDITORIAS MUNICIPALES	Río Negro
GOBERNADOR INGENIERO VALENTIN VIRASORO - AUDITORIAS MUNICIPALES	Corrientes
JUJUY - AUDITORIAS PROVINCIALES	Jujuy
LA BANDA - AUDITORIAS MUNICIPALES	Santiago del Estero
LA PAMPA - AUDITORIAS PROVINCIALES	La Pampa
LA RIOJA - AUDITORIAS MUNICIPALES	La Rioja
LA RIOJA - AUDITORIAS PROVINCIALES	La Rioja
MENDOZA - AUDITORIAS PROVINCIALES	Mendoza
MISIONES - AUDITORIAS PROVINCIALES	Misiones
NEUQUEN - AUDITORIAS PROVINCIALES	Neuquén
RIO GRANDE - AUDITORIAS MUNICIPALES	Tierra del Fuego
RIO NEGRO - AUDITORIAS PROVINCIALES	Río Negro
ROSARIO - AUDITORIAS MUNICIPALES	Santa Fe
SALTA - AUDITORIAS MUNICIPALES	Salta
SALTA - AUDITORIAS PROVINCIALES	Salta

SAN CARLOS DE BARILOCHE - AUDITORIAS MUNICIPALES	Río Negro
SAN FRANCISCO - AUDITORIAS MUNICIPALES	Córdoba
SAN JUAN - AUDITORIAS PROVINCIALES	San Juan
SAN LUIS - AUDITORIAS MUNICIPALES	San Luis
SAN LUIS - AUDITORIAS PROVINCIALES	San Luis
SANTA CRUZ - AUDITORIAS PROVINCIALES	Santa Cruz
SANTA FE - AUDITORIAS MUNICIPALES	Santa Fe
SANTA FE - AUDITORIAS PROVINCIALES	Santa Fe
SANTIAGO DEL ESTERO - AUDITORIAS MUNICIPALES	Santiago del Estero
SANTIAGO DEL ESTERO - AUDITORIAS PROVINCIALES	Santiago del Estero
TIERRA DEL FUEGO - AUDITORIAS PROVINCIALES	Tierra del Fuego
TRELEW - AUDITORIAS MUNICIPALES	Chubut
TUCUMAN - AUDITORIAS PROVINCIALES	Tucumán

Tabla 3: Regiones incluidas en las auditorías de la Red Federal

Al incluir todas las provincias, y una significativa concentración del número de principales ciudades, inducen a pensar que los resultados obtenidos puedan ser representativos del ámbito Nacional, al menos, en algunos aspectos.

Considérese esta conclusión como una intuición más que una verdad absoluta, especialmente teniendo en cuenta la dificultad en encontrar fuentes de datos exhaustivas.

2.2.1.2.4 - El intervalo temporal de captación de datos

En primer lugar, es importante concretar el intervalo temporal en que se procederá a la captación de auditorías. Si bien se puede conceptualizar el estudio de manera que abarque un solo año de auditorías, dos razones aconsejan realizar la captación durante un intervalo de tiempo lo suficientemente amplio como para proceder a efectuar varias mediciones a una misma problemática.

La primera de las razones tiene en cuenta que la contrastación del indicador variabilidad requiere el seguimiento temporal de las auditorías de cada programa, para así calcular el número de auditorías finalizadas o suspendidas y la magnitud promedio de los distintos cambios de estados. Bajo una captación de datos estática sería imposible valorar nada acerca de este indicador.

En segundo lugar debe tenerse en cuenta que un proceso de captura de datos que se repita en el tiempo permite captar mejor las diferentes problemáticas analizadas durante un período, y por ende obtener una deducción más precisa de los factores causales de las mismas.

El estudio plantea un intervalo temporal de cuatro años, siendo realizada una captación del planeamiento completo en cada una de ellas. El intervalo comprende entre Enero del 2011 y Diciembre del 2014. Cada año se captaban para cada planeamiento, todas sus auditorías en cada uno de los tribunales de cuentas que formaban la Red Federal.

Cabe tener en cuenta las razones que indujeron a concretar un intervalo de cuatro años. Por un lado, éste es suficientemente amplio como para permitir una comparación significativa de las problemáticas del Estado durante un mandato. Por otro, no ha sido posible captar información providente de años anteriores al 2011, ya que a partir de ese año se informatizó la actividad de Red Federal.

Las siguientes tablas demuestran la cantidad de auditorías de muestras captadas por año por unidad de auditoría interna (UAI):

Unidad de Auditoría Interna	Total de Auditorías 2011
MINISTERIO DE DESARROLLO SOCIAL	54
MINISTERIO DE EDUCACION	60
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS	20
MINISTERIO DE SALUD	72
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL	39
TOTAL DE AUDITORÍAS	245

Tabla 4: Cantidad de auditorías de muestras año 2011

Unidad de Auditoría Interna	Total de Auditorías 2012
MINISTERIO DE AGRICULTURA, GANADERIA Y PESCA	14
MINISTERIO DE DESARROLLO SOCIAL	49
MINISTERIO DE EDUCACION	65
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS	17
MINISTERIO DE SALUD	58
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL	45
TOTAL DE AUDITORÍAS	248

Tabla 5: Cantidad de auditorías de muestras año 2012

Unidad de Auditoría Interna	Total de Auditorías 2013
MINISTERIO DE AGRICULTURA, GANADERIA Y PESCA	12
MINISTERIO DE DESARROLLO SOCIAL	49
MINISTERIO DE EDUCACION	72
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS	20
MINISTERIO DE SALUD	72
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL	31
TOTAL DE AUDITORÍAS	256

Tabla 6: Cantidad de auditorías de muestras año 2013

Unidad de Auditoría Interna	Total de Auditorías 2014
CONSEJO NACIONAL DE COORDINACION DE POLITICAS SOCIALES	9
MINISTERIO DE AGRICULTURA, GANADERIA Y PESCA	20
MINISTERIO DE DESARROLLO SOCIAL	56
MINISTERIO DE EDUCACION	52
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS	21
MINISTERIO DE SALUD	86
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL	36
TOTAL DE AUDITORÍAS	280

Tabla 7: Cantidad de auditorías de muestras año 2014

2.2.1.2.5 – Muestra de Organismos

La captación de auditorías supone previamente haber definido una muestra representativa de organismos encargados de las mismas. Este estudio requiere dos muestras distintas, una representativa de las Unidades de Auditoría Interna y otra de los Organismos de Control Local.

Cada Unidad de Auditoría Interna (UAI) representa a un Ministerio de la Nación Argentina. La cuál es el encargado de proponer cada año programas nacionales para solucionar las dificultades de la sociedad actual. Cada Ministerio posee programas por el cual su fin califica con el de la actividad del mismo. Las auditorías son realizadas sobre los programas, por tal motivo es de suma importancia poseer una muestra de las Unidades de Auditoría Interna para la medición.

A continuación se detallan todas las UAIs incluidas por la Red Federal hasta el 2014:

Unidad de Auditoría Interna
CONSEJO NACIONAL DE COORDINACION DE POLITICAS SOCIALES
MINISTERIO DE AGRICULTURA, GANADERIA Y PESCA
MINISTERIO DE DESARROLLO SOCIAL
MINISTERIO DE EDUCACION
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS
MINISTERIO DE SALUD
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL

Tabla 8: UAIs incluidas en la Red Federal

Un Organismo de Control Local (OCL) es un ente responsable de auditar el programa generado por el Ministerio en su municipio o provincia correspondiente. Son los encargados de encontrar falencias en los programas como desviaciones de la planificación, diferencias en el monto presupuestario, incumplimiento de los objetivos e irregularidades en la ejecución entre otras; como así también medir el impacto social de las políticas aplicadas. Un OCL puede ser municipal (solo genera auditorías del programa en un Municipio) o provincial (genera auditorías a un nivel más alto abarcando toda una provincia). Al ser el responsable de generar las auditorías de la medición es de suma importancia poseer una muestra de los Organismos de Control Local.

A continuación se detallan todos los OCLs incluidos en la Red Federal hasta el 2014:

OCL	Provincia
------------	------------------

BUENOS AIRES - TRIBUNAL DE CUENTAS PROVINCIALES	Buenos Aires
CABA - AUDITORIA GENERAL	Capital Federal
CABA - SINDICATURA GENERAL	Capital Federal
CATAMARCA - TRIBUNAL DE CUENTAS PROVINCIALES	Catamarca
CHACO - TRIBUNAL DE CUENTAS PROVINCIALES	Chaco
CHUBUT - TRIBUNAL DE CUENTAS PROVINCIALES	Chubut
COMODORO RIVADAVIA - TRIBUNAL DE CUENTAS MUNICIPALES	Chubut
CORDOBA - TRIBUNAL DE CUENTAS PROVINCIALES	Córdoba
CORRIENTES - TRIBUNAL DE CUENTAS PROVINCIALES	Corrientes
ENTRE RIOS - TRIBUNAL DE CUENTAS PROVINCIALES	Entre Ríos
FORMOSA - TRIBUNAL DE CUENTAS PROVINCIALES	Formosa
GENERAL ROCA - TRIBUNAL DE CUENTAS MUNICIPALES	Río Negro
GOBERNADOR INGENIERO VALENTIN VIRASORO - TRIBUNAL DE CUENTAS MUNICIPALES	Corrientes
JUJUY - TRIBUNAL DE CUENTAS PROVINCIALES	Jujuy
LA BANDA - TRIBUNAL DE CUENTAS MUNICIPALES	Santiago del Estero
LA PAMPA - TRIBUNAL DE CUENTAS PROVINCIALES	La Pampa
LA RIOJA - TRIBUNAL DE CUENTAS MUNICIPALES	La Rioja
LA RIOJA - TRIBUNAL DE CUENTAS PROVINCIALES	La Rioja
MENDOZA - TRIBUNAL DE CUENTAS PROVINCIALES	Mendoza
MISIONES - TRIBUNAL DE CUENTAS PROVINCIALES	Misiones
NEUQUEN - TRIBUNAL DE CUENTAS PROVINCIALES	Neuquén
RIO GRANDE - TRIBUNAL DE CUENTAS MUNICIPALES	Tierra del Fuego
RIO NEGRO - TRIBUNAL DE CUENTAS PROVINCIALES	Río Negro
ROSARIO - TRIBUNAL DE CUENTAS MUNICIPALES	Santa Fe
SALTA - TRIBUNAL DE CUENTAS MUNICIPALES	Salta
SALTA – SINDICATURA GENERAL	Salta
SALTA - AUDITORÍA GENERAL	Salta
SAN CARLOS DE BARILOCHE - TRIBUNAL DE CUENTAS MUNICIPALES	Río Negro
SAN FRANCISCO - TRIBUNAL DE CUENTAS MUNICIPALES	Córdoba
SAN JUAN - TRIBUNAL DE CUENTAS PROVINCIALES	San Juan
SAN LUIS - TRIBUNAL DE CUENTAS MUNICIPALES	San Luis
SAN LUIS - TRIBUNAL DE CUENTAS PROVINCIALES	San Luis
SANTA CRUZ - TRIBUNAL DE CUENTAS PROVINCIALES	Santa Cruz
SANTA FE - TRIBUNAL DE CUENTAS MUNICIPALES	Santa Fe
SANTA FE - TRIBUNAL DE CUENTAS PROVINCIALES	Santa Fe
SANTIAGO DEL ESTERO - TRIBUNAL DE CUENTAS MUNICIPALES	Santiago del Estero
SANTIAGO DEL ESTERO - TRIBUNAL DE CUENTAS PROVINCIALES	Santiago del Estero

TIERRA DEL FUEGO - TRIBUNAL DE CUENTAS PROVINCIALES	Tierra del Fuego
TRELEW - TRIBUNAL DE CUENTAS MUNICIPALES	Chubut
TUCUMAN - TRIBUNAL DE CUENTAS PROVINCIALES	Tucumán
TUCUMAN - DIRECCIÓN GENERAL DE AUDITORÍA	Tucumán

Tabla 9: OCLs incluidos en la Red Federal

2.2.1.2.5 – Muestra de Programas

Para tener perfectamente definida la captación de auditorías, es necesario además de los parámetros anteriormente tratados, considerar una muestra de programas.

Un programa está compuesto por Políticas Sociales de Estado, y lógicamente, tiene asignado auditorías, que es la variable central de la captación de datos. En una dimensión superior, el hecho de captar las auditorías de los programas por OCLs en cada Ministerio, permite caracterizar mejor el impacto social, y da a pie a analizar el cumplimiento de las políticas. A su vez, el intervalo temporal fijado para la captación, también permite caracterizar mejor el cumplimiento en dicho período. Pero cabe tener en cuenta que, tras las distintas dimensiones presentadas, está la muestra de Programas que es el objeto directamente relacionado con las auditorías.

A continuación se detallan los programas incluidos en cada año:

Programas Año 2011
Plan Nacional de Desarrollo Local y Economía Social - Manos a la Obra
Plan Nacional de Seguridad Alimentaria
Plan Nacional Familias - Asistencia Directa a Organizaciones
Desarrollo e Innovación Tecnológica
Fondo Nacional de Incentivo Docente
Gestión Educativa
Infraestructura y Equipamiento
Mejoramiento de la Educación Rural
Políticas Socio Educativas
Programa Federal de Construcción de Viviendas Techo Digno
Atención de la Madre y el Niño – Plan Nacer y Plan Nacional a favor de la Madre y el

Programas Año 2011
Niño
Atención Médica a Beneficiarios de Pensiones No Contributivas (PROFE)
Atención Primaria de la Salud
Desarrollo de Estrategias en Salud Social y Comunitaria – Médicos Comunitarios
Fortalecimiento de la Capacidad del Sistema Público de Salud (ex REMEDIAR)
Prevención y Control de Enfermedades Endémicas
Programa de Lucha contra el SIDA y Enfermedades de Transmisión Sexual
Programa de Prevención y Control de Enfermedades y Riesgos Específicos – Vacunaciones
Sanidad Escolar
Acciones de Capacitación Laboral
Acciones de Empleo

Tabla 10: Programas año 2011

Programas Año 2012
Programa Social Agropecuario (PSA)
Plan Nacional de Desarrollo Local y Economía Social - Manos a la Obra
Plan Nacional de Seguridad Alimentaria
Plan Nacional Familias – Asistencia Directa a Organizaciones
Fondo Nacional de Incentivo Docente (FONID)
Gestión Educativa – Plan de Finalización de Estudios Primarios y Secundarios (FINES)
Gestión Educativa – Programa Nacional de Inclusión Educativa (PIIE)
Políticas Socioeducativas – Coros y Orquestas
Programa de Mejoramiento de la Educación Rural (PROMER)
Programa Nacional de Infraestructura y Equipamiento
Programa Nacional para la Educación Técnico Profesional (INET)
Desarrollo Infraestructura Habitacional Techo Digno
Lucha contra el SIDA y ETS
Prevención y Control de Enfermedades Endémicas
Programa Federal de Salud - PROFE
Programa Nacional de Control de Enfermedades Inmunoprevenibles (Inmunizaciones)
Provisión de Medicamentos Genéricos para APS (ex-REMEDIAR)
Seguro Universal de Maternidad e Infancia - Plan Nacer
Acciones de Promoción del Empleo
Actividades de Capacitación y Formación Profesional

Tabla 11: Programas año 2012

Programas Año 2013
Programa Social Agropecuario (PSA)
Plan Nacional de Desarrollo Local y Economía Social - Manos a la Obra
Plan Nacional de Seguridad Alimentaria
Plan Nacional Familias - Asistencia Directa a Organizaciones
Programa Ingreso Social con Trabajo - Argentina Trabaja
Fondo Nacional de Incentivo Docente (FONID)
Fondo Nacional para la educación Técnico Profesional (INET) - Innov. y Desar. de la Formac. Tecnol.
Gestión Educativa - Plan de Finalización de Est. Primarios y Secundarios (FINES)
Gestión Educativa - Programa Nacional de Inclusión Educativa (PIIE)
Integral Escuelas - Sistema de Transferencias de Recursos Educativos (SITRARED)

Programas Año 2013
Programa de Mejoramiento de la Educación Rural (PROMER)
Programa Nacional de Infraestructura y Equipamiento
Desarrollo Infraestructura Habitacional Techo Digno
Lucha contra el SIDA y ETS
Prevención y Control de Enfermedades Endémicas
Programa Incluir Salud - (ex- PROFE)
Programa Nacional de Control de Enfermedades Inmunoprevenibles (Inmunizaciones)
Provisión de Medicamentos Genéricos para APS (ex-REMEDIAR)
Seguro Universal de Maternidad e Infancia - Plan Nacer
Acciones de Promoción del Empleo
Actividades de Capacitación y Formación Profesional

Tabla 12: Programas año 2013

Programas Año 2014
Plan de Políticas Integrales de Acción para Personas con Discapacidad
Programa de Desarrollo de las Economías Regionales
Programa de Servicios Agrícolas Provinciales (PROSAP)
Plan Nacional de Desarrollo Local y Economía Social - Manos a la Obra
Plan Nacional de Seguridad Alimentaria
Plan Nacional Familias - Asistencia Directa a Organizaciones
Programa Ingreso Social con Trabajo - Argentina Trabaja
Fondo Nacional de Incentivo Docente (FONID)
Fondo Nacional para la Educación Técnico Profesional (INET) - Innov. y Des. de la Form. Tecnológica
Integral Escuelas - Sistema de Transferencias de Recursos Educativos (SITRARED)
Desarrollo Infraestructura Habitacional Techo Digno
Lucha contra el SIDA y Enfermedades de Transmisión Sexual
Prevención y Control de Enfermedades Transmisibles por Vectores
Prevención y Control de Enfermedades y Riesgos Específicos
Programa Atención Primaria de la Salud - Desarrollo de la Salud Sexual y la Procreación Responsable
Programa Desarrollo de las Estrategias en Salud Familiar y Comunitaria - Médicos de Cabecera
Programa Federal Incluir Salud - (ex-PROFE)
Programa Funciones Esenciales de Salud Pública
Programa Nacional de Desarrollo de Seguros Públicos de Salud - Subprograma SUMAR (Plan Nacer)
Programa Reforma del Sector Salud - Subprogr. Prov. de Medicamentos Genéricos para APS (ex-REMEDIAR)
Acciones de Promoción del Empleo
Actividades de Capacitación y Formación Profesional

Tabla 13: Programas año 2014

2.2.1.2.6 – Las auditorías observadas, tipos de auditorías, estados de auditorías y la cantidad de horas observadas

Tomando en consideración los 4 años, los 84 programas pertenecientes a 7 Unidades de Auditoría Interna y los 41 Organismos de Control Local incluidos en las respectivas muestras, resultan a priori una total de 1029 observaciones de auditorías. Este es el número de auditorías que se podrán captar con la estructura del estudio que se ha planteado sin contar las auditorías No Planificadas.

