

Base de Datos

FES Aragón

MTI. Omar Mendoza González

Objetivos

- Conocer y comprender el concepto de Base de datos.
- Reconocer los elementos involucrados en el concepto de base de datos.
- Conocer los objetivos de los sistemas de base de datos.
- Conocer los diferentes niveles de abstracción de la información.

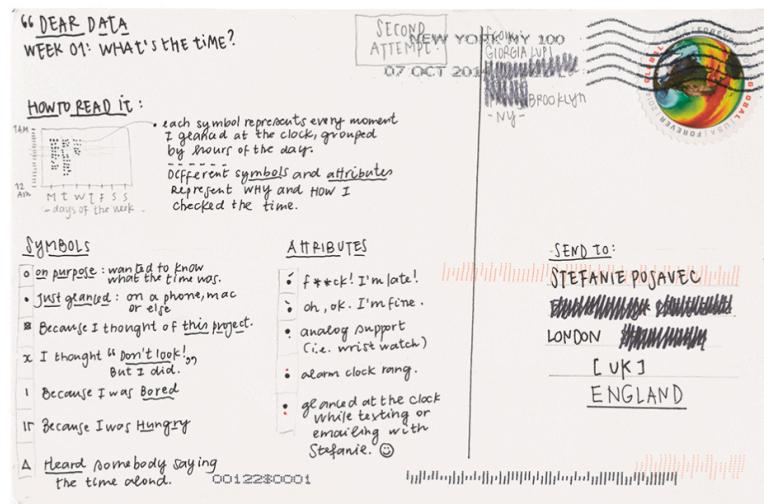
Explosión de datos

- La mayoría de los datos que recopilan, crean y gestionan las organizaciones hoy en día están representados en tres grandes grupos
 - Estructurados
 - Desestructurados
 - Semiestructurados
- Repartidos en bases de datos, documentos json, documentos XML, documentos de procesadores de texto, hojas de cálculo, imágenes y vídeos

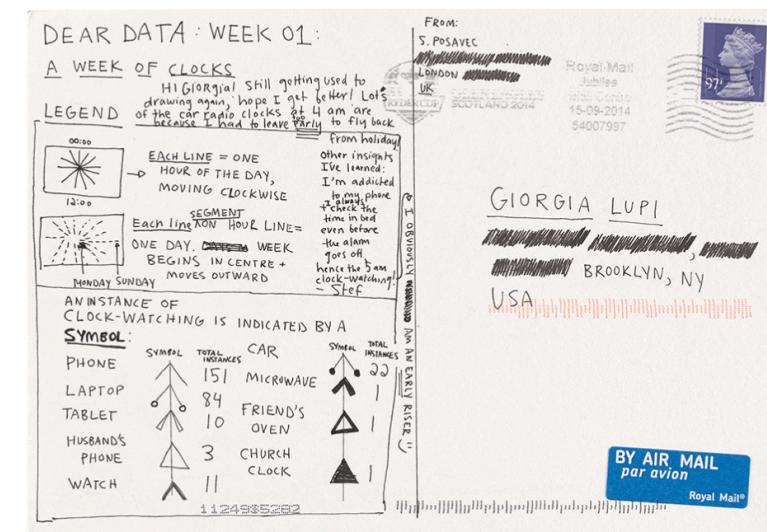
Conceptos Básicos

- Dato
 - Representación almacenada de objetos y eventos significativos
 - Puede estar estructurado:
 - números, texto, fechas, decimal, booleano
 - O no estructurado
 - imágenes, video, documentos
- Es la unidad básica de información y con significado en sí mismo (atómico).

Conceptos Básicos



Drawing her first postcard, Giorgia had an idea for her whole collection: from now on every time she tracks something related to Stefanie, or to Dear Data, she uses a special pen to represent it!



Conceptos Básicos

■ Información

- Datos procesados que aumenta el conocimiento en la persona que la utiliza
- Es un grupo organizado de datos procesados que integran un mensaje sobre un determinado ente o fenómeno; permitiendo que se adquiera el conocimiento necesario para la toma de decisiones

Conceptos Básicos

■ Características de la Información

- Datos
- Orden
- Veracidad
- Vigencia
- Validez
- Valor
- Uso

Conceptos Básicos

- Metadato
 - Datos que describen las propiedades y el contexto de los datos del usuario.

