

Práctica 1: Diseño e Implementación de un Data Mart

Almacenes y Minería de Datos
Curso 2020/2021

Grado en Ingeniería en Informática
Dpto. de Informática e Ingeniería de Sistemas
Universidad de Zaragoza
Escuela de Ingeniería y Arquitectura

Sergio Ilarri y Raquel Trillo

15 de octubre de 2020

1. Objetivos / tareas a realizar

En esta práctica se diseñará e implementará en Oracle un data mart orientado a la toma de decisiones en el ámbito de la aviación comercial de pasajeros. Concretamente se trabajará con información de vuelos, para trabajar conceptos básicos de almacenes de datos. En mayor detalle, los objetivos de esta práctica son:

- Siguiendo la metodología de Kimball, diseñar un data mart para el problema planteado e implementar el esquema relacional propuesto en Oracle.
- Analizar cómo podría incorporarse información de los pasajeros de cada vuelo e implementar una solución en Oracle.
- Proponer otro enunciado de un problema que requiera el desarrollo de un data mart y realizar el diseño del esquema en estrella.

2. Contexto del problema: vuelos comerciales en Estados Unidos

Se desea implantar un data mart para el análisis de información de vuelos comerciales en Estados Unidos. Se requiere que se almacenen los siguientes datos de cada uno de los vuelos: número de vuelo, aeropuerto origen y destino (código y nombre), ciudad origen y destino, estado origen y destino, operadora del vuelo (código y nombre completo), aerolínea (código y nombre), modelo del avión (código y nombre), fecha en la que se realiza el vuelo, tiempo real de vuelo, tiempo estimado de vuelo, y retardo producido en los casos en los que los vuelos se retrasen o adelanten.

En esta primera versión del data mart, el análisis se centrará en datos del rendimiento del vuelo: tiempo total de vuelo y retardo (si es que existe). Se desea saber, por ejemplo, en qué períodos existen mayores retrasos y cuáles son las causas que pueden influir en dichos retrasos (por ejemplo, el modelo de avión o un determinado aeropuerto con una gestión ineficiente).

3. Entrega de la práctica

La práctica se realizará en equipos de tres personas (salvo que existan problemas logísticos, que deberán comentarse y resolverse previamente con el profesor). Al finalizar la práctica se debe entregar un fichero .zip denominado *pN_NIP1_NIP2_NIP3.zip* (donde NIP1, NIP2 y NIP3 son los NIP de los autores de la práctica, con $NIP1 < NIP2 < NIP3$, y en *pN* la *N* representa el número de práctica) con el siguiente contenido:

1. Un fichero PDF denominado **informe.pdf**, que contendrá la memoria de prácticas (respuestas a las cuestiones planteadas, esquemas y código desarrollados, etc., según el caso). En la portada de dicho documento se indicarán los componentes del grupo de prácticas (nombre y apellidos y NIP de cada uno) y el número de práctica. El documento contendrá todo lo necesario para la evaluación de las prácticas. El informe de prácticas debe contener al final un apartado de conclusiones personales que incluirá, entre otras cosas, información sobre el tiempo invertido por cada miembro del grupo de prácticas en la realización de la misma.
2. Todos los fuentes, proyectos, ficheros de configuración y programas de prueba desarrollados, si es el caso, en una carpeta **fuentes**.

Al descomprimir el fichero .zip se deben extraer los ficheros y directorios necesarios en el directorio *pN_NIP1_NIP2_NIP3*. Es importante seguir las convenciones de nombrado y la estructura de ficheros y directorios descrita.

Para la entrega del fichero .zip, se utilizará Moodle del Anillo Digital Docente de la Universidad de Zaragoza. La fecha límite de entrega es el día anterior al de la siguiente sesión de prácticas en el laboratorio a las 23:59.

Como parte de los fuentes a entregar en esta práctica, se incluirán los ficheros de los esquemas en estrella desarrollados con DBDAP (o herramienta similar), así como el código SQL necesario para implementar los esquemas. Además, durante la implementación de los esquemas requeridos se debe efectuar la inserción de un conjunto de tuplas de ejemplo, suficientemente representativo para la realización de pruebas; en la práctica entregada se debe incluir un directorio denominado **insercionDatos** (dentro del directorio *fuentes*) con scripts que faciliten la inserción de su conjunto de pruebas.

4. Recomendaciones

Una vez realizadas las prácticas y entregadas, cada grupo debe presentárselas al profesorado de prácticas en la siguiente sesión de prácticas. Al realizar la presentación el profesorado le formulará cuestiones sobre las decisiones que ha tomado. La práctica debe entregarse en los términos indicados anteriormente y debe ser original (no debe haber sido copiada). Se considerarán los siguientes aspectos:

- La consecución de los objetivos planteados y la adecuada realización de las tareas correspondientes.
- La justificación de las afirmaciones incluidas en el informe.
- La adecuada inclusión de referencias a fuentes de información, si procede.
- La metodología para llevar a cabo la búsqueda de información necesaria para la realización de la práctica.
- La creatividad para la búsqueda de soluciones apropiadas y el correcto análisis de alternativas.
- La estructura y presentación del informe elaborado (completitud del trabajo, precisión, explicaciones adecuadas y completas, justificación de las decisiones tomadas, referencias adecuadas, coherencia de discurso, ausencia de errores tipográficos y ortográficos, etc.).

Hay que asegurarse de que, en caso de que la práctica incluya algún tipo de desarrollo, funciona correctamente en los ordenadores del laboratorio o puede probarse fácilmente sobre una máquina virtual base proporcionada por el profesorado. También es importante someter código limpio (donde se ha evitado introducir mensajes de depuración y comentarios que no proporcionan información relevante).

Es importante extraer conclusiones: resultados obtenidos, indagaciones realizadas, diferencias observadas entre las distintas alternativas y/o herramientas, dificultades encontradas, valoración de los aprendizajes o mejoras de habilidades conseguidas, opinión personal, etc. Se recomienda destacar especialmente aquellos aspectos que muestren la capacidad de indagación, búsqueda de información, autonomía, y curiosidad personal que se hayan desarrollado.