

Título:

IMPLEMENTACIÓN EN UNA BASE DE DATOS MULTIDIMENSIONAL DE LAS ESTADÍSTICAS LABORALES DEL INEM PARA SU CONSULTA Y ANÁLISIS EN LÍNEA

Procesador de textos:

MICROSOFT WORD 97 SR-1

Tema:

Oportunidades y desafíos para la Administración Pública en la Sociedad de la Información

Resumen:

En el presente trabajo se detalla el modelo estadístico del INEM como almacén de datos, y sus posibilidades de consulta; así como la solución técnica adoptada para su implementación:

- Breve descripción de las estadísticas del INEM
- Metodología de desarrollo
- Descripción de la solución técnica
- Posibilidades de consulta y análisis en línea

Autores:

Antonio Hernando Sánchez

- En 1993, ingresa por concurso-oposición libre en el Cuerpo Superior de Tecnologías de la Información de la Administración General del Estado: actualmente Jefe de Servicio de Sistemas Informáticos, responsable de la Unidad de Estadística del INEM (Subdirección General de Informática y Estadística)
- En 1977 ingresa en el Cuerpo de Profesores Agregados de Matemáticas para Instituto de Bachillerato por concurso-oposición libre, permaneciendo como profesor hasta 1993 (actualmente en excedencia con la condición de Catedrático de Matemáticas).
- En 1975, licenciatura en Ciencias (Sección Matemáticas) por la Universidad Complutense de Madrid.

Pascual Lausín Aceña:

- En 1992, ingresa por concurso-oposición libre en la Escala Media de Formación Ocupacional del INEM. Actualmente, Jefe de Servicio de Gestión Estadística del INEM (Subdirección General de Informática y Estadística)
- En 1985, licenciatura en Psicología por la Universidad Central de Barcelona.

IMPLEMENTACIÓN EN UNA BASE DE DATOS MULTIDIMENSIONAL DE LAS ESTADÍSTICAS LABORALES DEL INEM PARA SU CONSULTA Y ANÁLISIS EN LÍNEA.

1) Introducción. Breve descripción de las estadísticas.

De las estadísticas del Instituto Nacional de Empleo (INEM) se conocen generalmente las de paro registrado, pero existen otras 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 las estadísticas del INEM proceden en su totalidad de registros administrativos automatizados (8 millones mensuales). 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.

En los últimos años ha pasado la obtención automatizada de las estadísticas por tres fases bien diferenciadas.

En la **primera fase** se obtenían mediante unos contadores que iban acumulando los hechos observados. Los datos se reflejaban en informes sobre soporte de papel y no era posible obtener otros informes no predefinidos.

En la **segunda fase**, los informes eran obtenidos de bases de datos propiamente estadísticas que contenían la evolución anual de las variables observadas, que permiten análisis MOLAP. Como principales inconvenientes de este sistema están la poca capacidad de almacenamiento y la necesidad de que el usuario de la información recibiera formación sobre el uso de la herramienta de análisis.

Actualmente, el nivel conceptual de la estadística es el mismo que en la fase anterior, pero presenta las ventajas de permitir disponer de los datos de dos años a un nivel de detalle mucho más amplio, posibilitando así la realización de análisis interanuales; y de la menor fragmentación de los modelos multidimensionales existentes gracias al incremento en la capacidad de almacenamiento que supone disponer de los modelos en un mainframe.

En el presente trabajo se presentan brevemente la metodología seguida en el desarrollo, la implementación técnica adoptada, las

posibilidades de consulta y análisis, así como las vías de ampliación previstas para el futuro.

2) Metodología de desarrollo.