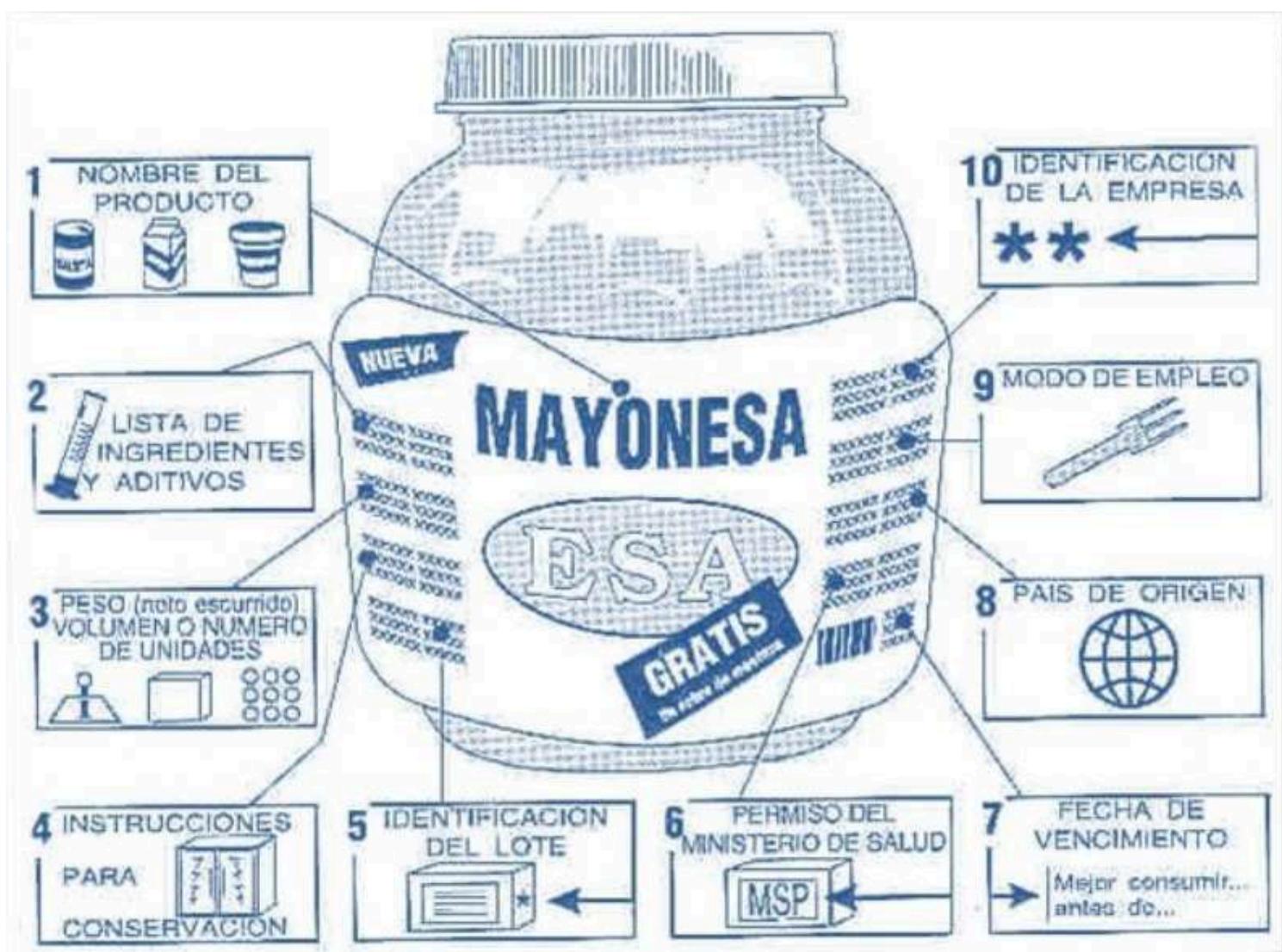
Metadatos

- En el actual contexto de Big Data, Internet de las cosas y cloud computing los metadatos han adquirido una relevancia inédita.
- Frente a ingentes cantidades de información que crecen de forma exponencial, la *gestión de los metadatos* puede ser de gran ayuda para ganar en eficiencia operativa y/o tomar mejores decisiones.

