



EL SISTEMA DE INFORMACIÓN DEL INEM PARA ANÁLISIS: VISIÓN GENERAL DE SU IMPLANTACIÓN Y POSIBILIDADES DE UTILIZACIÓN E INFORMACIÓN DISPONIBLE.

Antonio Hernando Sánchez.

*Funcionario del Cuerpo Superior de Sistemas y Tecnologías de la Información,
Consejero Técnico de Informática destinado como Jefe del Área de Estadística del INEM.*

INTRODUCCIÓN:

El INEM, junto con el resto de los Servicios Públicos de Empleo (SPEs), ofrece a los ciudadanos la posibilidad del encuentro entre empleadores y trabajadores que buscan empleo, de mejorar la ocupabilidad de los trabajadores y de incrementar la experiencia y formación profesional. Por ello es de suma importancia para su gestión conocer las repercusiones que tienen las acciones adoptadas para mejorar el desempeño de sus funciones, esto se puede conseguir actualmente a través del análisis de los datos disponibles y así poder tomar la decisión basada en datos reales. Estos datos ya no tienen que ser los utilizados en el sistema operacional dedicado a la gestión, si no que se puede contar con un almacén de datos histórico o banco de datos estadístico y así obtener una perspectiva más amplia.



Para satisfacer sus necesidades de información las distintas unidades de análisis del Organismo tienen a su disposición un almacén de datos, de cuyo estudio es objeto esta comunicación, que iniciado en 1999 se puede considerar que ha alcanzado su madurez atendiendo a su perspectiva histórica y tamaño, al número de usuarios reales y potenciales y las funciones que desempeña.

El objetivo de esta comunicación es dar a conocer la experiencia del INEM en la solución tecnológica adoptada y a la vez divulgar el potencial de los datos disponibles para otras instituciones interesadas. En el presente trabajo se presentan brevemente la metodología seguida en el desarrollo, la implementación técnica adoptada, las posibilidades de consulta y análisis.

SITUACIÓN DE PARTIDA.

En los últimos años se ha pasado por tres fases bien diferenciadas en la elaboración de informes.

En la primera fase se obtenían a partir de unos contadores que iban acumulando los hechos observados o directamente de los datos del sistema operacional. Los datos se reflejaban en informes sobre soporte de papel desde sistemas de impresión de modo carácter y no era posible obtener otros informes no predefinidos previamente, por la rigidez de las consultas, lo que provocaba que hubiese que desarrollar programas a petición con el incremento de tiempo de respuesta que conlleva.

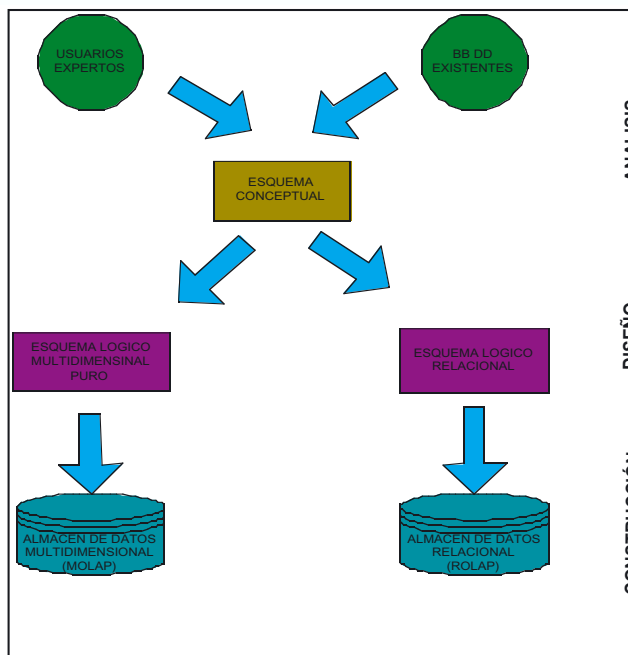
En la segunda fase, los informes eran obtenidos automáticamente a partir de microbases de datos propiamente estadísticas que contenían la evolución anual de las variables observadas. Como principales inconveniente de este sistema estaba la poca capacidad de almacenamiento, la necesidad de que el usuario de la información recibiera formación sobre el uso de la herramienta de análisis, se encontraban bajo sistemas propietarios diversos y los datos estaban altamente redundados, además de ser escasos en muchos ámbitos.