La metodología empleada en desarrollo de la estadística se puede ver en el siguiente esquema:

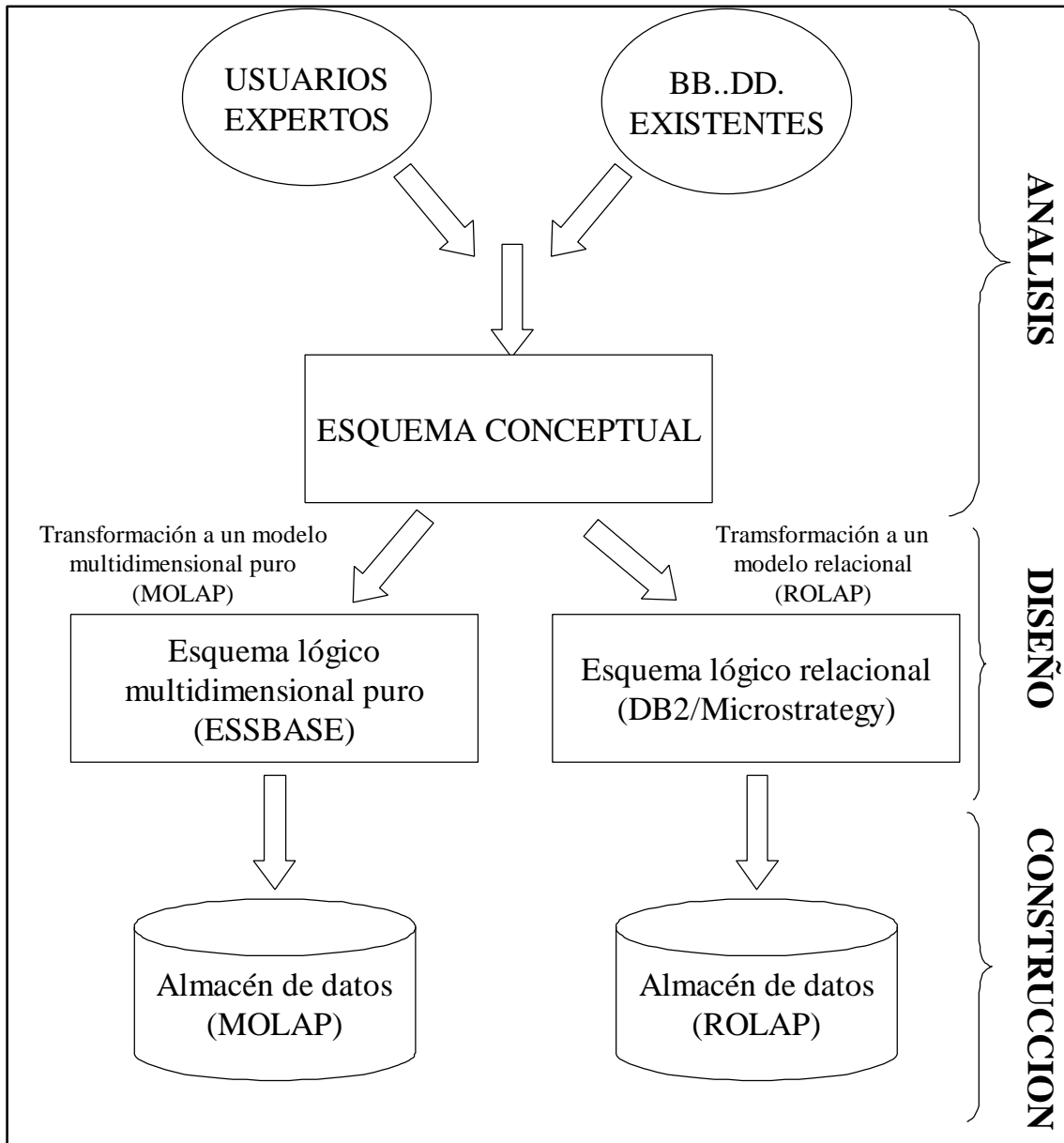


Figura 1. Visión general de la metodología.