Las siguientes tablas demuestran la cantidad de auditorías planificadas de muestras captadas por año por unidad de auditoría interna (UAI):

Unidad de Auditoría Interna	Total de Auditorías 2011
MINISTERIO DE DESARROLLO SOCIAL	54
MINISTERIO DE EDUCACION	60
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS	20
MINISTERIO DE SALUD	72
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL	39
TOTAL DE AUDITORÍAS	245

Tabla 14: Total de auditorías año 2011

Unidad de Auditoría Interna	Total de Auditorías 2012
MINISTERIO DE AGRICULTURA, GANADERIA Y PESCA	14
MINISTERIO DE DESARROLLO SOCIAL	49
MINISTERIO DE EDUCACION	65
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS	17
MINISTERIO DE SALUD	58
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL	45
TOTAL DE AUDITORÍAS	248

Tabla 15: Total de auditorías año 2012

Unidad de Auditoría Interna	Total de Auditorías 2013
MINISTERIO DE AGRICULTURA, GANADERIA Y PESCA	12
MINISTERIO DE DESARROLLO SOCIAL	49
MINISTERIO DE EDUCACION	72
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS	20
MINISTERIO DE SALUD	72

MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL	31
TOTAL DE AUDITORÍAS	256

Tabla 16: Total de auditorías año 2013

Unidad de Auditoría Interna	Total de Auditorías 2014
CONSEJO NACIONAL DE COORDINACION DE POLITICAS SOCIALES	9
MINISTERIO DE AGRICULTURA, GANADERIA Y PESCA	20
MINISTERIO DE DESARROLLO SOCIAL	56
MINISTERIO DE EDUCACION	52
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS	21
MINISTERIO DE SALUD	86
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL	36
TOTAL DE AUDITORÍAS	280

Tabla 17: Total de auditorías año 2014

Las auditorías podrán calificarse en según su tipo y su estado.

Los tipos de auditoría hacen referencia a si fue tomada en cuenta su ejecución en la planificación anual del OCL correspondiente, y se puede dividir en dos grupos:

- **Planificadas:** Aquellas auditorías que fueron tomadas en cuenta en la planificación anual del OCL.
- **No Planificadas:** Aquellas auditorías que no fueron tomadas en cuenta en la planificación anual del OCL, pero por alguna razón es necesaria su ejecución en el transcurso del año.

Con el estado de la auditoría se puede observar en qué situación se encuentra una auditoría en un momento determinado en la línea del tiempo, y a su vez el resultado de la misma. Los estados son los siguientes:

- **No iniciada:** La auditoría fue incluida en la planificación anual pero todavía no ha sido estudiada su ejecución.
- **Iniciada:** Capta las auditorías desde la planificación de su ejecución hasta el instante antes del informe de su finalización.

- **Finalizada – sin observaciones:** La auditoría ha sido finalizada y no se ha presentado un informe de irregularidades en el programa auditado.
- **Finalizada – con observaciones:** La auditoría ha sido finalizada y se ha presentado un informe de irregularidades en el programa auditado.
- **Solicitud de Baja:** El OCL por algún motivo ha solicitado suspender la ejecución de la auditoría y está en espera de la respuesta de los agentes responsables para la anulación.
- **Anulada:** Se ha llevado a cabo la solicitud de baja de la auditoría por el OCL y la misma se encuentra suspendida sin posibilidad de poder seguir adelante para su ejecución.

Cada auditoría posee una cantidad de horas de observación, la cual representa la duración de la ejecución de la auditoría expresada en horas. Dicho valor también es objeto de medición y representa la segunda métrica del cubo junto a la cantidad de auditorías.

2.2.1.3 – Diseño de la solución

En los últimos años, la cantidad de auditorías se ha incrementado considerablemente. En el 2014 el total de auditorías fue de 280. Lo cual representa un aumento del 14% con respecto al del año 2011 que fue de 245. Esto ha originado la necesidad de dotar a la Sindicatura General de la Nación con herramientas de software que faciliten no solo la operación, sino su control y seguimiento por los niveles estratégicos.

Sin embargo y a pesar de contar con sistemas de información que automatizan en gran medida la administración de las auditorías y que permiten a las áreas usuarias ejecutar en forma precisa y oportuna el seguimiento, los ejecutivos necesitan herramientas de software que les permita conocer en forma rápida datos puntuales como avances de las auditorías, grado de cumplimiento de las mismas, presupuestos asignados, entre otros.

Por lo tanto, proporcionar una herramienta de software a las áreas estratégicas de la Sindicatura General de la Nación, que cubra los requerimientos de información necesarios para tomar decisiones, es un proyecto que puede propiciar un mejor aprovechamiento de los recursos disponibles y una mayor transparencia en el uso de los mismos, dos aspectos que por su importancia pueden resultar de gran impacto para los diferentes niveles ejecutivos.

En este apartado se habla sobre el diseño del data warehouse para la SIGEN, que busca satisfacer la necesidad de información arriba descrita. Se hace una breve descripción del problema a solucionar y se mencionan los objetivos.

También se mencionan cuáles fueron las necesidades de los usuarios que dieron origen al proyecto y los requerimientos de la solución que se detectaron durante el desarrollo de las entrevistas llevadas a cabo con el personal de mando de la SIGEN.

Se habla también sobre la descripción funcional y se presenta el diseño del data warehouse de la solución, así como la estructura dimensional con la cual se implementó.

2.2.1.3.1 – Planeamiento del problema

La enorme cantidad de información que manejan los sistemas transaccionales u operacionales y que se almacena en bases de datos relacionales, no se está aprovechando al máximo en la toma de decisiones a niveles estratégicos de la Sindicatura General de la Nación. La normatividad vigente y los sistemas transaccionales en operación (mismos que están diseñados en apego a la normatividad), establecen los mecanismos de control necesarios para evitar un mal manejo de los recursos disponibles.

La información viaja desde los niveles operativos hasta los niveles tácticos y estratégicos a través de los canales de comunicación establecidos. Pero hoy en día, los ejecutivos, requieren de

herramientas ágiles que les permita visualizar, en forma clara y oportuna, métricas y parámetros de control con base en los cuales puedan tomar decisiones rápidamente. La preocupación principal se centra en la necesidad de contar con información oportuna disponible sobre la inversión en proyectos.

Por lo tanto, se tiene la necesidad de desarrollar una herramienta de software para los niveles estratégicos de la Sindicatura General de la Nación, que les proporcione información suficiente y oportuna para tomar decisiones con respecto a las inversiones sobre proyectos que se desarrollan a lo largo de todo el país, para conocer de manera precisa y en cualquier momento el avance de auditorías sobre los mismos y así poder encontrar irregularidades y desviaciones de la planificación.

Actualmente, los sistemas de información institucionales proporcionan datos útiles y confiables para la operación cotidiana de la SIGEN. Esto involucra a los niveles táctico y operativo. Sin embargo, en muchas ocasiones el personal de las áreas estratégicas debe solicitar información a diversas áreas y después de ello, un grupo de personas integra toda esa información y la ordena para presentar los informes ejecutivos necesarios para la oficina del Síndico y para la Presidencia de la Nación.

En resumen, el requerimiento de información que se presenta, responde a la necesidad de conocer en forma rápida y precisa el presupuesto asignado, el compromiso y avance del proyecto y el cumplimiento de las tareas.

2.2.1.3.2 – Objetivos del proyecto

Objetivo general

El objetivo general de este trabajo es desarrollar una herramienta de software usando tecnología de inteligencia de negocios, enfocada al seguimiento de proyectos comprendidos en el plan nacional, que cuente con las características de confiabilidad y oportunidad bajo los estándares tecnológicos adoptados por la Sindicatura General de la Nación, para el apoyo a la toma de decisiones en los niveles estratégicos.

Objetivos específicos.

Los objetivos específicos de este proyecto son los siguientes:

- Identificar la información y las fuentes de datos que utiliza el personal de las áreas de auditorías de proyectos dándole un formato uniforme para la entrada del data warehouse.
- Desarrollar una base de datos dimensional para la consolidación de la información histórica a nivel nacional.
- Desarrollar reportes personalizables que los usuarios utilicen para la elaboración de sus informes ejecutivos.

2.2.1.3.3 – Requerimientos de la solución

Derivado de las entrevistas realizadas al personal que componen las áreas de la Red Federal en la Sindicatura General de la Nación, se detectaron las siguientes necesidades de información.

1. **Contar con una única fuente de información de auditorías. La integración de datos deberá ser mucho más sencilla y rápida.** La integración de la información de auditorías entre los diferentes organismos locales se realiza en forma periódica y puede ser mensual, trimestral, semestral y anual. Anualmente, a cada organismo local se le asigna una

cantidad de auditorías, el cual ejerce de manera independiente y es responsable de su cumplimiento. Sin embargo, la SIGEN en coordinación con las UAIs, analizan los estados de las auditorías en forma periódica con el fin de encontrar tendencias, determinar áreas de oportunidad y en general, verificar si lo que se planeó al inicio de un ejercicio es congruente contra la ejecución de las actividades de los organismos locales. Al tener fuentes de información distintas, es complicado determinar una sola verdad de los datos y por lo tanto, la SIGEN demora demasiado en consolidar la información de todos los organismos.

2. **Reducir los tiempos de respuesta para que los usuarios de la alta gerencia de la organización, cuenten con reportes confiables de los proyectos cuando ellos los necesiten.** La línea de mando en la Sindicatura General de la Nación es jerárquica, por lo que cuando el ejecutivo requiere reportes de avances de los proyectos, esta solicitud puede tardar desde un par de horas hasta algunos días de acuerdo a las cargas de trabajo y disponibilidad del personal para atender el requerimiento. El inconveniente con esta situación es cuando los ejecutivos necesitan información confiable para tomar una decisión y no cuentan con ésta en el momento en el que se necesita, se pueden originar diferentes problemas como atrasos en la ejecución de obras públicas, mala inversión, no contar a tiempo con los datos necesarios para la planeación de actividades en la ejecución de un trabajo, entre otros. Para resolver esta situación, es necesario controlar la información, por medio de herramientas confiables, a los niveles ejecutivos de la Sindicatura General de la Nación, con el fin de que dispongan de todos los datos necesarios para la toma de decisiones en el momento en el que los requieran.
3. **Integrar la información de las auditorías de los diferentes organismos locales en forma rápida y sin depender de la operación de las áreas administrativas de la Red Federal.** Es decir, la herramienta que se elabore debe garantizar la integridad de la información con el fin de evitar lo más posible la intervención de los responsables que la generaron. De esta manera se pretende liberar a los ejecutivos de tiempo muy valioso que

pueden dedicar al desarrollo de las actividades sustantivas de sus áreas, en lugar de dedicar tiempo de revisión para realizar una validación de datos.

4. **Integrar información de auditorías para establecer parámetros de control y seguimiento con la información de otros organismos.** Una de las ventajas del uso de aplicaciones para la inteligencia de negocios es precisamente el poder definir parámetros de medición. Por esta razón, es necesario que la solución propuesta permita dar seguimiento a los siguientes parámetros:

- Cantidad de auditorías ejecutadas contra cantidad de auditorías planificadas.
- Cantidad de horas ejecutadas contra cantidad de horas planificadas.
- Auditorías ejecutadas y planificadas por organismo de control local.
- Horas ejecutadas y planificadas por organismo de control local.
- Auditorías ejecutadas y planificadas por ministerio.
- Horas ejecutadas y planificadas por ministerio.
- Auditorías ejecutadas y planificadas por proyecto.
- Horas ejecutadas y planificadas por proyecto.
- Auditorías ejecutadas y planificadas por provincia.
- Horas ejecutadas y planificadas por provincia.
- Presupuesto por proyectos.
- Presupuesto por ministerio.

5. **Integrar información histórica de la operación anual de la Sindicatura General de la Nación. La información histórica de la Sindicatura General de la Nación permite**

evaluar cuales fueron los resultados obtenidos por año en comparación con el planeamiento anual. También deberá permitir dar seguimiento a parámetros como:

- Cantidad de auditorías ejecutadas contra cantidad de auditorías planificadas anual.
- Cantidad de horas ejecutadas contra cantidad de horas planificadas anual.
- Auditorías ejecutadas y planificadas por organismo de control local anual.
- Horas ejecutadas y planificadas por organismo de control local anual.
- Auditorías ejecutadas y planificadas por ministerio anual.
- Horas ejecutadas y planificadas por ministerio anual.
- Auditorías ejecutadas y planificadas por proyecto anual.
- Horas ejecutadas y planificadas por proyecto anual.
- Auditorías ejecutadas y planificadas por provincia anual.
- Horas ejecutadas y planificadas por provincia anual.
- Presupuesto por proyectos anual.
- Presupuesto por ministerio anual.

6. **Conocer en cualquier momento, el grado de cumplimiento de las auditorías y la cantidad de irregularidades encontradas en el proyecto.** El seguimiento de las auditorías es otra de las preocupaciones de la alta gerencia. Identificar en cualquier momento las irregularidades en los proyectos en los que participan los distintos ministerios nacionales tiene como objetivo poder evaluar el avance de las actividades, el cumplimiento de las mismas y una buena asignación presupuestaria.

- 7. Simplificar la consolidación de la información de todas las áreas necesarias de la Sindicatura General de la Nación y los organismos de control local, así como la generación de la amplia variedad de reportes que se necesitan elaborar.**

2.2.1.3.4 – Descripción funcional de la solución

Actualmente, se encuentran en funcionamiento diversos sistemas institucionales que soportan la operación diaria de la Sindicatura General de la Nación. Dos de ellos son el Sistema de Entes y el Sistema de Sisio Red Federal.

El Sistema de Sisio Red Federal es el sistema mediante el cual se realiza la gestión de todo lo referente al seguimiento de las auditorías. El sistema cubre desde la elaboración de las matrices de planificación de cada Órgano de Control Local hasta la incorporación de los informes de auditoría. El dueño normativo del Sistema de Sisio Red Federal es el área de Red Federal de la Sindicatura General de la Nación.

Por otro lado, el Sistema de Entes es el sistema de información de apoyo a la Sindicatura General de la Nación la cual cubre la carga y obtención de información en detalle de cada Ente perteneciente a la Nación Argentina y sus relaciones. Por medio de este sistema se puede obtener y entender la estructura institucional de todo el Estado.

Ambos Sistemas se encuentran en operación y sus bases de datos almacenan una gran cantidad de información. El modelo de referencia que se muestra en la ilustración 14, es un bosquejo en el cual se ejemplifica el funcionamiento de la solución a desarrollar.