Metadatos

- Facilita búsquedas y análisis
- Facilita la estandarización
- Ayuda a la integración
- Gestión del cambio
- Más seguridad
- Mejora los informes
- Desarrollos más ágiles
- Mejor gobernanza de datos

Conceptos Básicos

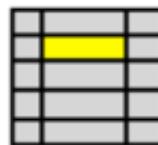


Bases de Datos

- Organización lógica de los datos
 - Sistemas basados en archivos
 - Concepto intuitivo de base de datos

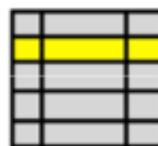
Definiciones de Bases de Datos

- Campo



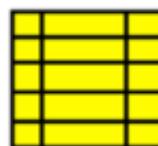
Unidad mínima.

- Registro



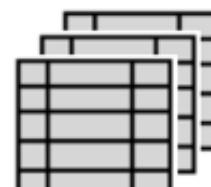
Conjunto de campos
relacionados de iguales o
diferentes tipos

- Archivo



Conjunto de registros
relacionados

- Directorio



Conjunto de archivos
relacionados

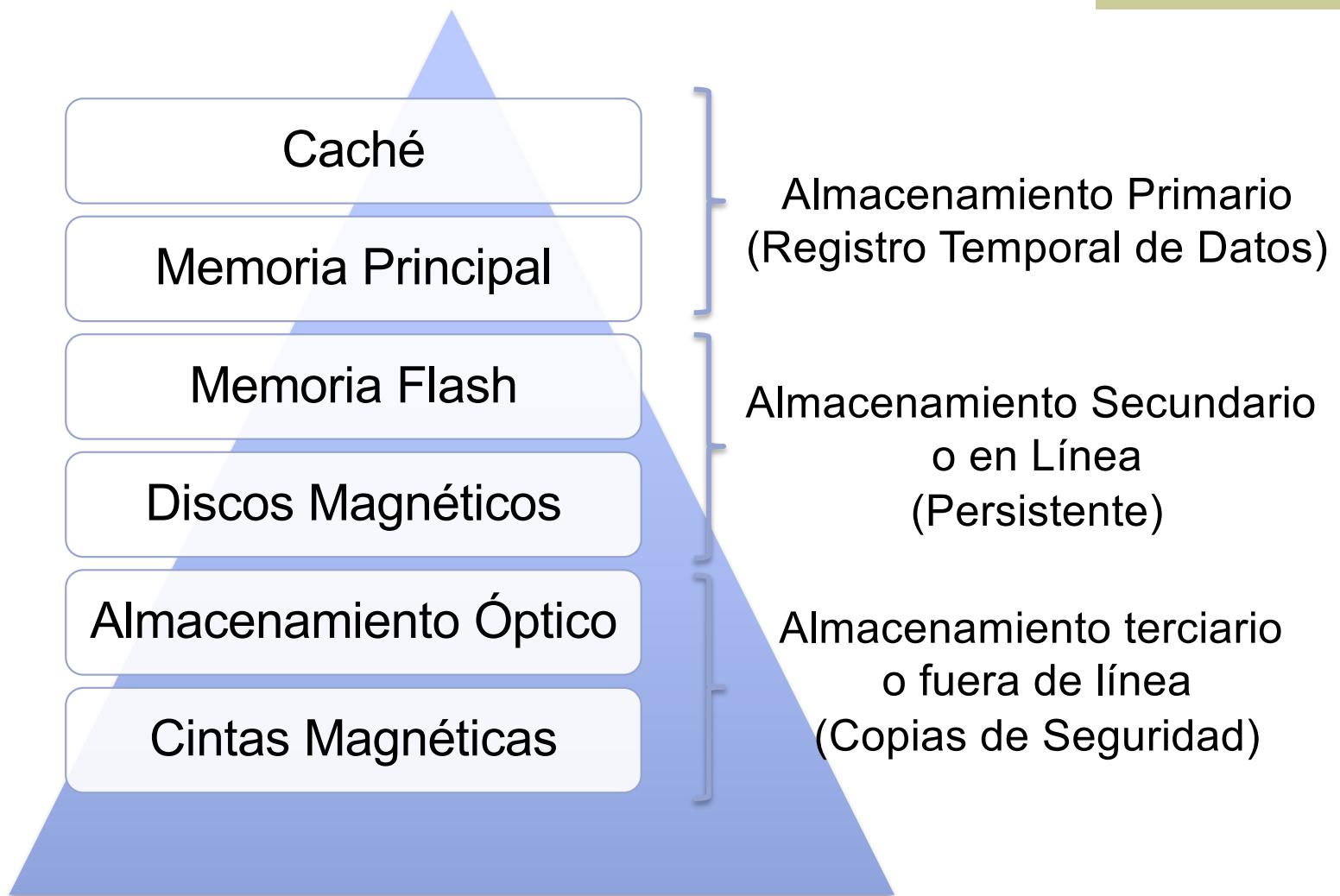
Aplicación de los Sistemas de BD

- Banca.
- Líneas aéreas.
- Universidades.
- Transacciones de tarjetas de crédito.
- Telecomunicaciones.
- Finanzas.
- Producción.
- Recursos humanos.