En la **fase de análisis** partiendo de los productos estadísticos ya establecidos y de las peticiones singulares que surgen se obtuvo una especificación detallada que pudiera satisfacer las distintas necesidades de información.

Para la definición del modelo lógico conceptual del almacén de datos estadísticos se ha seguido el modelo IDEA (Integración de Datos Elementales Agregados)¹. Mediante esta técnica se crea, partiendo del catálogo de requisitos definidos por el usuario, la estructura del almacenamiento de los datos bajo una visión multidimensional de los mismos. El resultado final es el **esquema de hecho** totalmente independiente de la implementación específica (MOLAP/ROLAP).

Básicamente el proceso de definición del esquema conceptual consiste en identificar las **dimensiones** del esquema de hecho, estructura fundamental del modelo multidimensional de datos, así como la **estructura de celda**.

Como resultado final se han obtenido siete esquemas de hecho: *contratos registrados, prórrogas de contratos, demandantes de empleo, puestos ofrecidos para gestión, colocaciones, prestaciones por desempleo y formación de alumnos*.

Las dimensiones son asociadas con las variables a analizar y pueden tener una jerarquía: *zona geográfica, edad, ocupación, ...* La estructura de celda está formada por una o varias **estructuras de subcelda** y unos **métodos**. En la estructura de subcelda se recogen los datos a analizar por el usuario, p.e.: *número de contratos registrados, duración total en días, ...* y las **funciones de síntesis**: *suma, media, porcentaje, ...* Los métodos se corresponden con los procedimientos que se aplican a varias estructuras de subceldas: *duración media de los contratos = duración total en días / número de contratos registrados*.

En el siguiente cuadro se representa gráficamente el esquema de hecho de contratos.

¹IDEA: Técnica para el diseño de modelos conceptuales de bases de datos multidimensionales creada dentro del Proyecto EINSTEIN en el Departamento de Informática de la Universidad CARLOS III de Madrid.