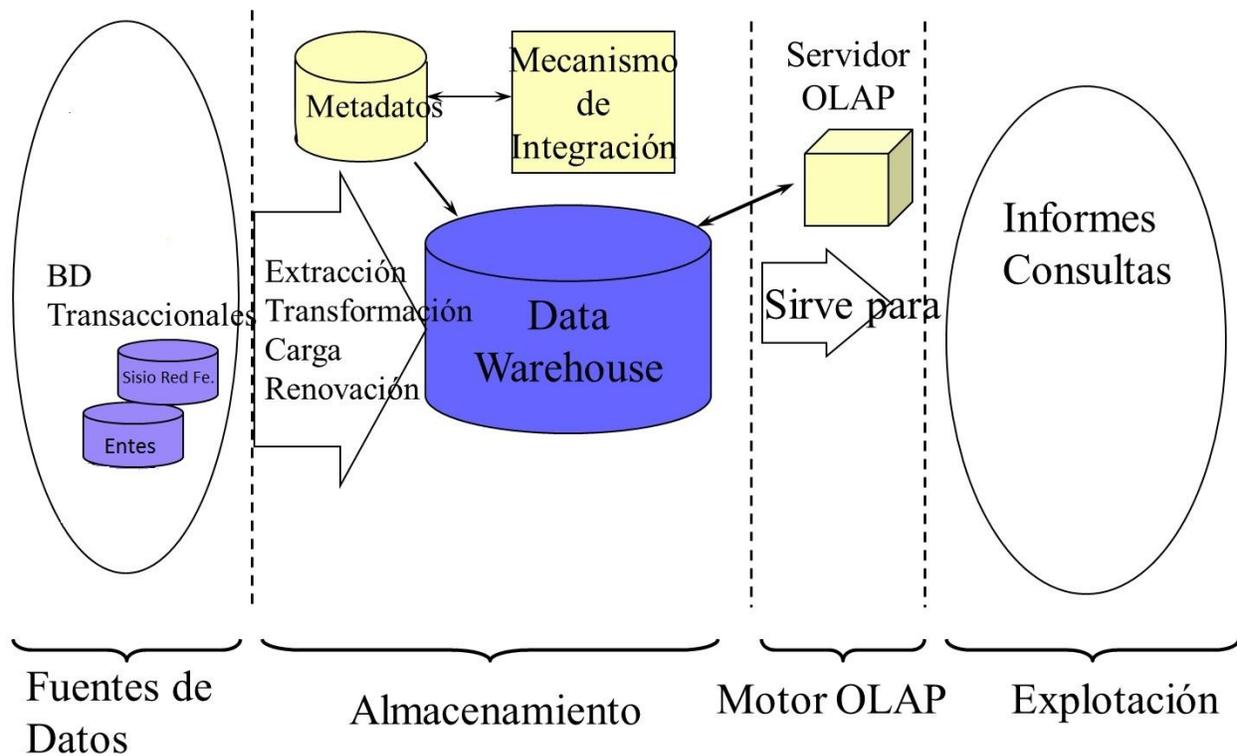


Ilustración 14 - Modelo de inteligencia de negocios propuesto para la solución

Como ya se ha comentado en este y otros apartados, se utilizarán las bases de datos del Sistema de Entes y el Sistema de Sisio Red Federal para extraer la información necesaria que se depositará en el data warehouse. La extracción de información de estos sistemas se hará por medio de procesos por lotes (archivos bat) que se conectarán a las fuentes de datos (bases de datos relacionales) y que extraerán la información y la dejarán en archivos planos (archivos de texto).

Adicional a la extracción de datos y previo a la carga de información al data warehouse, será necesario transformar la información en la estructura dimensional requerida para el almacenamiento de los datos. Para ello, se usará el software O3, el cual preparará al data warehouse para su carga.

Una vez preparada la estructura dimensional del data warehouse, se ejecutará un proceso de carga de los datos extraídos hacia los archivos planos. Por lo que al final de la carga de los datos, se

ejecutará un proceso que permitirá consolidar la información a los diferentes niveles del data warehouse.

Después de realizar el proceso de extracción, transformación y consolidación de datos hacia el data warehouse, el usuario podrá consultar su información por medio de herramientas de reportes.

Se proporcionará un conjunto de reportes y gráficos con indicadores estratégicos predefinidos por los usuarios que, en la actualidad, integran usando diferentes fuentes de datos. Esto último representa para el usuario un trabajo artesanal, ya que tiene que reunir datos de los sistemas en operación, de sus controles internos, de reportes de otras áreas y de trabajos de campo para preparar informes y otros documentos oficiales.

2.2.1.3.5 – Modelo Dimensional

En el desarrollo de un sistema de información basado en el modelo relacional, se utilizarían diagramas basados en la notación del modelo entidad-relación para identificar y diseñar la relación entre los datos de la organización. Para diseñar un data warehouse, se utiliza un modelo dimensional. Pero al igual que en el modelo entidad-relación, se debe partir del análisis de las necesidades de información y requerimientos del usuario.

En este proyecto, se utilizó el modelo de estrella para diseñar el repositorio de datos. El modelo lógico resultante contiene una sola tabla de hechos y un conjunto de tablas de dimensión. En las siguientes ilustraciones pueden observarse el diagrama de estrella obtenido especificando sus medidas con granularidad y el detalle de la estructura de cada dimensión.

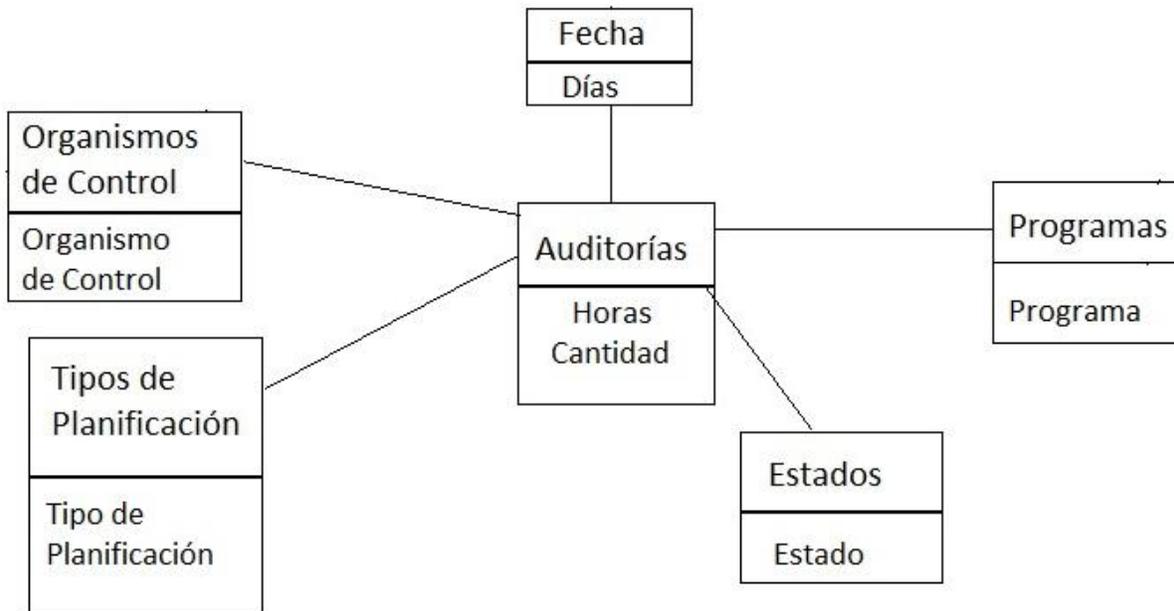


Ilustración 15 – Modelo Dimensional de la Solución



Ilustración 16 - Dimensiones con Estructura

En la siguiente tabla se describe cada una de las entidades de la ilustración:

Tabla	Atributos	Descripción
Auditorías	IdAuditoria IdPrograma IdTipoPlanificacion IdEstadoAuditoria IdOCL CantidadHoras Cantidad Fecha	Es la tabla de hechos. Contiene los atributos relaciones a las tablas de dimensiones y dos atributos medidas: Cantidad de Auditorías y Cantidad de Horas.
Fecha	Fecha	Dimensión de tiempo. Contiene el detalle diario de las operaciones y los acumulados mensual y anual.
Provincias	IdProvincia Provincia	Ubicación física del Organismo de Control
Tipos de Planificación	IdTipoPlanificacion TipoPlanificacion	Esta dimensión se utiliza para clasificar según el tipo de planificación de la auditoría. Puede ser planificada o no planificada.
Estados	IdEstadoAuditoria EstadoAuditoria	Esta dimensión se utiliza para clasificar según el estado en el que se encuentra una auditoría.
Programas	IdPrograma Programa IdMinisterio IdPresupuesto	Programa al que pertenece la auditoría. Hace referencia al programa a auditar por el Organismo de Control.
Ministerios	IdMinisterio Ministerio	Hace referencia al Ministerio que contiene el programa a auditar.
Niveles de	IdNivelPresupuesto	Es el nivel de presupuesto que posee la

Presupuesto	NivelPresupuesto	ejecución de un programa. Hace referencia al gasto destinado por el Gobierno para poner en marcha dicho programa.
Organismos de Control	IdOCL OCL IdProvincia	Son los encargados de planificar y llevar a cabo la auditoría.

Tabla 18 - Detalle del Modelo Dimensional de la Solución

2.2.3.6 - Descripción de los procesos desarrollados

El desarrollo de un data warehouse parte del diseño del repositorio de datos, el cual puede ser implementado utilizando una base de datos relacional. A continuación se detallan los procesos desarrollados para la construcción del data warehouse de la solución.

Extracción de datos

Una ventaja que se tuvo en el desarrollo de la tecnología necesaria para conectar las dos bases de datos, Sisio Red Federal y Entes, fue que las dos fuentes de datos a utilizar están desarrolladas sobre la misma plataforma de base de datos, SQL Server 2000. Por lo cual, no fue necesario desarrollar diferentes programas de extracción para diferentes ambientes como llega a ocurrir cuando se trata de implementar un data warehouse.

Los programas de extracción fueron desarrollados con Visual Basic .NET. Estos programas extraen la información de las fuentes de datos por medio de la ejecución de consultas y la guardan temporalmente en archivos planos o archivos de texto como registros con campos delimitados por el símbolo pipe (|), para que posteriormente mediante el software O3, se transformen e integren la información en el data warehouse.

Una vez que este programa de extracción de información es ejecutado, genera un archivo de texto con los datos indicados arriba. En la siguiente ilustración se muestra una parte del archivo generado después de la corrida del programa.

```
Organismos_Control.txt: Bloc de notas
Archivo Edición Formato Ver Ayuda
IdOCL|OCL|IdProvincia
1960|CABA - AUDITORIA GENERAL|CAP
2052|CABA - SINDICATURA GENERAL|CAP
3218|ROSARIO - TRIBUNAL DE CUENTAS MUNICIPAL|SFE
3219|SANTA FE - TRIBUNAL DE CUENTAS MUNICIPAL|SFE
3220|SANTIAGO DEL ESTERO - TRIBUNAL DE CUENTAS MUNICIPAL|SGO
3221|LA RIOJA - TRIBUNAL DE CUENTAS MUNICIPAL|LRJ
3222|SAN LUIS - TRIBUNAL DE CUENTAS MUNICIPAL|SLU
3223|SALTA - TRIBUNAL DE CUENTAS MUNICIPAL|SAL
3240|SAN CARLOS DE BARILOCHE - TRIBUNAL DE CUENTAS MUNICIPAL|RNO
3241|LA BANDA - TRIBUNAL DE CUENTAS MUNICIPAL|SGO
3242|SAN FRANCISCO - TRIBUNAL DE CUENTAS MUNICIPAL|CBA
3243|TRELEW - TRIBUNAL DE CUENTAS MUNICIPAL|CHU
3351|SALTA - SINDICATURA GENERAL|SAL
3723|SANTA FE - SINDICATURA GENERAL|SFE
4118|RIO GRANDE - TRIBUNAL DE CUENTAS MUNICIPAL|TDF
4480|GOBERNADOR INGENIERO VALENTIN VIRASORO - TRIBUNAL DE CUENTAS MUNICIPAL|COR
2053|LA PAMPA - TRIBUNAL DE CUENTAS|LAP
2054|RIO NEGRO - TRIBUNAL DE CUENTAS|RNO
2055|ENTRE RIOS - TRIBUNAL DE CUENTAS|ERI
2056|MISIONES - TRIBUNAL DE CUENTAS|MIS
2057|SALTA - AUDITORIA GENERAL|SAL
2393|CATAMARCA - TRIBUNAL DE CUENTAS|CAT
2394|CHUBUT - TRIBUNAL DE CUENTAS|CHU
2395|SANTA CRUZ - TRIBUNAL DE CUENTAS|SCR
2396|SANTIAGO DEL ESTERO - TRIBUNAL DE CUENTAS|SGO
2397|SAN JUAN - TRIBUNAL DE CUENTAS|SJU
2398|MENDOZA - TRIBUNAL DE CUENTAS|MZA
2399|LA RIOJA - TRIBUNAL DE CUENTAS|LRJ
2400|FORMOSA - TRIBUNAL DE CUENTAS|FOR
2401|CORDOBA - TRIBUNAL DE CUENTAS|CBA
2407|TUCUMAN - TRIBUNAL DE CUENTAS|TUC
2409|CHACO - TRIBUNAL DE CUENTAS|CHA
2410|BUENOS AIRES - TRIBUNAL DE CUENTAS|BUE
2411|TIERRA DEL FUEGO - TRIBUNAL DE CUENTAS|TDF
2412|SANTA FE - TRIBUNAL DE CUENTAS|SFE
2413|SAN LUIS - TRIBUNAL DE CUENTAS|SLU
2414|NEUQUEN - TRIBUNAL DE CUENTAS|NEU
2416|JUJUY - TRIBUNAL DE CUENTAS|JUJ
2417|CORRIENTES - TRIBUNAL DE CUENTAS|COR
3208|TUCUMAN - DIRECCION GENERAL DE AUDITORIA|TUC
```

Ilustración 17 - Muestra extracción de datos de las bases de datos relacionales

Integración y transformación de los datos

Esta fase consiste en integrar la información obtenida en el paso anterior, hacia un repositorio de datos coherente. En este trabajo se utilizará como repositorio de la solución una base de datos dimensional.

Es importante tener en cuenta que antes de iniciar con el proceso de integración de datos, es necesario preparar la estructura dimensional de la base de datos, a fin de que la importación de la información desde las fuentes de datos sea transparente y libre de errores.

Para llevar a cabo la integración de datos, se elaboraron dentro del programa O3 lo que se denominan reglas de carga, los cuales permiten definir la manera en la que se importará la información. Por medio de reglas de carga, pueden agregarse valores a la base de datos dimensional desarrollada. Pero si los datos no se encuentran en el formato correcto, entonces se necesita desarrollar reglas de carga especiales para esos valores.

Las reglas de carga son un conjunto de operaciones que se ejecutan sobre los datos para validarlos y transformarlos antes de ser integrados a la base de datos. Generalmente se crea una regla de carga por cada dimensión. Lo que hacen las reglas de carga es leer los datos a ser importados, los cambia con base en las reglas establecidas y los integra a la base de datos. Es importante tener en cuenta que las fuentes de datos no se modifican.

El proceso de transformación puede consistir en cualquier de las siguientes acciones o un conjunto de ellas:

- Ignorar campos o cadenas de la fuente de datos.
- Cambiar el orden de los campos moviendo, uniendo, dividiendo o creando nuevos campos a partir de los existentes en la fuente de datos.
- Mapeo de los datos desde el origen hacia la base de datos.
- Cambiar los valores de los datos provenientes de las fuentes de datos escalando o adicionando los valores existentes en la base de datos.
- Omitir o modificar valores nulos provenientes de la fuente de datos.

- Omitir registros inválidos para continuar con el proceso de integración de datos.
- Agregar nuevas dimensiones y miembros a la base de datos.
- Modificar dimensiones y miembros existentes en la base de datos.

Continuando con el ejemplo del archivo de texto de la Ilustración 17, describiré el proceso de construcción de la regla de carga que se deberá utilizar para transformar los datos antes de su integración a la base de datos. Cabe mencionar que este archivo contiene datos relacionados con los Organismos de Control Local de todo el país.

Para integrar la información de la fuente de datos (en este caso el archivo de texto), se debe especificar la relación uno a uno entre los campos que se encuentran en la fuente de datos contra las dimensiones en la base de datos. Esta relación se especifica en las reglas de carga y funciona cada vez que se requiera importar nuevos datos desde la misma fuente de datos, sin tener que modificar la estructura de la fuente ni la de la regla de carga. Esta asociación campo-dimensión es obligatoria y no se pueden integrar datos si no se especifica la relación de uno a uno entre el origen de los datos y la base de datos dimensional.

2.2.2 - Instrumentos de recolección y medición de datos

Como instrumento de medición se ha seleccionado la Observación. Consiste en el registro sistemático, válido y confiable de comportamiento o conducta manifiesta.

Se ha seleccionado este estilo de instrumento ya que son técnicas de medición no obstructivas, aceptan material no estructurado y además pueden trabajar con grandes volúmenes de datos.