Almacenamiento Físico de Datos

- Caché
- Memoria Principal
- Memoria Flash
- Discos Magnéticos
- Almacenamiento Óptico
- Cintas Magnéticas

Almacenamiento Físico de Datos



RAID

- Redundant Array of Independent Disks
- En el nivel más simple, un RAID combina varios discos duros en una sola unidad lógica.
- Así, en lugar de ver varios discos duros diferentes, el sistema operativo ve uno solo.
- Los RAIDs suelen usarse en servidores y normalmente (aunque no es necesario) se implementan con unidades de disco de la misma capacidad.

RAID 0

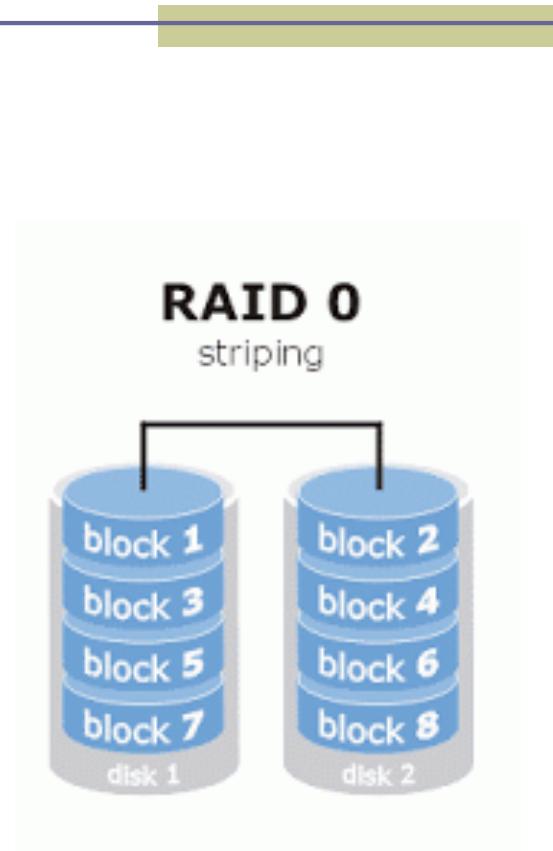
- La información se divide entre todos los discos del sistema, de forma que no se establece ningún tipo de redundancia.

- Ventajas:

- Alto rendimiento, tiempos de acceso muy bajos y posibilidad de acceso en paralelo.
- No tiene coste adicional.
- Se emplea toda la capacidad del disco.

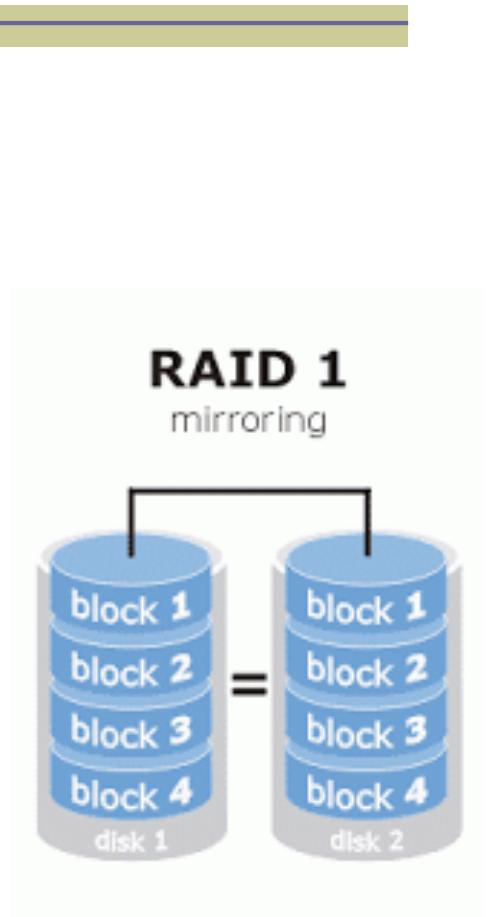
- Inconvenientes:

- No es verdaderamente un disco RAID ya que no presenta integridad de los datos.