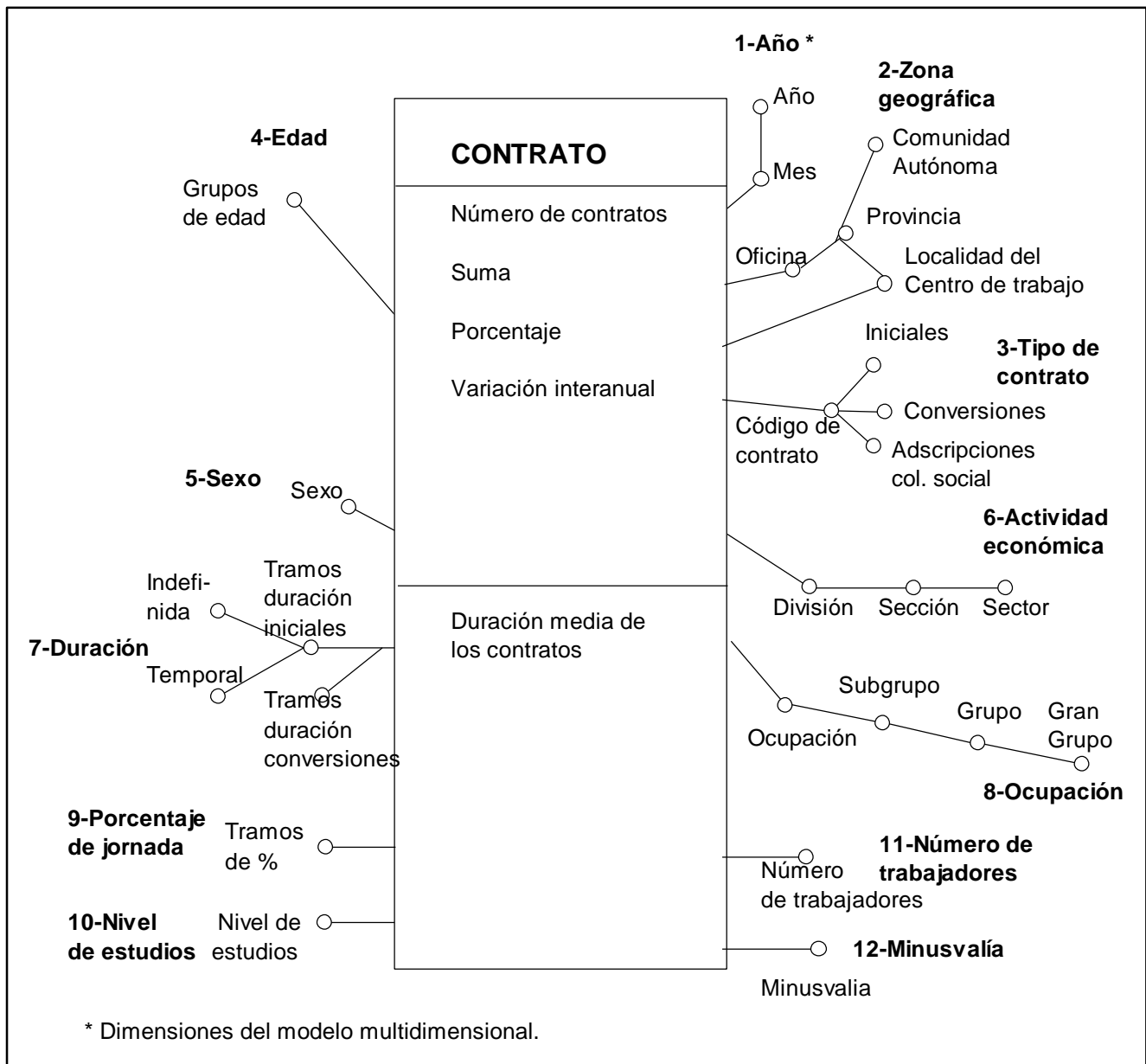


Figura 2. Esquema de hecho del modelo multidimensional de la estadística de contratos. Definición de las vistas de usuario.

Siendo un cubo de los modelos multidimensionales los valores que presenta el esquema de hecho en un momento determinado, la estructura de celda contiene los datos referidos a la combinación de los atributos correspondientes a sus coordenadas.

Durante la fase de diseño se definió el esquema lógico multidimensional adaptando el esquema conceptual a un esquema

convencional específico de los gestores que se utilizan: DB2 OLAP SERVER, como gestor multidimensional, y DB2 como gestor relacional, actuando conjuntamente con MICROSTRATEGY.

La generación del esquema convencional en este caso es muy simple, basta implementar la estructura de celda del esquema de hecho como una dimensión más. (Hacer referencia a ESSBASE sacar de la información de Francisco Jaivie garcía Hidalgo)

Como se puede observar, la implantación del esquema de hecho de Contratos en un sólo cubo sería prácticamente imposible por la poca manejabilidad del mismo, debido al elevado número de dimensiones que presenta y la gran dispersión de datos que tendría por el elevado número de celdillas posibles. Por ello la implementación física se ha realizado solamente de algunas proyecciones del hipercubo, estando reducidas a un número de dimensiones de 7 a 10 y en algunos casos se ha optado por hacer una unión de cubos para facilitar la manejabilidad de los mismos por los usuarios.