Para la construcción de un sistema de observación se desarrollaron los siguientes pasos:

- **Definir el universo de aspectos, eventos o conductas a observar:** Se ha definido como universo el sistema de Red Federal de la Sindicatura general de la Nación con sus actividades sobre las auditorías nacionales.
- **Establecer y definir las unidades de observación:** Se establecen como unidades de observación principales la cantidad y horas de auditorías. Tomando en consideración los 4 años, los 84 programas pertenecientes a 7 Unidades de Auditoría Interna y los 41 Organismos de Control Local incluidos en las respectivas muestras, resultan a priori una total de 1029 observaciones de auditorías.
- **Establecer y definir las categorías de observación:** Se han establecido 2 calificaciones de categorías de observación para cada auditoría: tipo de auditoría y Estado de la auditoría. Las mismas son detalladas en el apartado 2.2.1.2.6 – Las auditorías observadas, tipos de auditorías, estados de auditorías y la cantidad de horas observadas.

Para el procesamiento de los datos de la observación se ha elegido un instrumento ya desarrollado y disponible, el cual se adapta a los requerimientos del estudio en particular. Dicho instrumento es la herramienta O3, permitiendo la integración de datos a través de un modelo dimensional previamente diseñado.

Para asegurar una buena medición se ha implementado el siguiente procedimiento:

- Listar las variables que se pretenden medir u observar.
- Revisar su definición conceptual y comprender su significado.
- Revisar cómo han sido definidas operacionalmente las variables.
- Elegir el instrumento o los instrumentos (ya desarrollados) que hayan sido favorecidos por la comparación y adaptarlos al contexto de la investigación.
- Indicar el nivel de medición de cada ítem y el de las variables.
- Indicar cómo se habrán de codificar los datos en cada ítem y variable.

2.2.3 - Confiabilidad y validez de la medición

La confiabilidad y validez de la medición se difiere de la contrastación de las Hipótesis de Investigación. La misma se realiza en torno a dos herramientas estadísticas muy comunes: La Prueba de Hipótesis y el análisis del Intervalo de Confianza.

2.2.3.1 – Pruebas de Hipótesis

Por un lado, se realizará sobre la Hipótesis de Investigación, una Prueba de Hipótesis que permita su demostración científica. Conviene recordar que la filosofía subyacente a las Pruebas de Hipótesis se basa en la formalización de un proceso decisional, aplicando los siguientes pasos:

1. Construir el contraste de Hipótesis, poniendo en la hipótesis nula (H_0) aquello que se desea rechazar.
2. Especificar la distribución del estadístico bajo la hipótesis nula (H_0).
3. Aceptar un riesgo α (alfa) de rechazar H_0 siendo ésta cierta.
4. Delimitar una región crítica de rechazo de H_0 con probabilidad α y una región de aceptación de H_0 con probabilidad $1-\alpha$. La probabilidad α se establece en 0,05, lo que supone asumir un margen de error máximo del 5%. La siguiente figura muestra esto gráficamente.

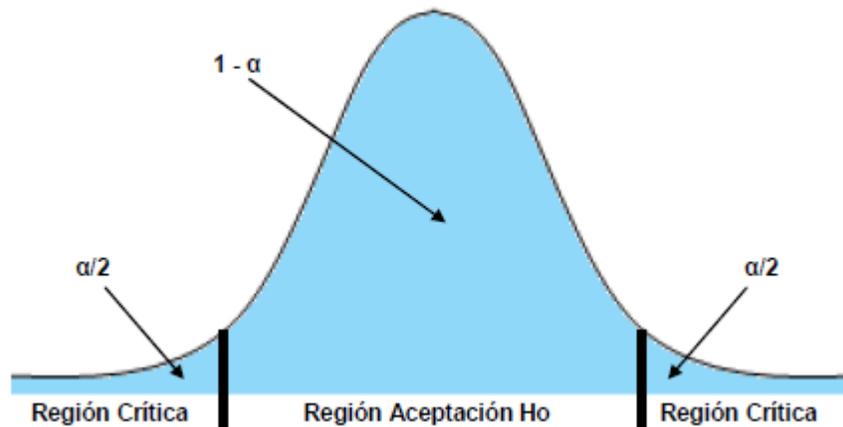


Ilustración 18: Planteamiento gráfico de la Prueba de Hipótesis

5. Observar en que región se sitúa el estadístico.
6. Concluir.

Del proceso presentado se deduce que la Prueba de Hipótesis es una herramienta de inferencia basada en plantear una pregunta cerrada, como por ejemplo, ¿son las horas de “auditorías finalizadas” igual a la de “no finalizadas”? La Prueba de Hipótesis se expresa en negativo, dado que el objetivo es rechazar la hipótesis nula.

2.2.3.2 - Intervalos de Confianza

Otra herramienta de inferencia útil se basa en los Intervalos de Confianza, los cuales se usan para plantear un pregunta abierta, como, ¿Qué valores el parámetros son ahora creíbles? Su análisis permite una interpretación en positivo, y más sutil que las Pruebas de Hipótesis (Cobo, 2003).

Así por ejemplo, cuando se rechaza la hipótesis nula en una Prueba de Hipótesis planteada bilateralmente, la conclusión formal es negar la igualdad de la nula, sin decantarse por ningún

lado. A efectos de este estudio, si se rechazase la hipótesis nula que supone la igualdad de horas de auditorías entre las “finalizadas” y “no finalizadas”, podría decirse que los estados tienen distintas horas, sin detallar en qué estado es mayor. Si además de la Prueba de Hipótesis se realiza el cálculo del Intervalo de Confianza, podrá estimarse los valores más probables de la media de cada estado, y consecuentemente, qué estado tendrá en general una cantidad de horas mayor.

En el caso en que no pueda demostrarse una Hipótesis de Investigación al no poder rechazarse la Hipótesis Nula planteada en la Prueba de Hipótesis correspondiente, el análisis basado en el Intervalo de Confianza facilitará deducir si esto se debe a una falta de potencia de la muestra y/o del método de inferencia, o bien se debe a que la realidad se comporta tal y como inducen a pensar los resultados.

2.2.4 - Métodos de análisis estadístico

Se determina que la escala de medición corresponde a las variables en estudio es nominal. La escala es categórica y las categorías son excluyentes, colectivamente exhaustivas y dicotómicas (no es posible ordenar las categorías). La medición es exclusivamente cualitativa.

Las categorías de las escalas son las siguientes:

- Planificada – No Planificada
- No Iniciada – Iniciada – Finalizada sin Observaciones – Finalizada con Observaciones – Solicitada Anulación – Anulada

Cabe destacar que para efectuar los cálculos necesarios para demostrar estadísticamente las Hipótesis de Investigación se ha utilizado el programa O3. En consecuencia, todos los gráficos y todas las tablas de resultados que se presentan en los dos capítulos siguientes, muestran los

resultados según calcula este programa. Asimismo, cabe destacar que en el próximo apartado se entrará en detalle de cuáles son los cálculos realizados.

2.2.4.1 - Distribuciones de frecuencias

Cuando los datos de una variable están dispersos, la dispersión sigue un cierto patrón. Inicialmente los datos no nos dicen nada por sí mismos, pero si los dividimos en clases o celdas ordenadamente, puede aclararse la forma de su dispersión, es decir, puede aclararse la forma de cómo están distribuidos. Esta forma de la distribución de los datos inherente a su variabilidad se denomina distribución de frecuencias.

Normalmente es posible ver la forma general de una distribución si se recogen cien o más valores y se prepara convenientemente una tabla de frecuencias con diez o veinte clases. Pero la distribución se puede ver aun con mayor claridad en forma de representación gráfica mediante un histograma de frecuencias. El histograma es una representación visual de los datos en la que pueden observarse más fácilmente tres propiedades esenciales de una distribución como son: Forma, tendencia central o acumulación y dispersión o variabilidad. De esta forma, el histograma da una idea del proceso, lo que un simple examen de los datos tabulados no hace.

Hay muchos métodos para construir histogramas. Cuando los datos son numerosos, es muy útil reunirlos en clases y se recomienda utilizar entre 4 y 20 clases (o celdas). A menudo conviene elegir un número total de clases igual aproximadamente a la raíz cuadrada del tamaño de la muestra. Las clases deben tener amplitud uniforme y se construye la primera de ellas comenzando con un límite inferior solo un poco menor que el valor más pequeño de los datos.

Se construye la última clase finalizando con un límite superior solo un poco mayor que el valor más grande de los datos. Para realizar el histograma se marcan las clases sobre el eje de abscisas,

y sobre cada clase se levanta un rectángulo de altura proporcional al número de observaciones de la variable (frecuencia absoluta) que caen en la clase.

El agrupamiento de los datos en clases condensa los datos originales, lo que da como resultado una pérdida de algo de detalle. Así, cuando el número de observaciones es relativamente pequeño, o cuando las observaciones solo toman pocos valores, puede construirse el histograma a partir de la distribución de frecuencias de los datos sin agrupar, dando lugar a los diagramas de barras.

Las distribuciones de frecuencias son la herramienta más sencilla y más utilizada y eficaz cuando estamos rodeados de montones de datos, que no nos dicen nada si no hacemos más que enumerarlos. Al expresar estos datos en forma de una distribución de frecuencias, ya nos proporcionan diversas ideas. Puesto que las distribuciones de frecuencias se utilizan muy a menudo en el control de calidad, es necesario conocer la finalidad de las mismas y su interpretación y uso.

Dada la importancia de las distribuciones de frecuencias, derivada de que en todo proceso hay un momento en el que nos encontramos con un conjunto de datos sobre las variables a tratar, es de gran importancia formalizar el proceso de recogida, ordenación y presentación de los datos que, en la mayoría de las ocasiones, aparecerán dispuestos en tablas de frecuencias de simple o doble entrada que servirán para analizar las distribuciones de las variables. Dada una variable X con valores x_1, x_2, \dots, x_N aparecen una serie de conceptos generales que se mencionan a continuación:

- Frecuencia absoluta n_i : Se denomina frecuencia absoluta del valor x_i de la variable X , el número de veces n_i que se repite ese valor.
- Frecuencia relativa f_i : Se denomina frecuencia relativa del valor x_i de la variable X la relación por cociente entre el número de veces que aparece el valor x_i y el número total de valores de la variable (N). O sea, $f_i = n_i / N$

- Frecuencia absoluta acumulada N_i Se denomina frecuencia absoluta acumulada del valor x_i a la suma de las frecuencias absolutas de los valores de la variable X anteriores o iguales a x_i . Su valor es $N_i = \sum_{j=1}^i n_j$.
- Frecuencia relativa acumulada F_i : Es la frecuencia absoluta acumulada dividida por el número total de valores de la variable. Su valor es $F_i = N_i / N$.

De todas estas definiciones se extraen las siguientes deducciones:

- La suma de las frecuencias absolutas sin acumular es igual al número total de elementos ($\sum n_j = N$)
- La última frecuencia relativa acumulada es el total de elementos (N).
- La suma de todas las frecuencias relativas acumuladas es igual a la unidad
- La última frecuencia relativa acumulada es la unidad

Al conjunto de valores que ha tomado una variable, junto con sus frecuencias, se le denomina distribución de frecuencias de la característica o variable. Para que una distribución de frecuencias quede determinada es necesario conocer todos los valores de la variable y uno cualquiera de los conceptos de frecuencia que acabamos de definir, ya que el paso de uno a otro es inmediato. Además, según la forma en que se presenten los valores de la variable será posible distinguir dos tipos de distribuciones de frecuencias:

- Las que no están agrupadas en intervalos, que surgen cuando la información se dispone asociando a cada valor o categoría de la variable su frecuencia.
- Aquellas cuyos valores observados generalmente aparecen agrupados en intervalos o clases $[L_i, L_{i-1}]$ debido al elevado número de observaciones, y, por tanto, las frecuencias correspondientes a cada intervalo se obtienen sumando las de los respectivos valores de la variable que contiene.

Cuando se trabaja con distribuciones agrupadas por intervalos o clases es necesario que las frecuencias observadas se asignen de alguna forma a los puntos del intervalo. Se podrá optar por suponer que los valores del intervalo se distribuyen uniformemente a lo largo de él o por considerar como representativo de todos los puntos del intervalo un único valor, por ejemplo, el punto medio del mismo, que denominaremos marca de clase (X_i) y que, en consecuencia, se obtendrá mediante $X_i = (L_{i-1} + L_i)/2$. Aunque la agrupación de valores tiene la ventaja de simplificar el manejo de la información, presenta en cambio un importante inconveniente consistente en la pérdida, en mayor o menor medida, de una parte de dicha información. La distribución de frecuencias de una variable suele presentarse ordenadamente mediante la tabla de frecuencias siguiente:

I_i	X_i	n_i	f_i	N_i	F_i
$[L_0, L_1]$	x_1	n_1	$f_1 = n_1/N$	$N_1 = n_1$	$F_1 = N_1/N$
$[L_1, L_2]$	x_2	n_2	$f_2 = n_2/N$	$N_2 = n_1 + n_2$	$F_2 = N_2/N$
$[L_2, L_3]$	x_3	n_3	$f_3 = n_3/N$	$N_3 = n_1 + n_2 + n_3$	$F_3 = N_3/N$
$[L_{k-1}, L_k]$	x_k	n_k	$f_k = n_k/N$	$N_k = n_1 + \dots + n_k = N$	$F_k = N_k/N = 1$

$$\sum n_i = N \quad \sum f_i = 1$$

Tabla 19: Tabla de Frecuencias

En cuanto al número de intervalos k a considerar puede tenerse en cuenta la fórmula de Sturges:

$$(K = E[3/2 + \log(N)/\log(2)]), \text{ o también tomar } K = \sqrt{N}.$$

3-RESULTADOS

3.1. – Modelo Resultante

La siguiente ilustración muestra la ventana de construcción del modelo generado mediante el software O3 por el cual se generó el cubo. Aquí se definieron las dimensiones con sus niveles, las medidas y todas las fuentes de datos, en este caso proveniente de archivos de texto.

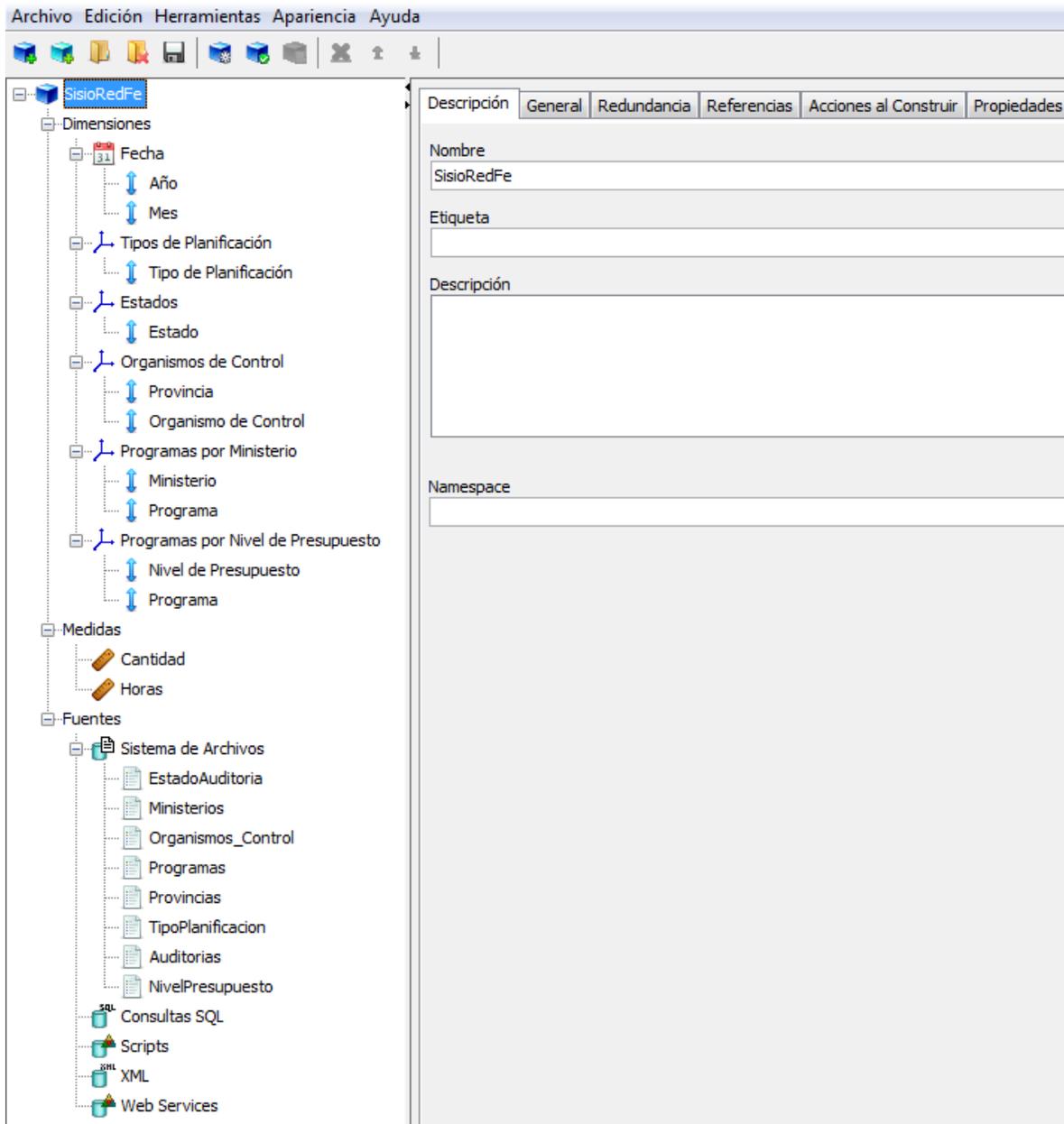


Ilustración 19 - Modelo generado con O3

3.2. – Cubo Resultante

En el presente apartado se podrá visualizar la navegación del cubo generado por el modelo anteriormente mencionado con sus gráficos y tablas asociadas.