Bajo el anterior escenario se decidió implantar un sistema de información laboral que integrase todos los datos del INEM para desempeñar las funciones de conservar y mantener los datos estadísticos, cubrir las necesidades de información del Organismo, alimentar de forma automática las publicaciones e informes periódicos, tener un acceso a la información ágil y sencillo para cualquier usuario y tener los procesos de carga y producción automatizados o semiautomatizados.

DESARROLLO

La metodología empleada en el desarrollo del almacén de datos del INEM puede verse esquemáticamente en el siguiente gráfico:

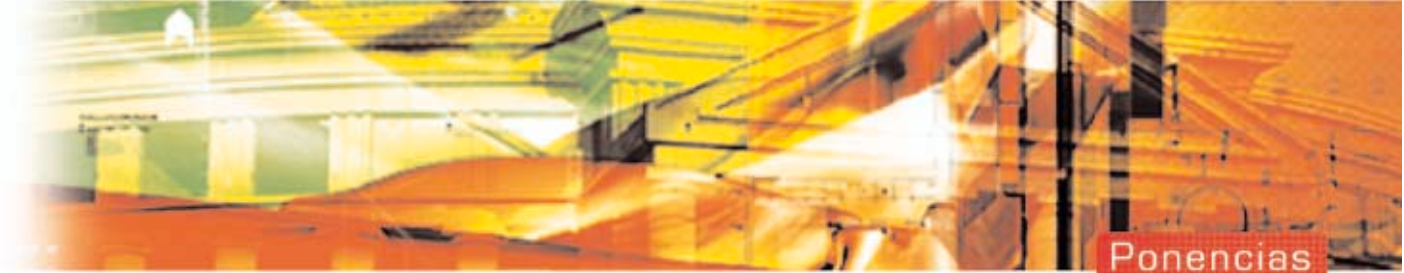




Durante la fase de análisis se hizo una descripción de lo que sería el almacén de datos y una delimitación de su alcance estableciendo la información estratégica a incluir. La determinación de los requisitos generales y del alcance del sistema se llevó a cabo en sesiones de trabajo conjuntas entre los analistas y los grupos de usuarios expertos, además de una encuesta masiva a todas las unidades de la organización susceptibles de usar la información con el fin de detectar sus necesidades de información, definiendo el origen de los datos y las transformaciones necesarias. Como resultado final se obtuvo el modelo conceptual del almacén de datos, identificando las variables y sus jerarquías, las métricas y obteniendo los esquemas de hecho mas apropiados a la definición elegida. Así mismo se definió y dimensionó el entorno tecnológico requerido en base al rendimiento esperado y al tamaño previsto.