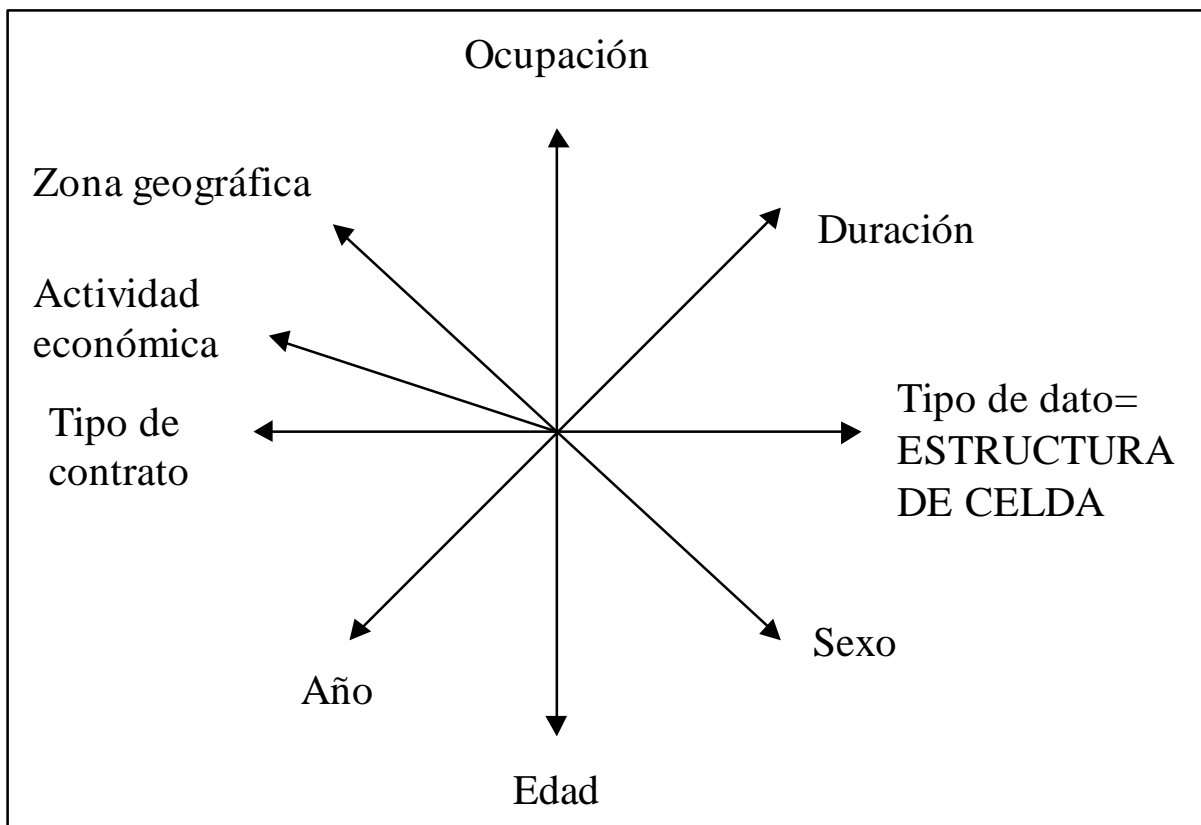


Figura 3. Esquema Lógico multidimensional puro de uno de los cubos de contratos.

3) Solución técnica adoptada :

La arquitectura adoptada, descrita en la figura siguiente, permite descargar todo el proceso de preparación de la estadística mensual del INEM desde el host IBM a un equipo SUN ULTRA-ENTERPRISE 10000 con 40 procesadores en el dominio de Explotación, y S.O. SOLARIS UNIX SYSTEM V R. 4; dedicado a contener el almacén de datos del INEM.