La ilustración 20 muestra como ejemplo una representación en formato gráfico de los datos de un cubo. El mismo pertenece al análisis de datos de cantidad de auditorías por Ministerio por año. Por medio de la barra de navegación que se observa en la parte superior de la ventana se puede agregar o eliminar cualquier componente de datos del cubo. Arrastrando y soltando las dimensiones se puede conformar la visión de los datos que se desee. Al final de la barra se encuentra las medidas seleccionables por medio de un combo. También se puede observar debajo de cada dimensión un combo con la granularidad de las mismas. Seleccionando una opción se puede bajar o subir el nivel de la dimensión, y al mismo tiempo es posible aplicar filtros para cada una de ellas.

De la misma manera y con las mismas opciones de navegación se puede visualizar también una tabla con los mismos datos en la ilustración 21.

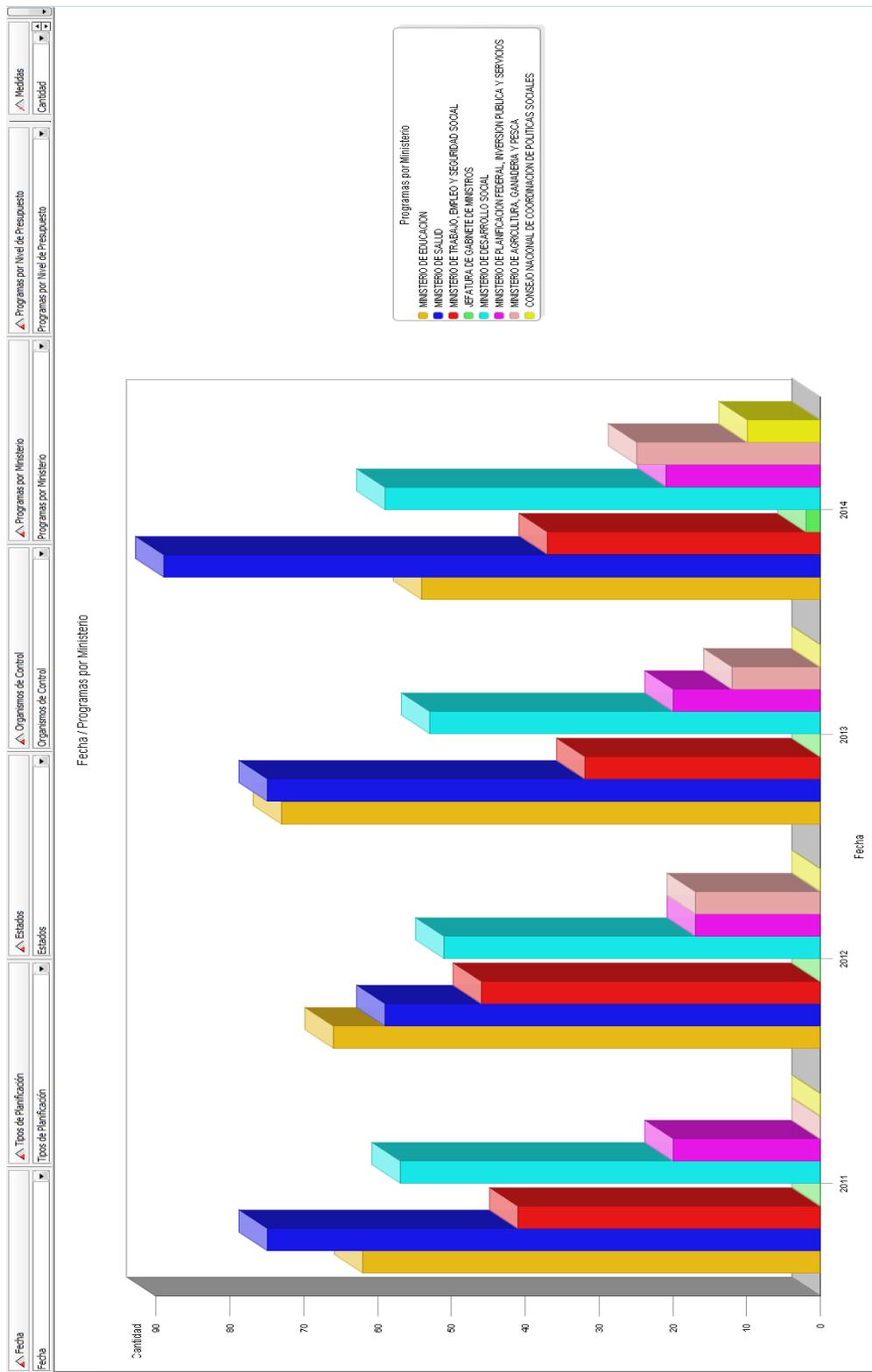


Ilustración 20 - Gráfico Ejemplo del Cubo

Fecha / Programas por Ministerio

Programas por Ministerio	Fecha			
	Año			
Ministerio	2011	2012	2013	2014
MINISTERIO DE EDUCACION	62,00	66,00	73,00	54,00
MINISTERIO DE SALUD	75,00	59,00	75,00	89,00
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD...	41,00	46,00	32,00	37,00
JEFATURA DE GABINETE DE MINISTROS	0,00	0,00	0,00	2,00
MINISTERIO DE DESARROLLO SOCIAL	57,00	51,00	53,00	59,00
MINISTERIO DE PLANIFICACION FEDERAL, INVERS...	20,00	17,00	20,00	21,00
MINISTERIO DE AGRICULTURA, GANADERIA Y PE...	0,00	17,00	12,00	25,00
CONSEJO NACIONAL DE COORDINACION DE POLI...	0,00	0,00	0,00	10,00

Ilustración 21- Tabla Ejemplo del Cubo

En las siguientes ilustraciones se muestran la tabla y el gráfico de cantidad total de auditorías realizadas en un año. Comprende la cantidad de auditorías que fueron realizadas entre el año 2011 y 2014 inclusive por todos los Organismos de Control en todos los programas pertenecientes a los Ministerios. La información se seccionó por año.

Medidas -> Cantidad / Fecha

Fecha	Medidas
	Medidas
Año	Cantidad
2011	255,00
2012	256,00
2013	265,00
2014	297,00

Ilustración 22 - Tabla Total de auditorías por año

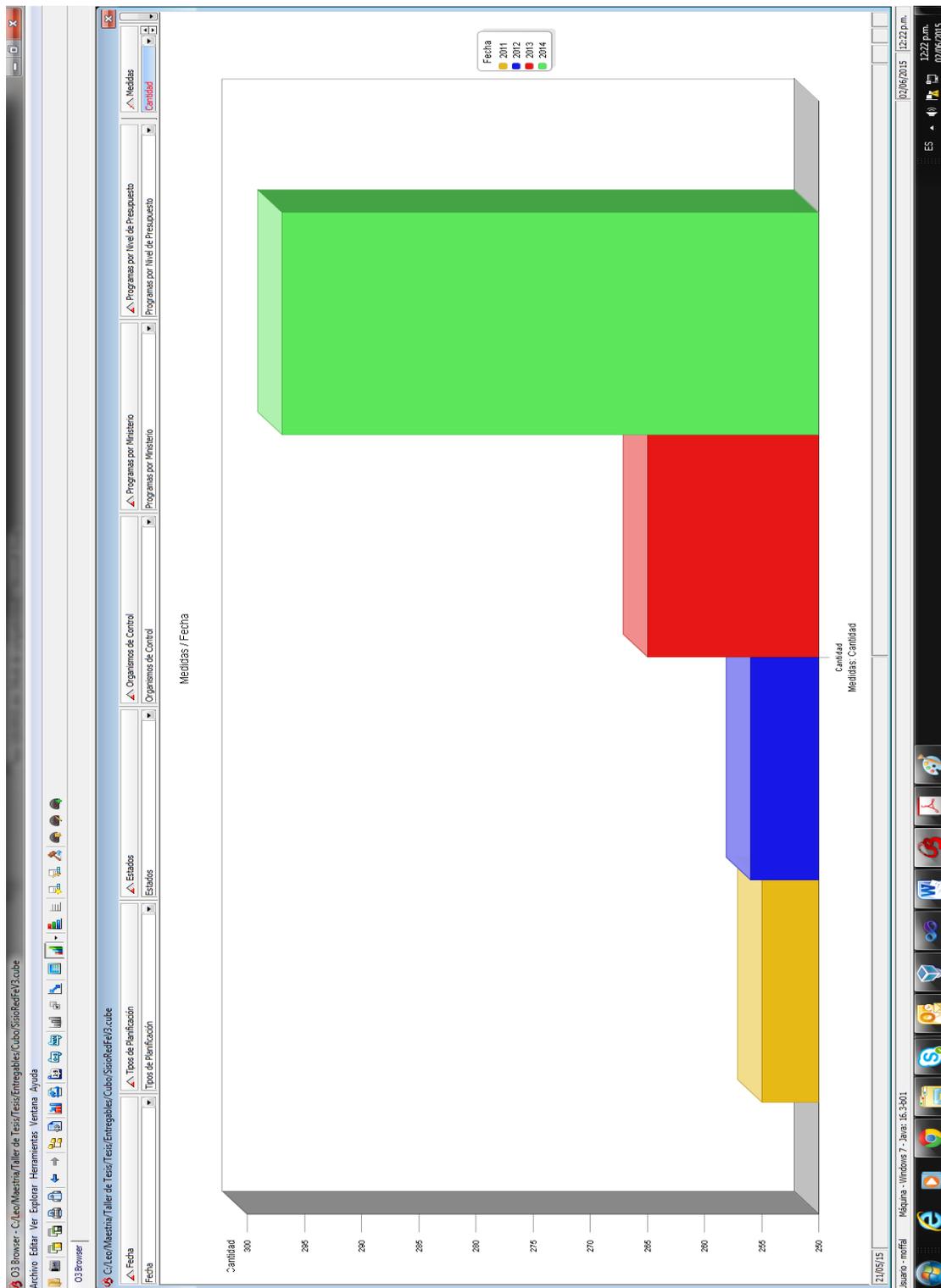


Ilustración 23 - Gráfico Total de auditorías por año

En las siguientes ilustraciones se muestran la tabla y el gráfico de la cantidad total de horas de auditorías realizadas en un año. Comprende el total de horas de auditorías que fueron realizadas entre el año 2011 y 2014 inclusive por todos los Organismos de Control en todos los programas pertenecientes a los Ministerios. La información se seccionó por año.

Medidas -> Horas / Fecha

Medidas	
Fecha	Medidas
Año	Horas
2011	53.737,00
2012	52.555,00
2013	54.650,00
2014	57.600,00

Ilustración 24 - Tabla Total de horas de auditorías por año

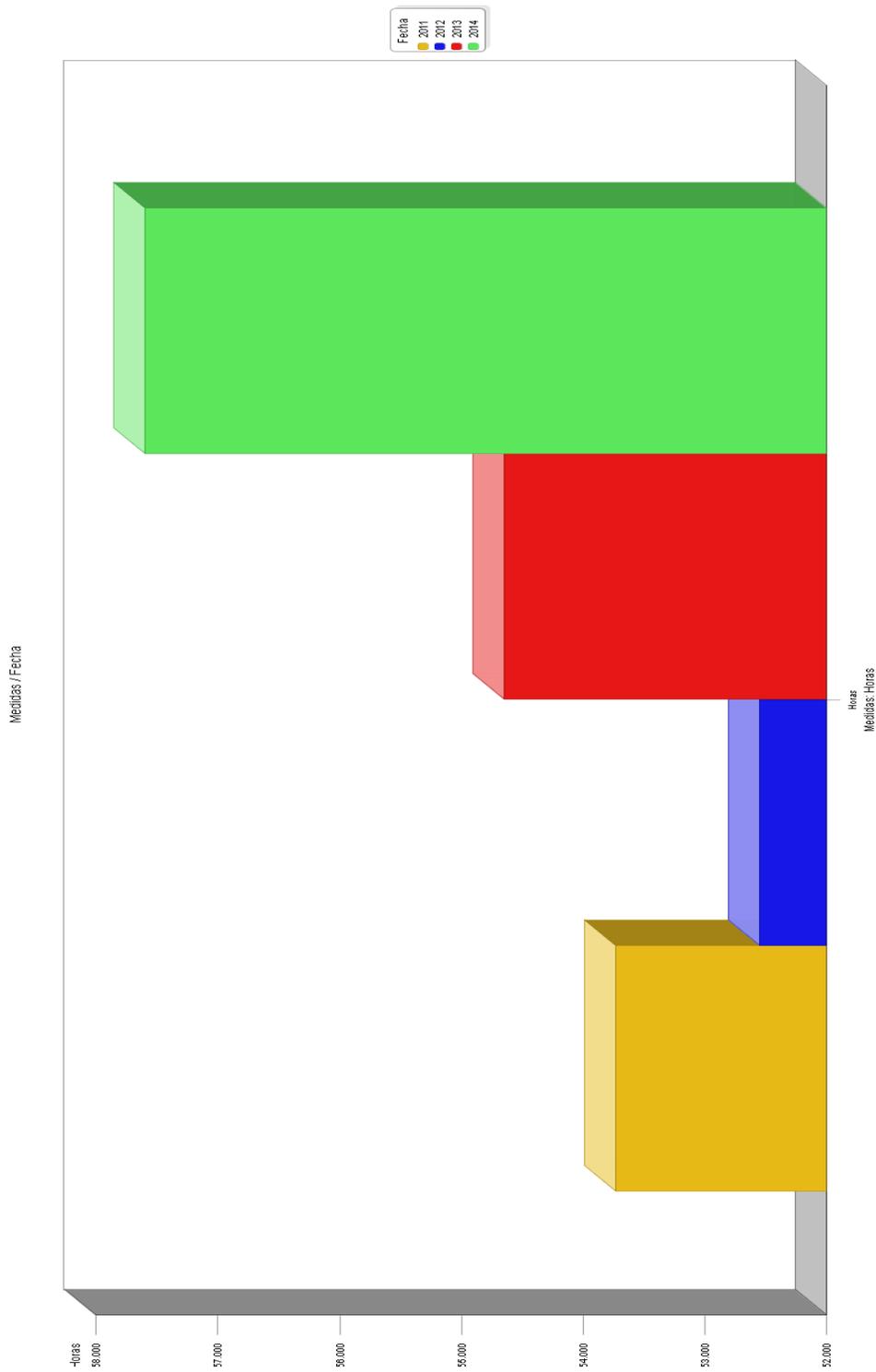


Ilustración 25 - Gráfico Total de horas de auditorías por año

En las siguientes ilustraciones se muestran la tabla y el gráfico del total de auditoría por año, calificadas por sus estados. Comprende la cantidad de auditorías que fueron realizadas entre el año 2011 y 2014 inclusive por todos los Organismos de Control en todos los programas pertenecientes a los Ministerios. La información se seccionó por año y estado de la auditoría.