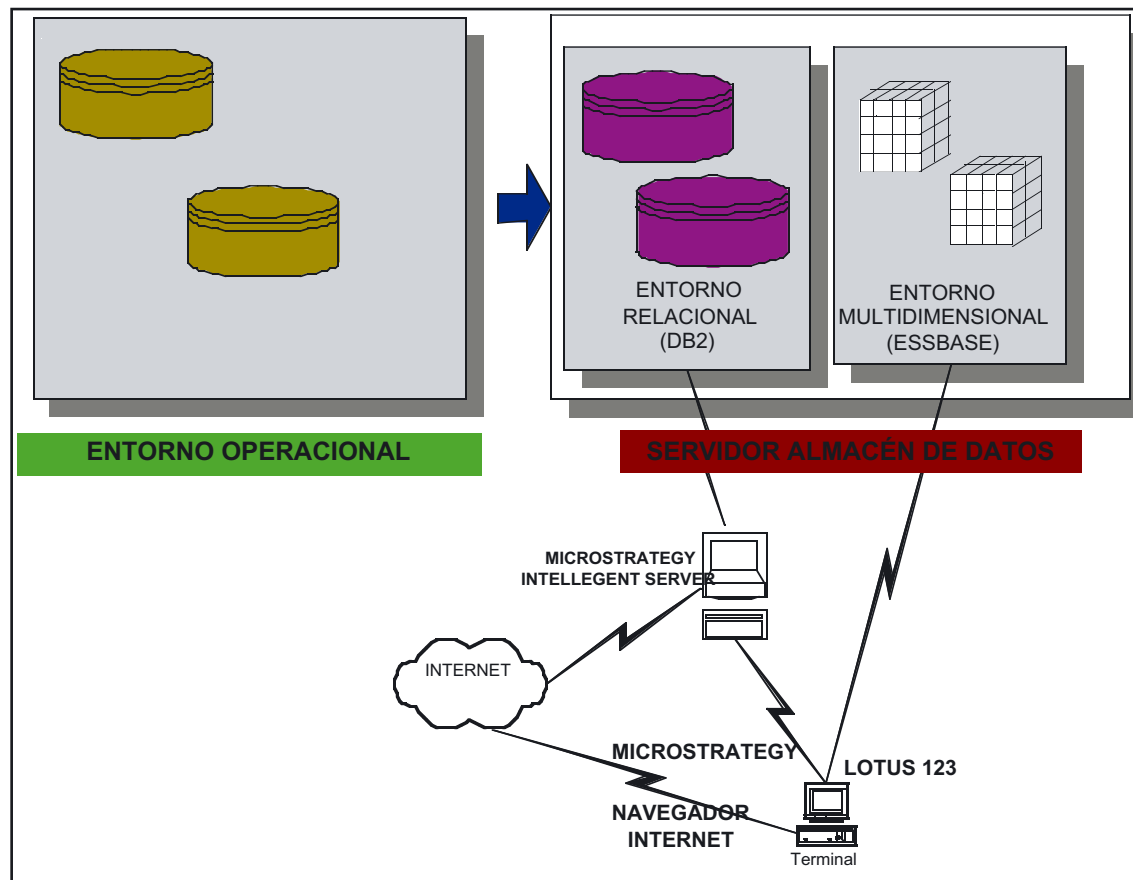
Partiendo del modelo conceptual se diseñaron los esquemas lógicos relacional o multidimensional, según los tipos de datos. Por un lado se diseñó un almacén de datos primarios o atómicos referentes a todos los hechos observados de uso exclusivo para el análisis en profundidad por el INEM con alto grado de confidencialidad para implementar bajo un gestor de base de datos relacional. Por otro, se diseñó el sistema multidimensional que recoge en bases de datos agregados la información estadística que alimentan las series históricas del Instituto, el acceso es rápido y sencillo y es susceptible de ser abierta para consultas externas.

En el período de construcción las actividades de implementación del almacén de datos vinieron determinadas por la solución tecnológica elegida, que fue resuelta como es preceptivo para las compras de la Administración mediante tres concursos públicos: para el hardware, para los gestores de bases de datos y para las herramientas de usuario final. En esta fase las principales actividades llevadas a cabo fueron: la migración de datos desde el operacional y de archivos históricos almacenados en ficheros de texto, determinación de las funciones de extracción transformación y carga de los datos periódicos e implementación de los procedimientos de usuario. En este último apartado se diferenciaron dos perfiles distintos de usuarios, que más adelante se describirán con mas extensión, uno para informes estadísticos, que se resuelven a través del entorno multidimensional puro, y el otro para informes de gestión y análisis que se obtienen a través del entorno relacional.



VISIÓN GENERAL DE LA ARQUITECTURA ADOPTADA

La arquitectura implantada quedó como se muestra en el siguiente cuadro.





Características técnicas del sistema:

Servidor del almacén de datos: SUN ULTRAENTERPRISE 10000. 40 procesadores y SO. SOLARIS 8 y una capacidad de 2503,6 Gbytes.

Gestor relacional: DB2

Gestor multidimensional: ESSBASE

Herramienta de análisis: MICROSTRATEGY

Volúmenes de datos totales:

Datos atómicos (Gestor relacional): 280 millones de registros con un espacio total de disco de 110 Gbytes.

Datos agregados (Gestor multidimensional): 60 Gbytes de espacio total.

Una vez al mes se inician los procesos de extracción de datos del operacional, su depuración y carga en el servidor del almacén de datos.

Como se ha visto en el apartado anterior se dispone de dos entornos de almacenamiento separados que determinan dos tipos de accesos distintos:

- Los usuarios de perfiles estadísticos pueden acceder a datos de las series históricas incluidas en las estadísticas publicadas por el INEM. Su utilización está orientada a obtener los informes tradicionales de las estadísticas y se hace mediante aplicaciones diseñadas para tal fin y desarrolladas en LOTUS 123 que actúan como clientes de las bases de datos multidimensionales, dónde a partir de un índice el usuario puede obtener su propio informe restringiendo al nivel de detalle que se desee. El acceso es rápido y sencillo incluso para usuarios poco expertos, además el usuario experimentado puede construir sus propios informes



manejando para ello el asistente integrado en el menú de la hoja de cálculo. La utilización de la hoja de cálculo conocida por el usuario conlleva la ventaja en cuanto a la simplicidad de la formación, que en breves momentos le permite utilizar todo el potencial del gestor multidimensional, además de la facilidad de exportar datos a cualquier tipo de documento. Mediante este tipo de acceso se obtienen la mayoría de los informes y publicaciones estadísticas periódicas en formato papel o formato digital.