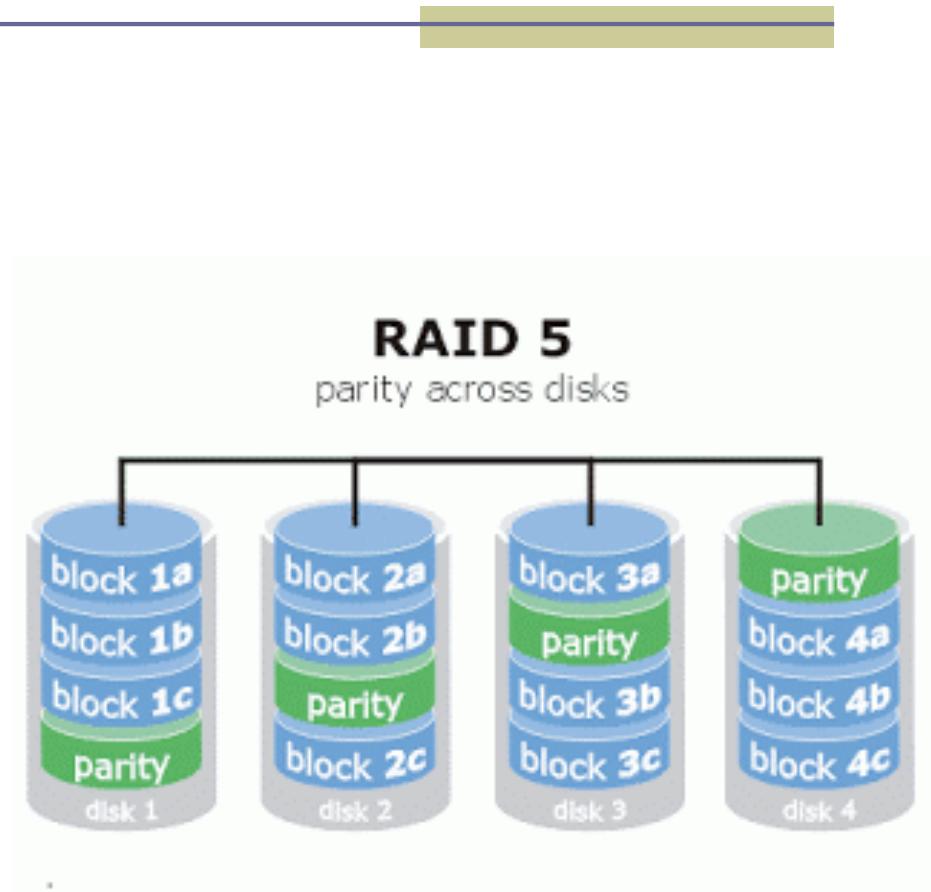
RAID 1

- También conocido como MDA (Mirrored Disk Array).
- Los discos se asocian por parejas y cada una de ellas almacenará la misma información.
- Cada pareja está formada por un disco primario, donde se leen y se escriben los datos, y un disco espejo, donde solamente se escriben las modificaciones y en el que se leerán datos cuando el primario falle.
- Ventajas:
 - En caso de error de uno de los discos se recuperan todos los datos.
 - Es la arquitectura más rápida que presenta tolerancia a fallos.
 - Con un mínimo de dos discos es suficiente.
- Inconvenientes:
 - Es bastante caro, ya que se emplea el doble de espacio del necesario.



RAID 5

- Es un sistema de discos independientes con integración de códigos de error mediante paridad.
- Los datos y la paridad se guardan en los mismos discos, por lo que se consigue aumentar la velocidad de demanda.
- La paridad nunca se guarda en los discos que contienen los datos que han generado dicha paridad, ya que, en el caso de que uno de ellos se estropeara, bastaría con regenerar los discos para que el dato pudiera volver a restablecerse.
- Ventajas: alto rendimiento en aplicaciones con gran demanda de velocidad. No se desaprovecha ningún disco exclusivamente para almacenar códigos de paridad. Se pueden recuperar los datos.
- Inconvenientes: bajo rendimiento en escrituras. Se requiere un mínimo de tres discos.