Fecha / Estados

Estados	Fecha				
	Año				
Estado	2011	2012	2013	2014	Total
No Iniciada	9,00	7,00	5,00	7,00	28,00
Iniciada	5,00	9,00	10,00	27,00	51,00
Finalizada sin Observaciones	28,00	48,00	151,00	149,00	376,00
Solicitud de Baja	1,00	0,00	0,00	0,00	1,00
Anulada	18,00	27,00	31,00	49,00	125,00
Con Observaciones	194,00	165,00	68,00	65,00	492,00
Total	255,00	256,00	265,00	297,00	1.073,00

Ilustración 26 - Tabla Total de auditorías por año según su estado

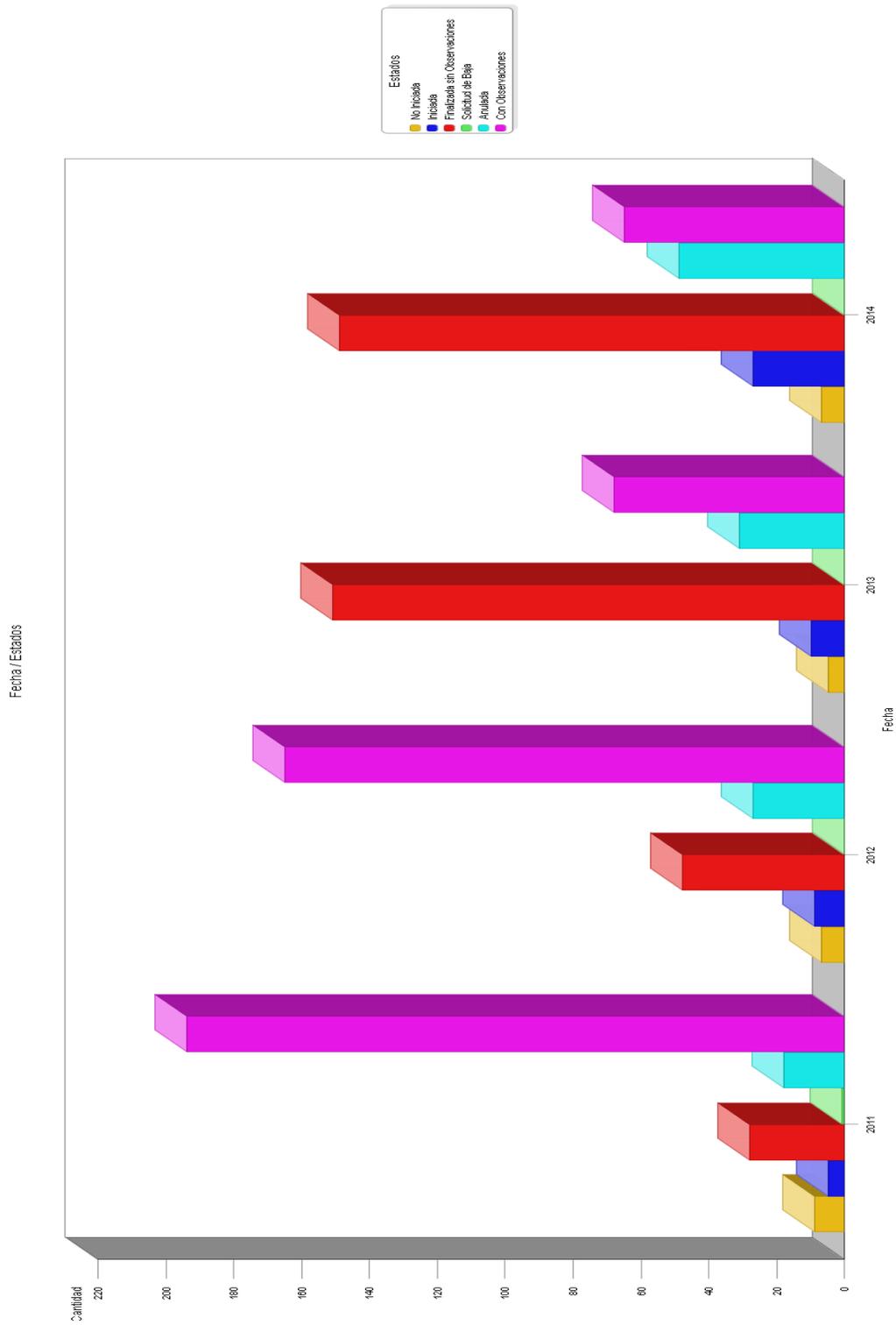


Ilustración 27 - Gráfico Total de auditorías por año según su estado

En las siguientes ilustraciones se muestran la tabla y el gráfico del total de auditorías por nivel de presupuesto, calificadas por sus estados. Comprende la cantidad de auditorías que fueron realizadas entre el año 2011 y 2014 inclusive por todos los Organismos de Control en todos los programas pertenecientes a los Ministerios. La información se seccionó por nivel de presupuesto y estado de la auditoría. A continuación se detallan los valores referenciales de los niveles de presupuesto.

Nivel de Presupuesto	Valores Referenciales
Nivel Bajo	De \$ 0 a \$ 99.999.999
Nivel Medio	De \$ 100.000.000 a \$ 999.999.999
Nivel Alto	Mayor o igual a \$ 1.000.000.000

Tabla 20- Valores referenciales del Nivel de Presupuesto

Programas por Nivel de Presupuesto / Estados

Estados	Programas por Nivel de Presupuesto			
	Nivel de Presupuesto			
	Nivel Alto	Nivel Bajo	Nivel Medio	Total
No Iniciada	14,00	6,00	8,00	28,00
Iniciada	12,00	17,00	22,00	51,00
Finalizada sin Observaciones	132,00	106,00	138,00	376,00
Solicitud de Baja	0,00	1,00	0,00	1,00
Anulada	36,00	47,00	42,00	125,00
Con Observaciones	157,00	132,00	203,00	492,00
Total	351,00	309,00	413,00	1.073,00

Ilustración 28 - Tabla Total de auditorías por nivel de presupuesto según sus estados

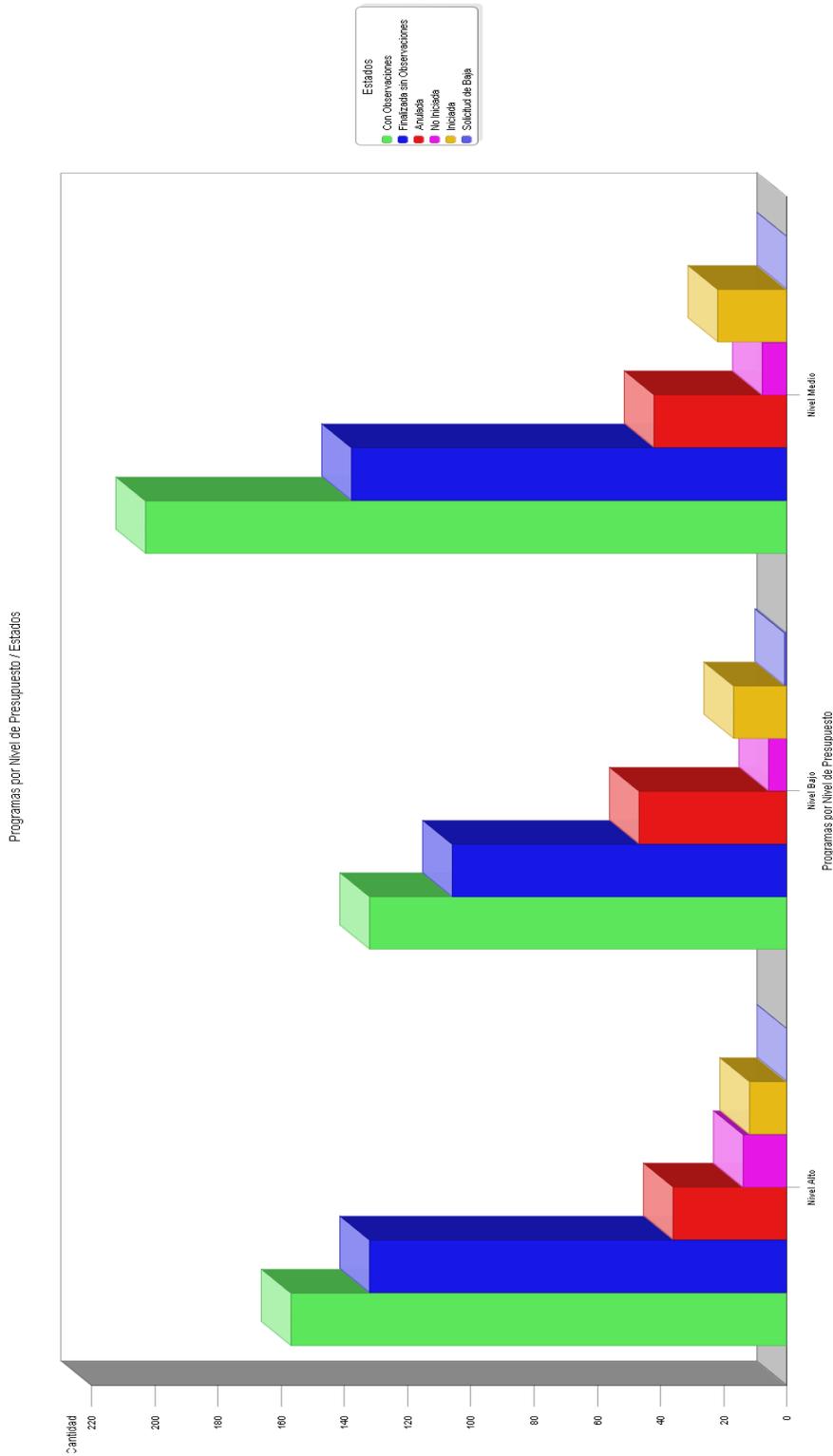


Ilustración 29 - Gráfico Total de auditorías por nivel de presupuesto según sus estados

En las siguientes ilustraciones se muestran la tabla y el gráfico del total de auditorías por año agrupadas por Ministerio. Comprende la cantidad de auditorías que fueron realizadas entre el año 2011 y 2014 inclusive por todos los Organismos de Control en todos los programas pertenecientes a los Ministerios. La información se seccionó por año y Ministerio.

Fecha / Programas por Ministerio

Programas por Ministerio	Fecha					
	Año	2011	2012	2013	2014	Total
Ministerio						
MINISTERIO DE EDUCACION		62,00	66,00	73,00	54,00	255,00
MINISTERIO DE SALUD		75,00	59,00	75,00	89,00	298,00
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL		41,00	46,00	32,00	37,00	156,00
JEFATURA DE GABINETE DE MINISTROS		0,00	0,00	0,00	2,00	2,00
MINISTERIO DE DESARROLLO SOCIAL		57,00	51,00	53,00	59,00	220,00
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS		20,00	17,00	20,00	21,00	78,00
MINISTERIO DE AGRICULTURA, GANADERIA Y PESCA		0,00	17,00	12,00	25,00	54,00
CONSEJO NACIONAL DE COORDINACION DE POLITICAS SOCIALES		0,00	0,00	0,00	10,00	10,00
Total		255,00	256,00	265,00	297,00	1.073,00

Ilustración 30 - Tabla Total de auditorías por año según Ministerio

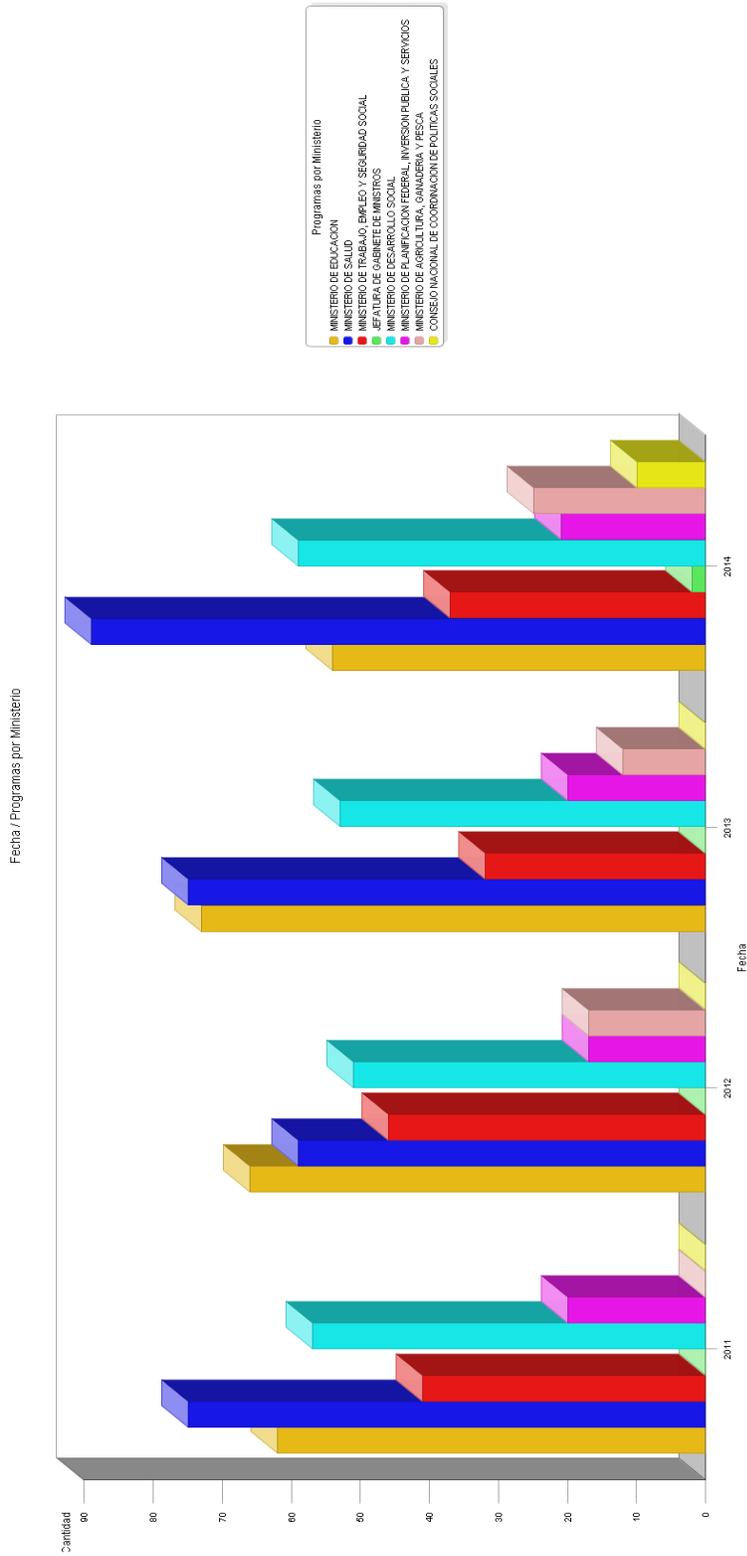


Ilustración 31 - Gráfico Total de auditorías por año según Ministerio

En las siguientes ilustraciones se muestran la tabla y el gráfico del total de auditorías por ministerio clasificándolas según sus estados. Comprende la cantidad de auditorías que fueron realizadas entre el año 2011 y 2014 inclusive por todos los Organismos de Control en todos los programas pertenecientes a los Ministerios. La información se seccionó por estado de la auditoría y por Ministerio.

Estados / Programas por Ministerio

Programas por Ministerio	Estados									
	Estado		Finalizada sin Observaciones	Solicitud de Baja	Anulada	Con Observaciones	Total			
Ministerio	No Iniciada	Iniciada								
MINISTERIO DE EDUCACION	5,00	12,00	85,00	0,00	11,00	142,00	255,00			
MINISTERIO DE SALUD	2,00	16,00	130,00	0,00	14,00	136,00	298,00			
MINISTERIO DE TRABAJO, EMPLEO Y SEGURIDAD SOCIAL	8,00	5,00	39,00	0,00	24,00	80,00	156,00			
JEFATURA DE GABINETE DE MINISTROS	0,00	1,00	1,00	0,00	0,00	0,00	2,00			
MINISTERIO DE DESARROLLO SOCIAL	11,00	9,00	71,00	0,00	48,00	81,00	220,00			
MINISTERIO DE PLANIFICACION FEDERAL, INVERSION PUBLICA Y SERVICIOS	2,00	5,00	34,00	1,00	7,00	29,00	78,00			
MINISTERIO DE AGRICULTURA, GANADERIA Y PESCA	0,00	2,00	13,00	0,00	19,00	20,00	54,00			
CONSEJO NACIONAL DE COORDINACION DE POLITICAS SOCIALES	0,00	1,00	3,00	0,00	2,00	4,00	10,00			
Total	28,00	51,00	376,00	1,00	125,00	492,00	1.073,00			

Ilustración 32 - Tabla Total de auditorías por ministerio según sus estados

Estados / Programas por Ministerio

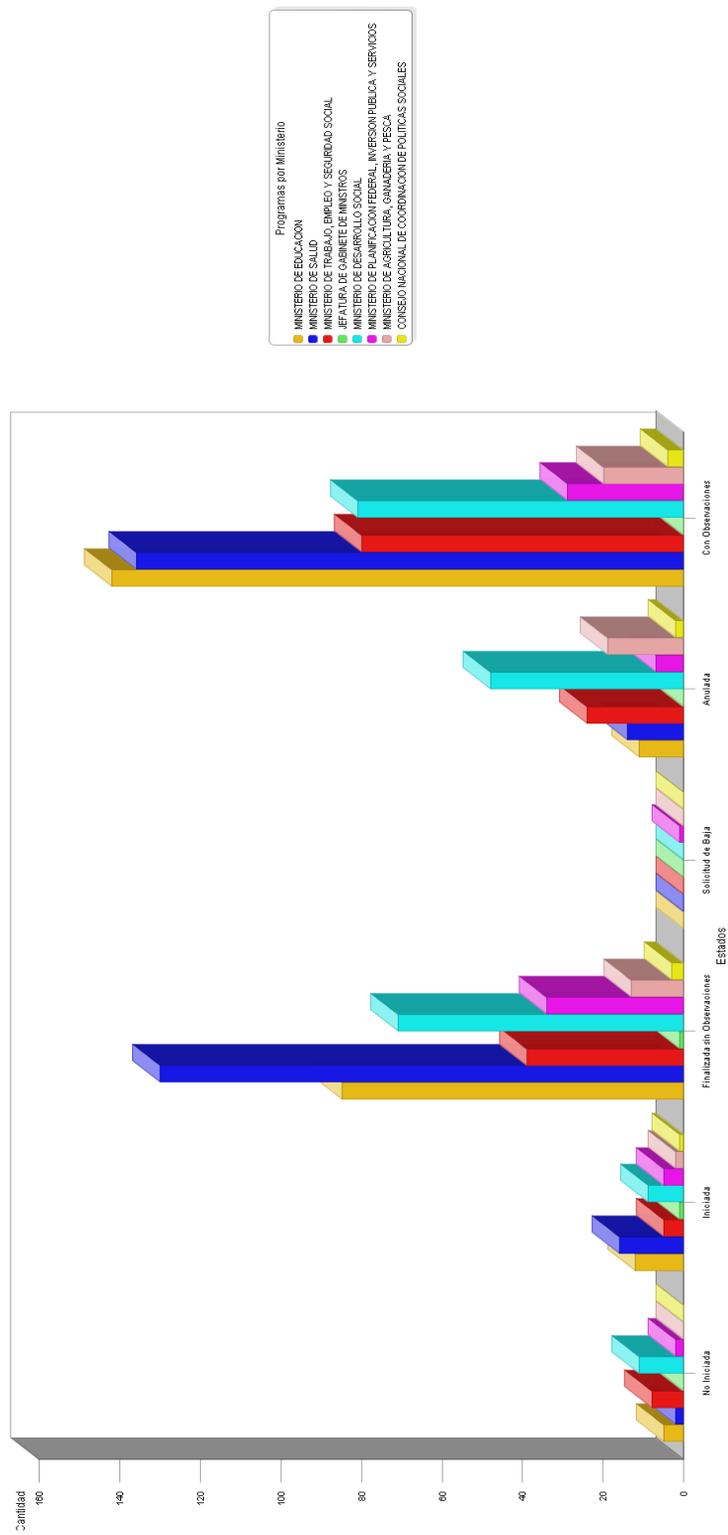


Ilustración 33 - Gráfico Total de auditorías por ministerio según sus estados

En las siguientes ilustraciones se muestran la tabla y el gráfico del total de auditorías por año para el Ministerio de Educación, clasificándolas según sus estados. Comprende la cantidad de auditorías que fueron realizadas entre el año 2011 y 2014 inclusive por todos los Organismos de Control en todos los programas pertenecientes a solo el Ministerio de Educación. La información se seccionó por año y estado de la auditoría.