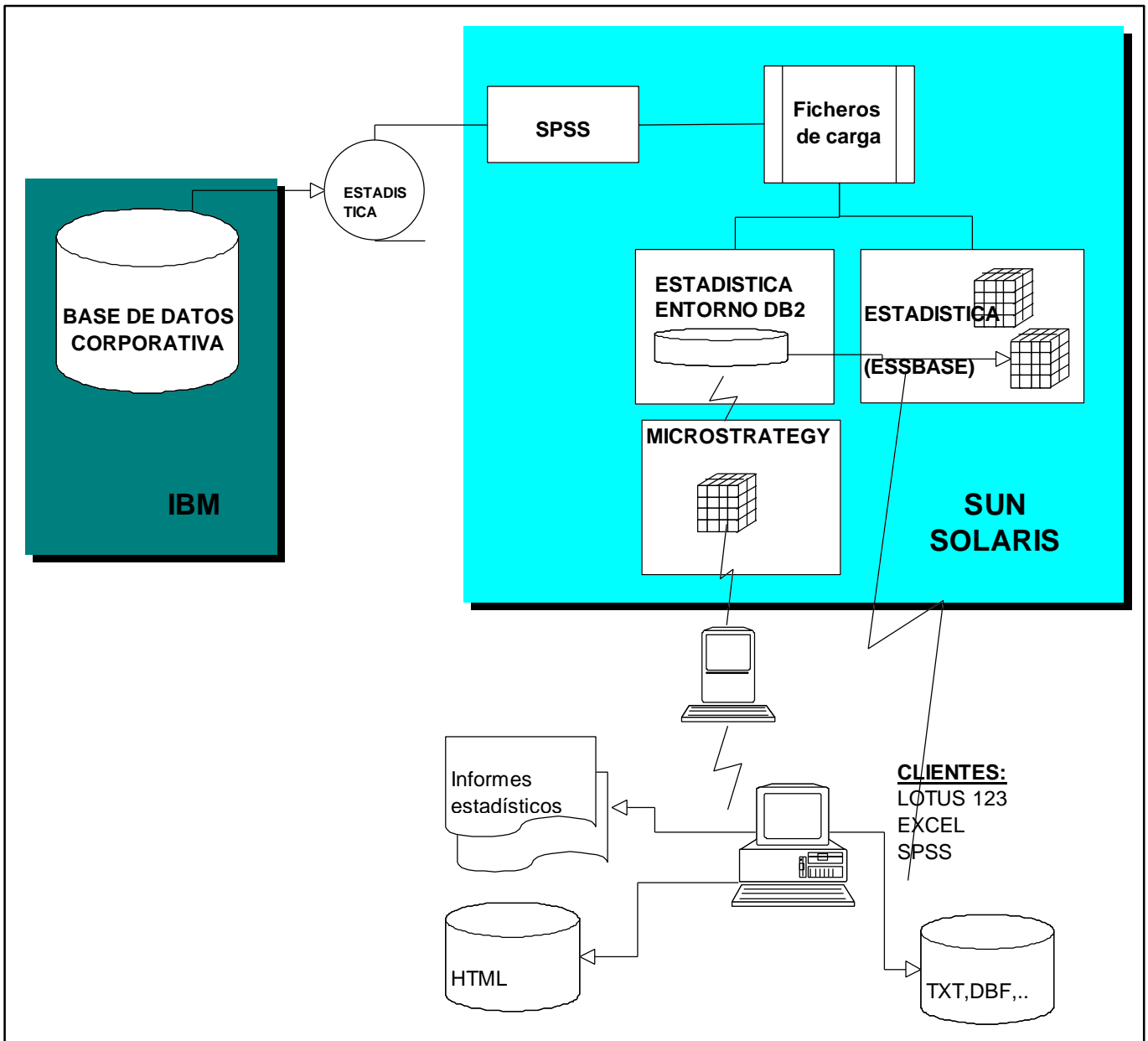


Figura 4. Arquitectura de la solución.

Para la realización de la estadística mensual correspondiente a los diferentes dominios (contratos, empleo, formación,...), se extraen el último día de cada mes los ficheros planos correspondientes a la situación en ese momento de la base de datos operacional.