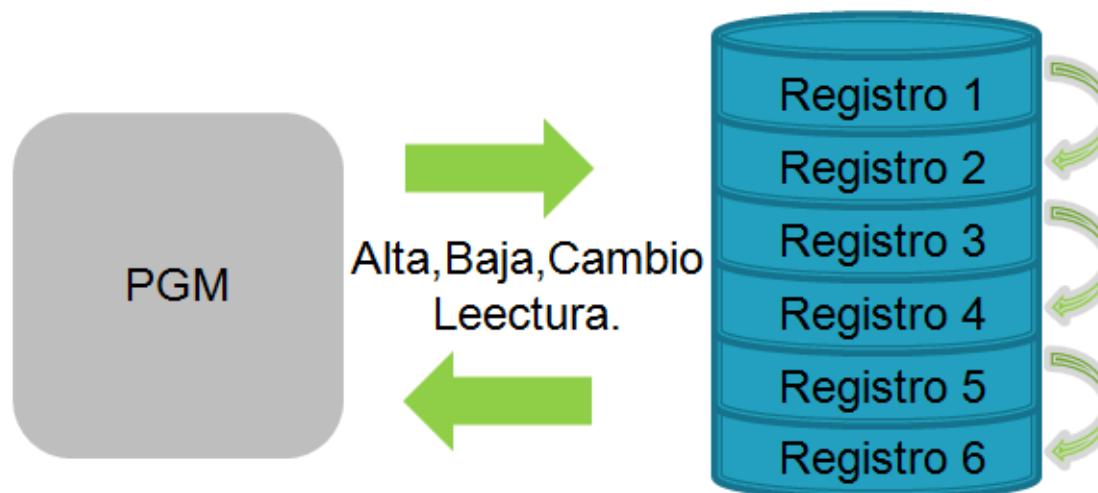
Algoritmos de planificación de peticiones

- Los algoritmos de planificación de peticiones de lectura y escritura a discos se encargan de registrar dichas peticiones y de responderlas en un tiempo razonable.
- Los algoritmos más comunes para esta tarea son:
 - Primero en llegar, primero en ser servido (FIFO)
 - Primero el más cercano a la posición actual
 - Por exploración (algoritmo del elevador)
 - Por exploración circular

Métodos de acceso en los sistemas de archivos

■ Acceso secuencial

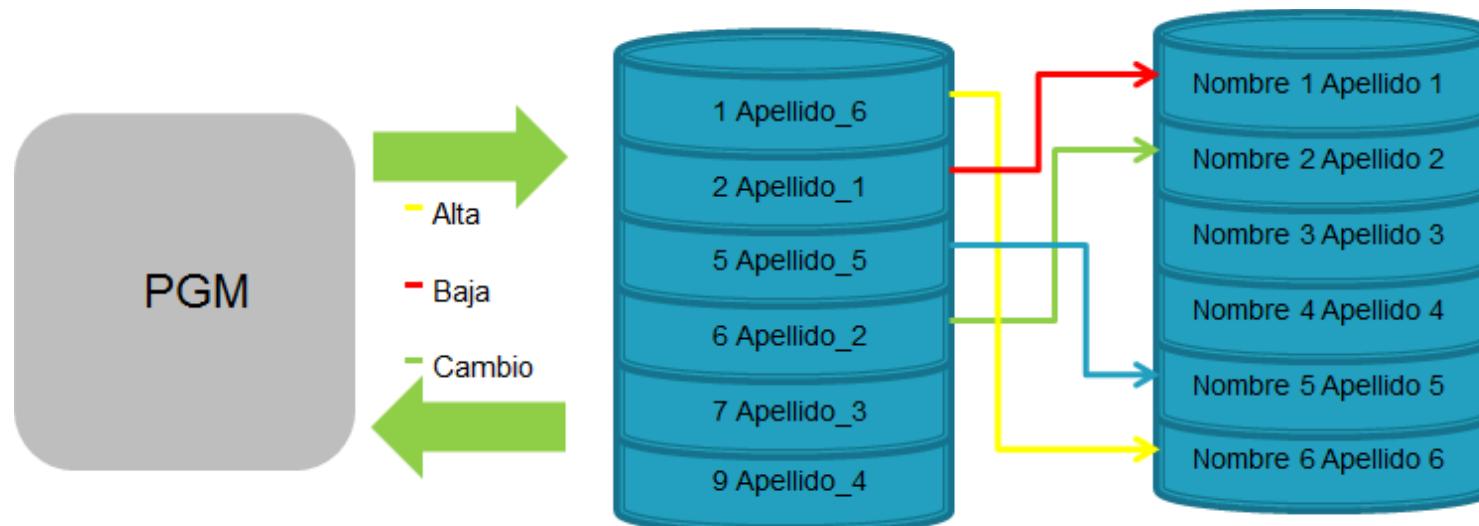
- Es el método más lento y consiste en recorrer los componentes de un archivo uno en uno hasta llegar al registro deseado.
- En un archivo organizado secuencialmente, los registros quedan ordenados de manera consecutiva cuando se utiliza el archivo como entrada.



