

Ejercicio 1: Métodos de obtención de actualizaciones

La obtención de actualizaciones es un paso del proceso de **ETL** (*Extract-Transform-Load*) utilizado en *Data Warehousing*. En esta etapa se extraen los datos del sistema de almacenamiento. Para no consumir gran cantidad de tiempo en extraer toda la información, extraeremos los datos que se hayan añadido con cada nueva actualización después de la primera carga inicial.

Existen dos enfoques para la extracción de datos:

- **Diferido:** se establece un periodo de tiempo a partir del cual se miran las actualizaciones de los datos. Cada cierto tiempo se recogerá un resumen de los datos para ser actualizados. Con este enfoque, perdemos detalle de las actualizaciones realizadas, pero ahorramos en número de accesos a la base de datos.
- **Inmediato:** actualizamos los datos con cada cambio producido. En este caso tenemos un registro completo de todas las modificaciones, pero si se hacen actualizaciones de forma frecuente, abusaremos de las peticiones a la base de datos.

A la hora de extraer los datos existen diversos métodos que comentaremos a continuación:

- **Comparación de imágenes:** enfoque diferido. Mantenemos una imagen del estado de los datos y cada cierto tiempo generamos una copia nueva que compararemos con la anterior para ver si se han producido cambios.
- **Huella de tiempo:** enfoque diferido. Se mantiene un registro de la última fecha de actualización de la base de datos que actualizaremos cada cierto tiempo al añadir una nueva versión de los datos actualizados,
- **Generados por las aplicaciones:** enfoque inmediato. La aplicación mantiene un registro detallado de todos los movimientos que se van produciendo sobre los datos.
- **Mediante disparadores:** enfoque inmediato. Creamos disparadores sobre el **SGBD** que se accionan cuando se incluyen nuevos datos y guardan un registro de los nuevos movimientos.
- **Log file:** enfoque inmediato. Se hace uso de los archivos de inicio del **SGBD** (*log files*) para recoger los movimientos de las actualizaciones.

Referencias:

- *C.Jensen, T.Pedersen and C. Thomsen.* Multidimensional Databases and Data Warehousing.

Alumno: Daniel Bolaños Martínez