- El perfil de usuario analista, está orientado al análisis de la información, partiendo de lo general para llegar al detalle satisfaciendo los nuevos requerimientos que surgen a medida que avanza el análisis. Se utiliza la herramienta de MICROSTRATEGY WEB para la creación de informes accesible desde cualquier navegador, que desde una visión multidimensional de los datos accede a la base de datos relacional. La herramienta es de fácil uso y la información obtenida es exportable a cualquier tipo de documentos. Los usuarios pueden obtener informes, predefinidos o elaborados en el momento, en forma tabular o gráfica. Todos los informes elaborados por el usuario pueden ser compartidos con el resto. Tiene la ventaja de que es independiente del navegador utilizado y no se necesita descargar previas ni ninguna otra configuración especial en el ordenador del usuario. También existe la posibilidad de definir usuarios con menos derechos y no poder realizar informes, sólo ejecutar los ya definidos, con lo que es posible limitar por esta vía el acceso a los datos confidenciales.
- El perfil de usuario avanzado, con el acceso a través de MICROSTRATEGY DESKTOP combina todas las funcionalidades anteriores con el potencial de desarrollo y administración que ofrece la herramienta.



INFORMACIÓN DISPONIBLE

De las estadísticas del Instituto Nacional de Empleo (INEM) se conocen generalmente las de paro registrado, pero existen otras posibilidades también interesantes para el conocimiento del mercado de trabajo que presentan los datos de la gran cantidad de información que se recoge en el INEM; bien de sus tareas administrativas propias o a través de registros o comunicaciones obligatorias.



Los datos que alimentan el almacén de datos, como se ha visto, proceden en su totalidad de registros administrativos automatizados (8 millones mensuales aproximadamente). Su mayor ventaja es la enorme cantidad de observaciones que se computan, que hace que sean significativos a cualquier nivel de desglose, por eso son especialmente útiles para el estudio referido a zonas geográficas pequeñas (o de variables con muchos valores posibles, como por ejemplo las ocupaciones) donde las encuestas de ámbito nacional no dan resultados fiables. Otra de sus ventajas es la rapidez con que se obtienen resultados, actualmente los primeros datos relativos a un mes finalizado son publicados el segundo día hábil del mes siguiente.

La información disponible se ha organizado en tres áreas:

- Demandas y ofertas de empleo, que recogen los datos de las personas que demandan un empleo y de los puestos ofertados por las empresas, así como de la relación entre ambos: las colocaciones. Para los demandantes se almacena el stock a final de cada mes y flujos habidos en el mes, pudiéndose analizar además de por los datos personales del demandante por la duración ininterrumpida de la demanda, ocupaciones solicitadas, perfil laboral del demandante, situación laboral, servicios recibidos orientados a la colocación,...
- Contratos, que se refiere a los contratos laborales que se han realizado en el país que por ley han de ser comunicados. Contratos iniciales, temporales prorrogados o convertidos en indefinidos, según duración, ocupación contratada, geografía del centro de trabajo, datos del empleador: actividad económica, número de trabajadores, programas de empleo,....
- Formación ocupacional, con información sobre la formación a trabajadores desempleados, según familia profesional, duración del curso, lugar de impartición del curso, nivel del curso, evaluación del alumno, ...



Como información común a las tres áreas para cada persona de la que se tiene registrada información se conoce su edad, sexo, nivel formativo, nacionalidad, lugar de residencia a nivel municipal y código postal e identificador personal que permite relacionar información referida a esa persona en cada una de las áreas. Así interrogantes como los siguientes tienen ahora respuesta inmediata:

¿Cuántos alumnos formados en un período han tenido empleo como consecuencia de la formación? , ¿Cuál es el número medio de demandantes por municipio en una determinada región? ¿Cuál son las ocupaciones mas contratadas?,...

Los períodos de tiempo almacenados, varían desde los cuatro años para los datos relacionados con los demandantes de empleo hasta siete para contratos registrados.

Para mas información relacionada con el significado de los datos se puede consultar <http://www.inem.es/>