Fecha / Estados

Estados	Fecha				
	Año				
Estado	2011	2012	2013	2014	Total
No Iniciada	1,00	3,00	0,00	1,00	5,00
Iniciada	0,00	2,00	3,00	7,00	12,00
Finalizada sin Observaciones	3,00	5,00	45,00	32,00	85,00
Anulada	3,00	3,00	3,00	2,00	11,00
Con Observaciones	55,00	53,00	22,00	12,00	142,00

Ilustración 34 - Tabla Total de auditorías por año según sus estados para el Ministerio de Educación

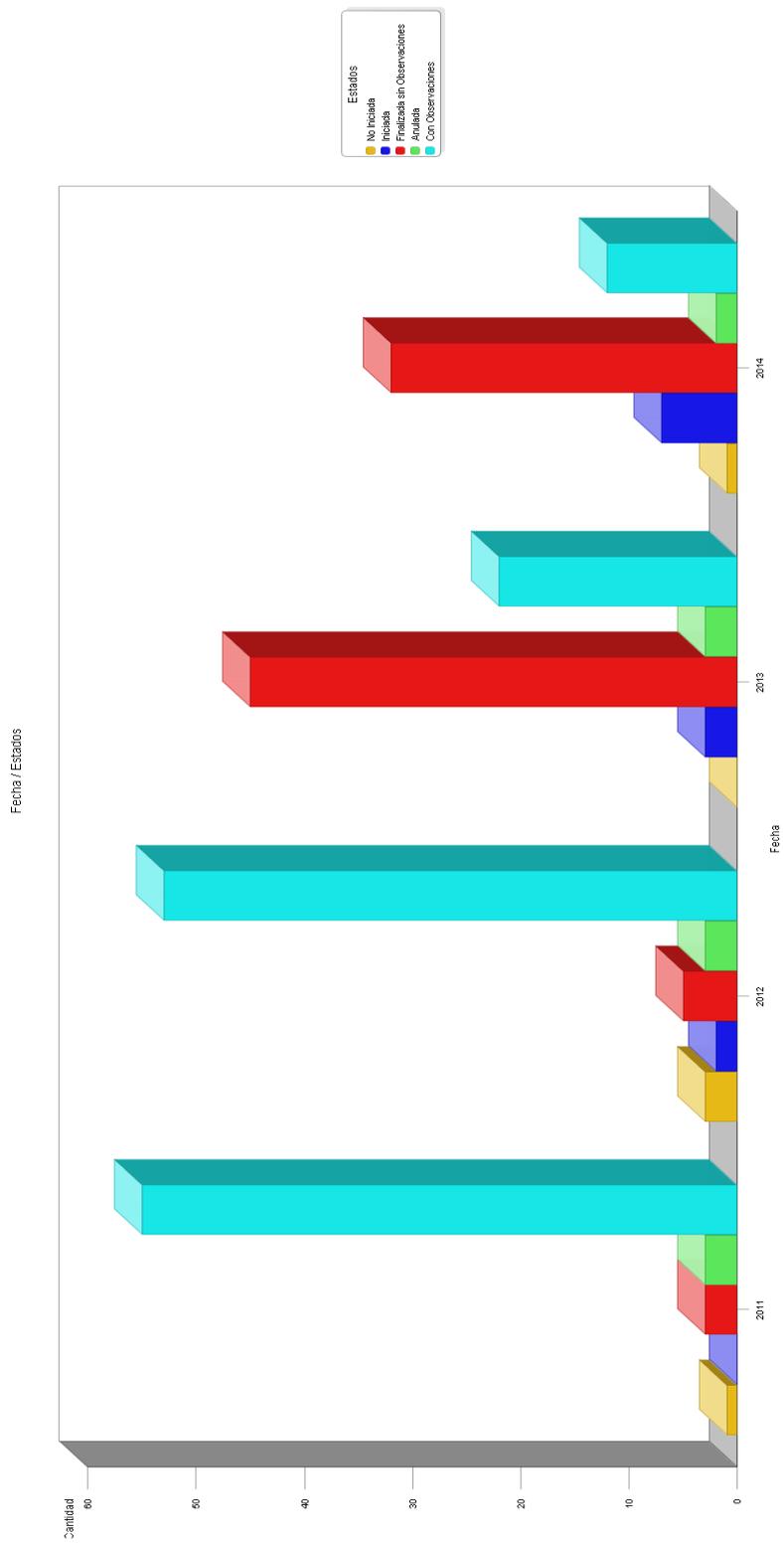


Ilustración 35 - Gráfico Total de auditorías por año según sus estados para el Ministerio de Educación

En la siguiente ilustración se muestra la tabla del total de auditorías por provincia clasificándolas según sus estados. Comprende la cantidad de auditorías que fueron realizadas entre el año 2011 y 2014 inclusive por todos los Organismos de Control en todos los programas pertenecientes a los Ministerios. La información se seccionó por estado de la auditoría y por provincia.

Estados / Organismos de Control

Organismos de Control	Estados							
	Estado	No Iniciada	Iniciada	Finalizada sin Observaciones	Solicitud de Baja	Anulada	Con Observaciones	Total
Capital Federal	Provincia	0,00	2,00	7,00	0,00	2,00	14,00	25,00
Santa Fe		1,00	0,00	16,00	0,00	8,00	35,00	60,00
Santiago del Estero		9,00	13,00	70,00	0,00	22,00	68,00	182,00
La Rioja		1,00	6,00	1,00	1,00	7,00	1,00	17,00
San Luis		0,00	5,00	31,00	0,00	6,00	14,00	56,00
Salta		5,00	5,00	14,00	0,00	13,00	35,00	72,00
Río Negro		8,00	1,00	25,00	0,00	8,00	38,00	80,00
Córdoba		1,00	0,00	23,00	0,00	2,00	16,00	42,00
Chubut		0,00	1,00	4,00	0,00	6,00	36,00	47,00
Tierra del Fuego		0,00	2,00	1,00	0,00	8,00	1,00	12,00
Corrientes		0,00	2,00	28,00	0,00	6,00	17,00	53,00
La Pampa		0,00	4,00	3,00	0,00	1,00	21,00	29,00
Entre Ríos		0,00	0,00	5,00	0,00	2,00	19,00	26,00
Misiones		0,00	0,00	4,00	0,00	4,00	22,00	30,00
Catamarca		0,00	1,00	13,00	0,00	1,00	8,00	23,00
Santa Cruz		0,00	1,00	14,00	0,00	2,00	5,00	22,00
San Juan		0,00	0,00	23,00	0,00	5,00	10,00	38,00
Mendoza		0,00	1,00	6,00	0,00	3,00	27,00	37,00
Formosa		0,00	1,00	24,00	0,00	8,00	10,00	43,00
Tucumán		0,00	2,00	4,00	0,00	6,00	60,00	72,00
Chaco		0,00	0,00	13,00	0,00	3,00	8,00	24,00
Buenos Aires		0,00	1,00	17,00	0,00	1,00	16,00	35,00
Neuquén		2,00	1,00	5,00	0,00	0,00	0,00	8,00
Jujuy		1,00	2,00	25,00	0,00	1,00	11,00	40,00

Ilustración 36 - Tabla Total de auditorías por provincia según estado

En la siguiente ilustración se muestra la tabla del total de auditorías por provincia clasificándolas según los ministerios. Comprende la cantidad de auditorías que fueron realizadas entre el año 2011 y 2014 inclusive por todos los Organismos de Control en todos los programas pertenecientes a los Ministerios. La información se seccionó por estado de la auditoría y por provincia.

Programas por Ministerio / Organismos de Control

Organismos de...	Programas por Ministerio										Total
	Provincia	Ministerio de Educación	Ministerio de Salud	Ministerio de Trabajo	Defensoría del Pueblo	Ministerio de Gabinete de Ministros	Ministerio de Desarrollo Social	Ministerio de Planificación Federal	Ministerio de Agricultura	Consejo Nacional de CO	
Capital Federal	10,00	11,00	0,00	0,00	1,00	3,00	0,00	25,00	0,00	0,00	25,00
Santa Fe	10,00	17,00	16,00	0,00	10,00	1,00	5,00	60,00	1,00	1,00	60,00
Santiago del Estero	43,00	63,00	23,00	0,00	36,00	11,00	5,00	182,00	1,00	1,00	182,00
La Rioja	0,00	4,00	5,00	0,00	6,00	1,00	1,00	17,00	0,00	0,00	17,00
San Luis	14,00	23,00	4,00	0,00	10,00	4,00	1,00	56,00	0,00	0,00	56,00
Salta	17,00	13,00	14,00	1,00	14,00	9,00	3,00	72,00	1,00	1,00	72,00
Río Negro	10,00	19,00	13,00	0,00	26,00	6,00	5,00	80,00	1,00	1,00	80,00
Córdoba	8,00	16,00	9,00	0,00	9,00	0,00	0,00	42,00	0,00	0,00	42,00
Chubut	10,00	9,00	10,00	0,00	8,00	4,00	4,00	47,00	2,00	2,00	47,00
Tierra del Fuego	3,00	3,00	2,00	0,00	3,00	0,00	1,00	12,00	0,00	0,00	12,00
Corrientes	16,00	18,00	7,00	0,00	6,00	3,00	3,00	53,00	0,00	0,00	53,00
La Pampa	8,00	10,00	3,00	0,00	4,00	3,00	0,00	29,00	1,00	1,00	29,00
Entre Ríos	4,00	5,00	4,00	0,00	9,00	0,00	4,00	26,00	0,00	0,00	26,00
Misiones	11,00	6,00	8,00	0,00	4,00	0,00	1,00	30,00	0,00	0,00	30,00
Catamarca	5,00	4,00	2,00	0,00	6,00	4,00	2,00	23,00	0,00	0,00	23,00
Santa Cruz	8,00	6,00	2,00	0,00	5,00	1,00	0,00	22,00	0,00	0,00	22,00
San Juan	7,00	8,00	3,00	0,00	13,00	4,00	3,00	36,00	0,00	0,00	36,00
Mendoza	9,00	7,00	7,00	0,00	7,00	4,00	2,00	37,00	1,00	1,00	37,00
Formosa	13,00	12,00	3,00	0,00	6,00	4,00	4,00	43,00	1,00	1,00	43,00
Tucumán	18,00	14,00	11,00	0,00	14,00	8,00	6,00	72,00	1,00	1,00	72,00
Chaco	5,00	4,00	6,00	0,00	5,00	3,00	1,00	24,00	0,00	0,00	24,00
Buenos Aires	10,00	15,00	4,00	0,00	3,00	1,00	2,00	35,00	0,00	0,00	35,00
Neuquén	4,00	2,00	0,00	0,00	2,00	0,00	0,00	8,00	0,00	0,00	8,00
Jujuy	12,00	9,00	0,00	1,00	13,00	4,00	1,00	40,00	0,00	0,00	40,00
Total	255,00	298,00	156,00	2,00	220,00	78,00	54,00	1.073,00	10,00	10,00	1.073,00

Ilustración 37 - Tabla Total de auditorías por provincia según ministerio

4-DISCUSIÓN

Para comenzar con este apartado es necesario responder si se han cumplido cada uno los objetivos específicos planteados al comienzo del trabajo. Para eso vamos a entrar en detalle uno por uno de ellos.

1. Realizar un diagnóstico de la situación actual y un levantamiento de los procesos involucrados.

Sabiendo que el dominio de datos en el Estado es mucho más complejo que en una empresa ya que las fuentes son mucho más diversas, que es necesario poder encapsular todos esos datos dispersos para obtener una medición eficaz, y que debemos llevar a cabo una estrategia aplicando un modelo dimensional correcto sin sacar del foco central el objetivo institucional; se llevó a cabo un estudio de las necesidades de la Red Federal de Control Público y se llegó a determinar que su fin es evaluar el desempeño de los programas sociales ejecutados por las Jurisdicciones de carácter nacional como la garante de cobertura de control en el ámbito nacional, a través de relevamientos y acciones directas en el campo de ejecución de la Política Social. Así mismo, se evaluaron los procesos involucrados y se consideró que por el incesante crecimiento del alcance de la Red Federal de Control Público, así como también del volumen de auditorías y programas a auditar, surgió la existencia de un sistema de información que sirviese de apoyo, tanto al proceso de planificación de la Red como al seguimiento de su ejecución. Siendo la base de datos de dicho sistema, junto con los datos de apoyo de Entes, suficiente para fuentes de información de nuestro data warehouse, se decidió estudiar el sistema operacional del Sisio Red Federal para consolidar la información más importante para la toma de decisiones de la alta gerencia.

2. Proponer técnicas para fijar objetivos claros y depurar los datos que sirvan como indicadores claves de desempeño útiles para la toma de decisiones de los organismos estatales.

Se llevó a cabo una estrategia aplicando un modelo dimensional sin sacar del foco central el objetivo institucional; de esta forma las herramientas del data warehouse permiten el análisis y presentación de estudios multidisciplinarios en forma dinámica, lo que conlleva a la rápida obtención de resultados a través de los cuales pueden identificarse patrones o tendencias de comportamiento en situaciones determinadas ayudando a resolver los temas sensibles de la población. En este caso práctico se tomó como foco el objetivo institucional de la Red Federal y para el muestreo se utilizaron los datos provenientes del sistema Sisio Red Federal de la Sindicatura General de la Nación y el sistema de Entes. Ya que allí se pudieron extraer los datos resultantes de las auditorías realizadas en los distintos puntos del país años tras años, pudiendo ser usados para evaluar el desempeño de los programas sociales y así afrontar los distintos tópicos problemáticos que la Nación debe enfrentar año tras año.

3. Determinar los principales elementos de los requerimientos que el Estado busca en una modelo dimensional de Inteligencia de Negocios.

Se pudieron determinar los siguientes requerimientos (Detallados anteriormente en el apartado 2.2.1.3.3 – Requerimientos de la solución):

- Contar con una única fuente de información de auditorías. La integración de datos deberá ser mucho más sencilla y rápida.
- Reducir los tiempos de respuesta para que los usuarios de la alta gerencia de la organización, cuenten con reportes confiables de los proyectos cuando ellos los necesiten.
- Integrar la información de las auditorías de los diferentes organismos locales en forma rápida y sin depender de la operación de las áreas administrativas de la Red Federal.
- Integrar información de auditorías para establecer parámetros de control y seguimiento con la información de otros organismos.

- Integrar información histórica de la operación anual de la Sindicatura General de la Nación. La información histórica de la Sindicatura General de la Nación permite evaluar cuales fueron los resultados obtenidos por año en comparación con el planeamiento anual.
- Conocer en cualquier momento, el grado de cumplimiento de las auditorías y la cantidad de irregularidades encontradas en el proyecto.
- Simplificar la consolidación de la información de todas las áreas necesarias de la Sindicatura General de la Nación y los organismos de control local, así como la generación de la amplia variedad de reportes que se necesitan elaborar.