Métodos de acceso en los sistemas de archivos

■ Acceso secuencial indexado

- Permite accesar cualquier sector o registro inmediatamente, por medio de llamadas al sistema como la de seek.
- Es la combinación de tipos de acceso que manejan los archivos secuenciales y un archivo relativo o de acceso directo.



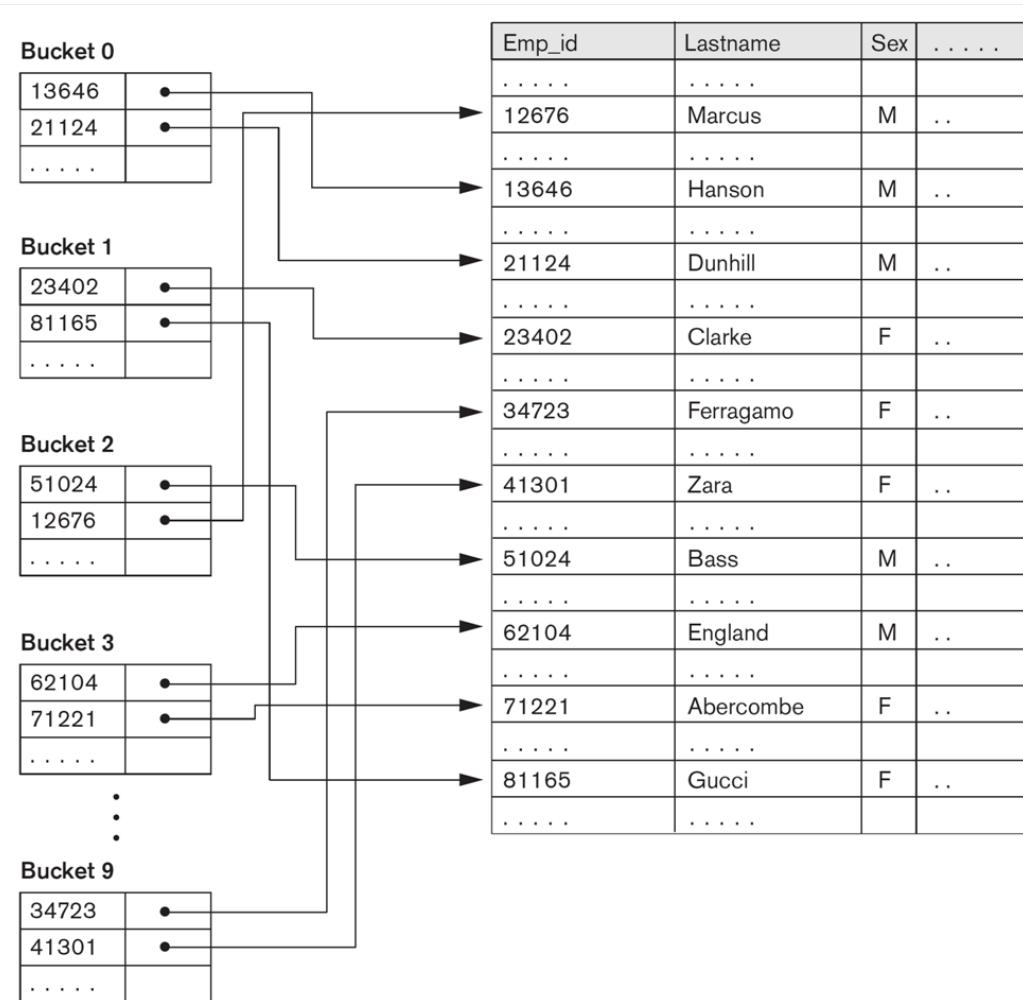
Métodos de acceso en los sistemas de archivos

■ *Acceso directo indexado*

- Es útil para grandes volúmenes de información o datos.
- Consiste en que cada archivo tiene una tabla de apuntadores, donde cada apuntador va a la dirección de un bloque de índices, lo cual permite que el archivo se expanda a través de un espacio enorme.