Una vez los ficheros son transferidos desde el host del INEM hasta el SUN SOLARIS, comienza la generación de la estadística:

- Mediante el paquete estadístico SPSS para SOLARIS se depuran los datos contenidos en los ficheros, verificando la calidad de los mismos, eliminando y subsanando los posibles errores que puedan afectar a las tablas estadísticas que se obtendrán en el paso final.
- La validación de los datos realizada da como primer resultado la obtención de ficheros de datos atómicos depurados, que se almacenan en formato propio de SPSS, y sirven de base para la carga en los gestores de bases de datos multidimensionales
- La propia herramienta SPSS prepara los ficheros que van a ser cargados en ESSBASE a partir de los previamente depurados, agregando los datos en función de las dimensiones que forman el modelo a alimentar; y obteniendo las métricas que conforman la estructura de subcelda para cada una de las combinaciones de dimensiones existentes.
- Un procedimiento automatizado dispara la carga simultánea de los distintos modelos de ESSBASE con los datos obtenidos, y ejecuta el cálculo de los mismos.
- Después de la finalización del cálculo, las tablas estadísticas están disponibles para su consulta e impresión desde los clientes.

Por otra parte, se está trabajando sobre una segunda opción, que estará disponible en el futuro, consistente en la carga de los ficheros depurados de datos atómicos obtenidos mediante SPSS sobre una base de datos DB2, para su consulta desde las herramientas de MICROSTRATEGY DSS Agent, Server, y Web, con tecnología ROLAP. Asimismo, se podrá acceder a DB2 a través de SQL, para la resolución

de peticiones puntuales, a las que no pueda darse respuesta con los modelos de ESSBASE.

En los puestos cliente se ha desarrollado una aplicación en Lotus Script para la obtención e impresión desde Lotus 1-2-3 de los libros estadísticos que se editan en el INEM. Desde Lotus se dispone de una opción de menú que permite su conexión directa con ESSBASE, actuando como cliente.

Igualmente, se han desarrollado los informes de ESSBASE para la generación de páginas estáticas en formato html; situando así las tablas estadísticas más consultadas en la Web de forma automatizada.

La solución elegida ha tenido en cuenta la infraestructura disponible en el INEM (el SUN SOLARIS dedicado a almacén de datos) y las aplicaciones ya existentes en el organismo: Lotus Millenium como paquete ofimático corporativo, DB2 como gestor de base de datos, ESSBASE formando parte del entorno de DB2, y SPSS como paquete estadístico. De esta forma, se ha aprovechado el software disponible, y la arquitectura propuesta ha tenido un coste cero en cuanto a la adquisición de productos para su puesta en marcha. Otra ventaja de la utilización de paquetes ya conocidos por los usuarios del INEM es la reducción de las necesidades de formación al mínimo, ya que por ejemplo, en Lotus 1-2-3, una rápida formación permite al usuario utilizar todo el potencial del gestor multidimensional.

4) Posibilidades de consulta y análisis en línea.

Como se ha mencionado anteriormente, los primeros datos estadísticos del INEM relativos a un mes finalizado son publicados el segundo día hábil del mes siguiente. Con el nuevo sistema de generación de estadísticas implantado, el usuario podrá disponer de prácticamente la totalidad de los datos desde ese momento, en lugar de tener que esperar unos días a la impresión del libro, como sucedía hasta ahora.

La aplicación desarrollada en Lotus permite a los usuarios autorizados del INEM consultar en línea toda la información estadística prediseñada para su impresión y publicación en los libros. El usuario a través del sencillo interfaz de ESSBASE para LOTUS puede navegar por los modelos multidimensionales para los que tenga acceso, extrayendo información al nivel de detalle que desee, simplemente seleccionando el botón de la dimensión elegida.

Lotus SmartSuite - 1-2-3 - [r:\Essbase\lotus\contratos.123]

Archivo Edición Ver Crear Rango Hoja Ventana Essbase ?