4. Implementar un caso práctico del Estado que permita reflejar la construcción de un modelo dimensional.

La enorme cantidad de información que manejan los sistemas transaccionales u operacionales y que se almacena en bases de datos relacionales, no se estaban aprovechando al máximo en la toma de decisiones a niveles estratégicos de la Sindicatura General de la Nación. La información viaja desde los niveles operativos hasta los niveles tácticos y estratégicos a través de los canales de comunicación establecidos. Pero los ejecutivos, requerían de herramientas ágiles que les permita visualizar, en forma clara y oportuna, métricas y parámetros de control con base en los cuales puedan tomar decisiones rápidamente. En muchas ocasiones el personal de las áreas estratégicas debía solicitar información a diversas áreas y después de ello, un grupo de personas integra toda esa información y la ordena para presentar los informes ejecutivos necesarios para la oficina del Síndico y para la Presidencia de la Nación. La preocupación principal se centra en la necesidad de contar con información oportuna disponible sobre la inversión en proyectos. Por lo tanto, se tuvo la necesidad de desarrollar una herramienta de software para los niveles estratégicos de la Sindicatura General de la Nación, que les proporcione información suficiente y oportuna para tomar decisiones con respecto a las inversiones sobre proyectos que se desarrollan a lo largo de todo el país, para

conocer de manera precisa y en cualquier momento el avance de auditorías sobre los mismos y así poder encontrar irregularidades y desviaciones de la planificación.

5. Evaluar y validar los resultados del caso práctico para corroborar la calidad del modelo y las ventajas que produce para la toma de decisiones.

Una vez obtenidos los resultados, fueron validados y evaluados llegando a las siguientes deducciones:

Como se muestran en la Ilustración 22 y la Ilustración 23 las auditorías fueron incrementando años tras años, esto se debe al intento de la Red Federal de cumplir uno de sus objetivos principales, evaluar el desempeño de los programas sociales cada vez con mayor precisión. Sin embargo, en la Ilustración 24 y la Ilustración 25 puede observarse notoriamente como el total de horas ejecutadas es mayor en el 2011 comparadas con las horas del 2012 (53737 horas y 52555 horas), tomando en cuenta que el total de auditorías para el 2011 es de 255 y de 256 para el 2012. Este comportamiento se genera debido a la experiencia que fueron ganando los agentes de control. En el 2011 les tomaba en promedio unas 211 horas aproximadamente ejecutar una auditoría, mientras en el 2012 el promedio de horas por auditoría es de 205. En el 2013 se mantiene el promedio en 206 horas aproximadamente y en el 2014 baja a 194 horas aproximadamente. Debido a la experiencia que fueron ganando los agentes de control, año tras años se pudieron realizar más auditorías en un tiempo menor de trabajo. Esta circunstancia notoriamente genera mayor eficiencia al momento de cumplir los objetivos de la Red Federal, ya que se genera más control ajustando cada vez más los recursos.

Otros resultados que nos muestra información importante para la toma de decisiones son los que se reflejan en la Ilustración 26 e Ilustración 27. En dichas ilustraciones puede observarse como en los primeros años las auditorías con observaciones (auditorías en donde se encontraron irregularidades en el desempeño de los programas) eran mucho mayores a las auditorías sin observaciones (donde no se encontraron irregularidad en el desempeño de los programas). Sin embargo, con el pasar de los años el resultado refleja que la situación fue

revertida notoriamente, siendo mucho mayor las auditorías sin observaciones para el año 2014. Esto se debe al control que fue ejerciendo año tras año la Red Federal a través de relevamientos y acciones directas en el campo de ejecución de la Política Social. No obstante, los resultados de las Ilustración 28 y la Ilustración 29 presentan un factor un poco alarmante para la alta gerencia. Se puede observar como las auditorías con observaciones son asociadas mayormente a los programas con niveles de presupuesto medio y alto. Esto significa que una gran porción de irregularidades fueron encontradas en los programas con mayor nivel de presupuesto, significando una gran pérdida de dinero para la Nación. Los ejecutivos deberían tomar este resultado como dato importante para realizar acciones correctivas sobre sus políticas sociales.

Por otra parte se puede observar en la Ilustración 30 e Ilustración 31 que los Ministerios en donde se ejercen más auditorías son el Ministerio de Educación y el Ministerio de Salud, debido a que sus programas son catalogados por la Nación como los más sensibles a nivel de Política Social. Sin embargo, en la Ilustración 32 y la Ilustración 33 se observa que en el Ministerio de Educación se encontraron una gran cantidad de observaciones en las auditorías. Aunque dichas irregularidades fueron bajando notoriamente con el pasar de los años, como demuestra la Ilustración 34 y la Ilustración 35.

Por último, la Ilustración 36 y la Ilustración 37 nos muestra como son repartidas las auditorías por Provincia. Estos resultados pueden ser muy útiles para la alta Gerencia a la hora de tomar medidas por región.

Investigaciones de otros autores

La mayoría de los trabajos sobre Inteligencia de Negocios están orientados a empresas privadas tales como las áreas de compras, finanzas o ventas; y muy pocas investigaciones son orientadas al ámbito público. Si bien en estas pocas investigaciones sobre Inteligencia de Negocios en el

Estado siempre se llega a la conclusión de que la Inteligencia de Negocios puede ser utilizada tanto en el ámbito privado como en el público, ninguna de ellas resaltan una metodología o estrategia para poder llegar a obtener resultados de calidad. En otras investigaciones, se realiza una demostración con un caso práctico sin entrar en el detalle de la estrategia utilizada. En cambio, en este trabajo se hizo énfasis en la diferencia de la estrategia entre una empresa pública y privada, y en cuales aspectos se deben tener en cuenta para aplicar adecuadamente Inteligencia de Negocios en el Estado.

Se destacó que una de las diferencias más importante entre los objetivos de empresas con fines financieros y en organismos del Estado es que para los segundos no es únicamente un balance de resultados financieros, sino que su medición tiene indicadores que sirven para medir el desempeño del Estado. Por ello la obtención de datos para una Inteligencia de Negocios en el Estado es mucho más compleja que en una empresa ya que las fuentes son mucho más diversas. El Estado es la suma de una gran variedad de dependencias, organismos, empresas estatales, organizaciones sociales y ciudadanas, entre otras, que trabajan en distinto ámbitos territoriales, y es necesario poder encapsular todos esos datos dispersos para obtener una eficaz medición. Para ello se insistió en utilizar la metodología de data marts, permitiendo así un mejor control de la información que se está abarcando. También se sugirió llevar a cabo una estrategia aplicando un modelo dimensional correcto sin sacar del foco central el objetivo institucional y dejando de lado cualquier información que no sea útil para alcanzar dicho objetivo. De esta forma, se mantendrá el foco en un conjunto más acotado del universo de datos del Estado. Por ejemplo, llevándolo al caso práctico de Red Federal, se analizaron los objetivos y se decidió utilizar sólo los datos relacionados con las auditorías ejecutadas en la Nación y descartar cualquier otro dato no útil para el cumplimiento de los mismos.

Mejoras a futuro

El caso práctico de Red Federal demuestra sólo una porción de lo que es capaz de generar la Inteligencia de Negocios en el Estado. Tenemos la suerte de que las auditorías, que se utilizaron para el desarrollo del data warehouse, se realizan sobre todos los programas sociales de carácter nacional que abarca el gobierno durante un año; por consecuencia podemos tomar una porción de los resultados para obtener una breve conclusión sobre las distintas políticas sociales aplicadas por el gobierno (abarcando todas las áreas en el que el gobierno participa como la educación, la salud, el desarrollo social entre otros). Pero si queremos un resultado más detallado sobre una problemática social, debemos aplicar Inteligencia de Negocios específicamente en el área del problema. Por ejemplo si queremos saber el nivel de la educación debemos aplicar inteligencia de negocios tomando como información las bases de datos provenientes del Ministerio de Educación entre otros.

Otra mejora a tener en cuenta a futuro es la utilización de data mining para predecir futuras tendencias y comportamientos. Si bien en este trabajo pudimos basarnos en eventos pasados provistos por las herramientas usuales de sistemas como soporte de la toma de decisión, el data mining va más allá de los eventos pasados provistos por aquellas herramientas. Es un proceso automático que permite extraer esquemas interesantes, no triviales y previamente desconocidos, de los datos y descubrir relaciones entre variables. Estas herramientas de data mining pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas. Exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar.

5-CONCLUSIONES

Se ha puesto en evidencia que el Estado puede sacar provecho de la Inteligencia de Negocios para la toma de decisiones sobre cuestiones actuales que afectan a la sociedad.

Hoy en día hay una gran dificultad para acceder a los datos de todos los organismos del Estado, y además de eso hay que tener en cuenta la diversidad de objetivos de cada organismo. Por esa razón no se piensa en un gigantesco data warehouse único para todo el Estado. Sino que debemos llevar a cabo una estrategia aplicando un modelo dimensional correcto sin sacar del foco central el objetivo de cada institución, encapsulando todos esos datos dispersos para obtener una medición eficaz. Esto justificó enfocar el estudio en la Red Federal de la Sindicatura General de la Nación.

Se creó un data warehouse centralizado en los objetivos de la Red Federal, sirviendo para evaluar el desempeño de los programas sociales ejecutados por las Jurisdicciones de carácter nacional. Durante el desarrollo del data warehouse, se identificaron las principales fuentes de datos que la Red Federal utiliza para obtener información y tomar decisiones. Se trata del sistema SISIO Red Federal y del sistema Entes, los cuales son dos de los sistemas de información institucionales cuyo uso es de carácter oficial y obligatorio para el organismo. De ahí que se puede afirmar que la información de la cual se alimenta el data warehouse es muy confiable para los ejecutivos a cargo de las tomas de decisiones. Consecuentemente se llegó a la conclusión de que la inteligencia de negocios aplicada sirvió para ayudar a la toma de decisiones gracias a la rápida obtención de resultados a través de los cuales pudieron identificarse patrones o tendencias de comportamiento en situaciones determinadas. Por ejemplo, se identificó que una gran porción de irregularidades fueron encontradas en los programas con mayor nivel de presupuesto, significando una gran pérdida de dinero para la Nación. Gracias a este resultado los ejecutivos podrán tomar medidas correctivas para disminuir el gasto público.

En resumen, el presente trabajo deja en evidencia que en la medida en que haya objetivos claros de la organización y se disponga de datos para ayudar a cumplir con los mismos, la Inteligencia de Negocios es aplicable también en el Estado ayudando a resolver los temas sensibles de la población.

LISTA DE ILUSTRACIONES

Ilustración 1: Proceso de Toma de decisiones	12
Ilustración 2: Inteligencia de Negocios	19
Ilustración 3: Data Warehouse Central	29
Ilustración 4: Data Warehouse Distribuido	30
Ilustración 5: Data Warehouse de 2 Niveles	30
Ilustración 6: Cubo de tres Dimensiones.....	42
Ilustración 7: Modelo Estrella	45
Ilustración 8: Modelo Copo de Nieve	46
Ilustración 9: Proceso OLAP.....	47
Ilustración 10: Representación Final del Proceso OLAP ante la Vista del Usuario	48
Ilustración 11: Infraestructura de un Análisis de Negocios con Data Mining	55
Ilustración 12: Estructura de la base de datos relacional de Sisio Red Federal	68
Ilustración 13: Estructura de la base de datos relacional de Entes.....	69
Ilustración 14 - Modelo de inteligencia de negocios propuesto para la solución	89
Ilustración 15 – Modelo Dimensional de la Solución.....	91
Ilustración 16 - Dimensiones con Estructura	91
Ilustración 17 - Muestra extracción de datos de las bases de datos relacionales	95
Ilustración 18: Planteamiento gráfico de la Prueba de Hipótesis.....	100
Ilustración 19 - Modelo generado con O3.....	107

Ilustración 20 - Gráfico Ejemplo del Cubo	109
Ilustración 21- Tabla Ejemplo del Cubo	110
Ilustración 22 - Tabla Total de auditorías por año	110
Ilustración 23 - Gráfico Total de auditorías por año	111
Ilustración 24 - Tabla Total de horas de auditorías por año.....	112
Ilustración 25 - Gráfico Total de horas de auditorías por año.....	113
Ilustración 26 - Tabla Total de auditorías por año según su estado	114
Ilustración 27 - Gráfico Total de auditorías por año según su estado	115
Ilustración 28 - Tabla Total de auditorías por nivel de presupuesto según sus estados	116
Ilustración 29 - Gráfico Total de auditorías por nivel de presupuesto según sus estados.....	117
Ilustración 30 - Tabla Total de auditorías por año según Ministerio	118
Ilustración 31 - Gráfico Total de auditorías por año según Ministerio	119
Ilustración 32 - Tabla Total de auditorías por ministerio según sus estados.....	121
Ilustración 33 - Gráfico Total de auditorías por ministerio según sus estados	122
Ilustración 34 - Tabla Total de auditorías por año según sus estados para el Ministerio de Educación	123
Ilustración 35 - Gráfico Total de auditorías por año según sus estados para el Ministerio de Educación	124
Ilustración 36 - Tabla Total de auditorías por provincia según estado	125
Ilustración 37 - Tabla Total de auditorías por provincia según ministerio	127

LISTA DE TABLAS

Tabla 1: Diferencias entre Sistemas Transaccionales y Basados en Data Warehouse.....	25
Tabla 2: Ventajas de un Data Warehouse	36
Tabla 3: Regiones incluidas en las auditorías de la Red Federal	71
Tabla 4: Cantidad de auditorías de muestras año 2011	72
Tabla 5: Cantidad de auditorías de muestras año 2012.....	73
Tabla 6: Cantidad de auditorías de muestras año 2013.....	73
Tabla 7: Cantidad de auditorías de muestras año 2014.....	73
Tabla 8: UAIs incluidas en la Red Federal	74
Tabla 9: OCLs incluidos en la Red Federal	76
Tabla 10: Programas año 2011.....	77
Tabla 11: Programas año 2012.....	77
Tabla 12: Programas año 2013.....	78
Tabla 13: Programas año 2014.....	78
Tabla 14: Total de auditorías año 2011	79
Tabla 15: Total de auditorías año 2012.....	79
Tabla 16: Total de auditorías año 2013.....	80
Tabla 17: Total de auditorías año 2014.....	80
Tabla 18 - Detalle del Modelo Dimensional de la Solución	93
Tabla 19: Tabla de Frecuencias.....	105

Tabla 20- Valores referenciales del Nivel de Presupuesto.....116

BIBLIOGRAFÍA

Bhattacharyy, S. B. (2009). *Data warehousing and data mining: an overview*. <http://www.sbbhattacharyya.info/>.

Eckerson, W. (2004). *Best Practices in Business Performance Management: Business and Technical Strategies*. The Data Warehouse Institute, www.dw-institute.com, p. 1-32.

Gardner, S.R. (1998). *Building the Data Warehouse*. *Association for Computer Machinery-Communications of the ACM*, 41(9), p. 52-60.

Gill, H. S. (1996). *Data warehousing La integración de la información para la mejor toma de decisiones*. Prentice Hall.

Gobierno de la provincia de Neuquén. (2003) *Plan Maestro de Gobierno Electrónico*.

Gonzales, R. (2008b). *Inteligencia de Negocios y Data Warehouse*. Manuscrito inédito, Universidad ESAN, Agosto 2008.

Groth, Robert. (1998). *Data Mining: A Hands On Approach for Business Professionals*. Prentice Hall.

Han, J. (2006). *Data Mining: Concepts and Techniques*. San Francisco, CA: Morgan Kaufmann.

Hayes, B. E. (2006). *Cómo medir la satisfacción del cliente*. México: Alfaomega.

Inmon, B. (2004). *The Logical Data Warehouse*. *DM Review*, 14(6), p. 67.

Inmon, W.H. (2005). *Building the Data Warehouse*. Indianapolis: Wiley Publishing, Inc.

Johnson, L.K. (2004). *Strategies for Data Warehousing*. MIT Sloan Management Review, 45(3), p. 9.

Kroenke, D. (2003). *Procesamiento de bases de datos*. Prentice Hall.

Mc Grimmon, K. (1990). *Decision Making and Problem Solving*.

McLeod Jr, Raymond (2000). *Sistemas de información gerencial, 7^a ed.* Prentice Hall Hispanoamericana.

Revista Política digital una publicación de NEXOS: Editorial NEXOS. Ejemplar 4, Pág. 79 .

Sperley, Eric. (1999) *The Enterprise Data Warehouse Planning, Building and Implementation*.
Pactrice Hall.

Sindicatura General de la Nación. *Información Institucional*.
<http://www.sigen.gob.ar/institucional.asp>.

Sindicatura General de la Nación. *Red Federal de Control Público*.
<http://www.sigen.gob.ar/red.asp>.

Thomas, J.H. (2001). *Business Intelligence - why?* eAI Journal, July, p. 47-49.