Métodos de acceso en los sistemas de archivos

■ Acceso directo indexado



Sistema de Archivos

- Un sistema de archivos consta de tipos de datos abstractos, que son necesarios para el almacenamiento, organización jerárquica, manipulación, navegación, acceso y consulta de datos.
 - FAT
 - FAT32
 - HSF
 - HSF+
 - EXT2
 - EXT3

Operaciones soportadas por el subsistema de archivos

- Crear (create)
- Borrar (delete)
- Abrir (open)
- Cerrar (close)
- Leer o Escribir (read, write)
- Concatenar (append)
- Localizar (seek)
- Leer atributos
- Poner atributos
- Renombrar (rename):

Organización Secuencial

- Los registros se almacenan físicamente en secuencia (uno detrás del otro).
- El acceso es muy rápido, sin embargo para localizar un registro en particular, se tiene que leer todos los registros anteriores.
- Existen dos métodos de almacenar los datos con este tipo de Organización:
 - Sin ningún orden en particular
 - Organizado por un campo clave

Organización Secuencial

Ventajas

- Uso eficiente de disco y memoria con un mínimo de “overhead”.
- El acceso más rápido cuando se requiere leer los registros secuencialmente.
- Mejor uso de archivos con múltiples registros. También permite múltiples formas de guardar los datos.
- Ideal cuando se desea acceder sobre el 80% de los registros que residen en el archivo.

Organización Secuencial

Desventajas

- Para poder llegar a un registro, se tiene que leer primero todos los anteriores.
- Sólo permite que los registros se añadan al final del archivo.
- Si el archivo esta ordenado por algún campo en particular, se requiere de un proceso adicional de reorganización.

Organización Directa o Relativa

- Un registro se almacena en una dirección basado en una fórmula que utiliza de referencia el campo clave (Key) el cual debe ser numérico.
- Se debe crear el archivo con los registros en blanco inicialmente en esta estructura.
- Se accede a través de un número secuencial el cual indica la localización física del registro.
- Se utiliza dos tipos de Organizacion Directa:
 - Key Addressing Techniques
 - Hashing Techniques

Organización Directa o Relativa

Ventajas

- Permite acceso tanto secuencial como directo de los registros.
- Permite insertar y eliminar registros de forma aleatoria.
- Provee la forma más rápida de poder llegar a un registro directamente, por lo tanto es la forma ideal si se desea llegar a un registro en particular.

Organización Directa o Relativa

Desventajas

- Los registros tienen que ser del mismo largo.
- La programación para manejar este tipo de archivo es compleja, sobre todo en el manejo de las colisiones.
- Generalmente requiere que se cree y separe el espacio antes de insertar los registros, esto consume espacio en disco que no se utiliza inmediatamente.

Organización Indexada

- Se compone de un archivo con datos y otro de índices (pueden estar juntos).
- El archivo de índices tiene el campo clave (key) de cada registro y su localización en el archivo principal.
- Provee una ventaja al cotejar si un record en particular existe ya que se puede lograr utilizando solo el índice únicamente.
- La organización del archivo se hace en el índice. Si se quiere tener acceso al archivo por otro campo, se puede crear otro índice que contenga ese campo y la localización del registro.
- Es más rápido leer el índice que el Archivo, por lo que ciertas búsquedas pueden ejecutarse más rápido que en otras estructuras.
 - Indexed Random (nonsequential) Organization
 - Indexed Sequential Organization

Organización Indexada

Ventajas

- Permite el acceso secuencial y directo (por medio de una *key*). Puede manejar múltiples índices, lo que le permite buscar por varios campos.
- Fácil de implementar la programación que maneja este tipo de archivo. No se tiene que programar el manejo de colisiones ni actualizar parámetros del archivo.
- Permite actualización del largo de los campos y de los índices.
- Fácil el manejo para insertar, modificar y eliminar registros.

Organización Indexada

Desventajas

- Requiere de espacio en disco adicional para poder manejar su estructura. (índices)
- A mayor cantidad de índices que se incluyan, mas espacio en disco ocupa y más lento se convierte el proceso de insertar y eliminar.
- Requiere de un uso mayor de los recursos (memoria RAM y procesador).