C:A1

Inicio H11A H21A H22A H23A H23B H24A H24B H25a H26A H33A H33B H71A H81A

| C | A | B | C | D | E | F | G | H | I | J | K | L | M |
|----|---|---|---------|------------|---------|--------|--------|---------------------------|--------|--------|--------|-------|------------|
| 1 | | DURACION | | PERIODO | | | | OCUPACION | | | | | |
| 2 | | TOTAL DURACION | | MARZO 2000 | | | | TOTAL GRUPO DE OCUPACION | | | | | |
| 3 | | ZONA | | DATO | | | | ACT.ECONOMICA | | | | | |
| 4 | | TOTAL ZONA GEOGRAFICA | | td1 | | | | TOTAL ACTIVIDAD ECONOMICA | | | | | |
| 5 | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | |
| 7 | | VOLVER AL MENU | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | |
| 10 | | 2.1M CONTRATOS DE TRABAJO SEGUN EDAD POR TIPO DE CONTRATO Y SEXO. MES | | | | | | | | | | | |
| 11 | | | | | | | | | | | | | |
| 12 | | | | | | | | | | | | | MARZO 2000 |
| 13 | | | | | | | | | | | | | |
| 14 | | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | | |
| 16 | | | TOTAL | < 20 | 20-24 | 25-28 | 30-34 | 35-38 | 40-44 | 45-48 | 50-54 | 55-58 | >=60 |
| 17 | | | | | | | | | | | | | |
| 18 | | | | | | | | | | | | | |
| 19 | | AMBOS SEXOS | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | |
| 21 | | INDEFINIDO ORDINARIO | 29.889 | 507 | 3.126 | 6.037 | 8.967 | 5.581 | 3.360 | 1.047 | 666 | 391 | 18 |
| 22 | | OBRA O SERVICIO | 376.705 | 29.064 | 87.279 | 77.119 | 56.956 | 43.622 | 32.212 | 22.327 | 15.937 | 8.967 | 3.23 |
| 23 | | EVENTUAL CIRC. PROD. | 412.644 | 32.291 | 106.148 | 88.284 | 60.324 | 45.155 | 32.877 | 21.084 | 14.820 | 7.881 | 3.78 |
| 24 | | INTERINIDAD | 53.976 | 1.620 | 11.930 | 13.446 | 8.895 | 7.249 | 5.649 | 2.853 | 1.568 | 614 | 15 |
| 25 | | INDEFINIDOS LEY 63/97 | 42.644 | 2.314 | 13.624 | 16.309 | 477 | 321 | 278 | 4.146 | 2.780 | 1.799 | 59 |

Arial 10 Sin estilo General Activo

Figura 5. Consulta desde la aplicación desarrollada en LOTUS 1-2-3 de un modelo multidimensional

Además, el usuario experimentado podrá construir sus propias tablas manejando directamente el asistente de ESSBASE integrado en el menú de Lotus; sin necesidad de utilizar el interfaz proporcionado por el Area de Estadística, que se refleja en la figura 5. Para guiar las búsquedas de estos usuarios entre los modelos multidimensionales existentes, estará a su disposición una página en la intranet en la que se detallará la relación de modelos existentes, y la estructura de los mismos.

Si el uso de LOTUS permite a los usuarios del INEM acceder a los datos estadísticos, los informes ESSBASE generados posibilitan la consulta de los cuadros más significativos a usuarios externos al INEM a través de la dirección Web <http://www.inem.es>.

Cuando se trate de peticiones "a medida", a la que no pueden dar respuesta las páginas impresas generadas en LOTUS o los informes generados en html, el acceso a los datos puede realizarse por distintas vías:

- Hasta el servidor ESSBASE, con LOTUS o SPSS (que dispone también de conector con el gestor), generando la nueva tabla en función de la petición.
- Hasta DB2, con las herramientas que proporciona MICROSTRATEGY para el análisis ROLAP, o a través de consultas SQL.

Está previsto además construir páginas Web dinámicas, en las que el usuario de Internet definirá la tabla a obtener, sin más limitaciones que las impuestas por las dimensiones contenidas en los modelos. Las peticiones a ESSBASE serán tratadas por un módulo residente en el servidor de Internet que posibilita la conexión del gestor con páginas ASP.