

SISTEMAS MULTIDIMENSIONALES
DOBLE GRADO EN INGENIERÍA INFORMÁTICA Y MATEMÁTICAS
UNIVERSIDAD DE GRANADA

Métodos de Extracción

Alberto Jesús Durán López

27 de febrero de 2020

Definimos **ETL** como el proceso de extraer, transformar y cargar los datos. Nos centramos en la parte de extracción.

1. Tipos de métodos

- **Diferido:** Los métodos diferidos son aquellos en los que la actualización de información a partir de las fuentes de datos no se hace en tiempo real, es decir, no nos fijamos en la historia completa de los datos sino en un momento concreto. Ejemplos de este método puede ser la actualización de la liga de fútbol (1 vez a la semana) o la actualización de la cartilla bancaria (cada vez que el usuario va al cajero).
- **Inmediato:** Se genera y refleja cualquier movimiento o modificación que se realice, como puede ser el número de personas conectadas a una base de datos (es importante contabilizar esto para que no hayan personas sospechosas intentando acceder de forma fraudulenta) o el número de asistentes a un concurrenciado concierto (para no sobrepasar el límite máximo y así evitar aglomeraciones). A diferencia de los métodos diferidos, la ventaja de los métodos inmediatos es que la información intermedia sí se registra.

2. Métodos de extracción

- **Comparación de imágenes:** Es un método **diferido** porque lo que pasa entre la toma de una foto y la posterior se pierde. Es decir, no tenemos los movimientos que ocurren entre medias)
- **Generados por las aplicaciones:** La aplicación que almacena los datos lleva y guarda los movimientos de lo que va ocurriendo. Por ello es un método **inmediato**. Desventaja: Las aplicaciones pueden quedar obsoletas. Si el usuario no tiene experiencia para *actualizar* la aplicación tenemos un problema.
- **Mediante disparadores:** En la asignatura DDSI realizamos pruebas con disparadores. Requiere que el SGBD permita su uso. Se define una acción a ejecutar cuando haya un cambio que esperamos. Se trata de un método **inmediato**.
- **Huella de tiempo:** Necesitamos que esté grabada la fecha de la última modificación. Se usa un enfoque **diferido** ya que sólo se guarda la última modificación.
- **Log file:** Método de extracción **inmediato** que hace uso de ficheros *log* para almacenar todos los movimientos. Desventajas: Es posible que se almacene información prescindible para el SMD luego es necesario un filtrado de datos.

En resumen, todos tienen sus ventajas e inconvenientes y no podríamos tomar uno de ellos como el mejor de todos ya que depende de las circunstancias del momento/sistema. Sin embargo, si nosotros mismos somos los desarrolladores de las aplicaciones, escogeríamos **Generados por las aplicaciones** como el mejor método de extracción ya que no tendríamos problema alguno en actualizar (si fuera necesario) la aplicación.

Referencias

- [1] APUNTES DE CLASE