

# Ficheros en MVS (I)

En el numero de diciembre de *Solo Programadores* fueron publicados dos artículos distintos sobre ficheros, uno estaba dedicado a los ficheros en DOS y otro a los ficheros en UNIX. Este hecho sirvió de estímulo al autor de este artículo para tocar el mismo tema ya que no solo considera interesante sino necesario que, a nivel de programador, se conozcan las herramientas que pone a su disposición el S.O. para manejar la información, que en el entorno de los grandes sistemas puede llegar a ser muy voluminosa.

## Introducción

En cualquier sistema operativo, la información se encuentra agrupada en conjuntos homogéneos, recibiendo cada uno de estos conjuntos la denominación de fichero, archivo o dataset.

En MVS, la etiqueta o nombre de un fichero se forma agrupando calificadores y separando estos con un punto. Cada uno de los calificadores puede tener una longitud máxima de 8 caracteres alfanuméricos y no puede empezar ni por dígito ni por un carácter especial, con excepción del \$ que es valido.

La longitud máxima del nombre de un fichero no puede superar los 44 caracteres, y debe tener al menos dos calificadores, exceptuando los ficheros temporales, cuyo nombre se forma anteponiendo un & a un único calificador.

El nombre del miembro de un fichero particionado (tipo de fichero que se describe posteriormente), se forma escribiendo el nombre del fichero, de acuerdo con las reglas descritas anteriormente, y entre paréntesis el nombre del miembro a procesar, que solo podrá tener hasta 8 caracteres.

Otro termino contemplado en MVS es el ALIAS para los nombres de ficheros, siendo estos *ALIAS* nombres alternativos dados a ficheros. Para las versiones anteriores al MVS/ESA este nombre alternativo solo podía especificarse para el primer cualificador. En cambio, con esta versión de MVS, el ALIAS puede referenciar los cuatro primeros calificadores si así se desea.

Estos ALIAS residen en el catalogo maestro, termino que también se vera posteriormente, y apuntan a un determinado catalogo de usuario, donde estará ya la ubicación física de un determinado fichero.

## DEFINICIONES

Antes de entrar en materia, conviene recordar algunas definiciones de términos comunes en estos temas:

Soporte: Lugar donde se encuentra almacenada la información.

Campo: Unidad elemental de información procesable de un registro.

Es la unidad mas pequeña de información que nos puede interesar.

Registro Lógico: Conjunto de campos. Es la unidad básica de información con significado propio para el usuario o informático.

Bloque o Registro físico: Es la unidad de transferencia entre los dispositivos y la memoria, o viceversa.

Factor de bloqueo: Numero de registros lógicos contenidos en un registro físico, o bloque.

Fichero o archivo: Conjunto de registros lógicos referidos a un mismo tema

Volumen: Conjunto de información a la que se puede acceder en un instante determinado, en un cierto dispositivo.

## Características de los ficheros que influyen en su organización

Tamaño: viene determinado por el número de registros de que consta. Para este concepto debe tenerse en cuenta, no solo el tamaño actual, sino el previsible.

Volatilidad: es el número de altas y bajas que se procesan en cada pasada del fichero. Por tanto, un fichero que sea muy volátil deberá ser reorganizado muy a menudo.

Actividad: La actividad es la frecuencia con la que se actualiza o referencia un fichero. . Así, si un fichero sufre muchas actualizaciones, diremos que es muy activo.

Magnitud de actividad: en un proceso por lotes se denomina magnitud de actividad al número de pasadas por unidad de tiempo.

Porcentaje de actividad: es el porcentaje de registros que se actualizan o refieren en cada pasada. Cuando el número de transacciones a realizar sea mayor del 10 % no compensa realizar procesos al azar, sino que es mejor hacer un proceso batch secuencial.

Distribución de actividad: Es la distribución que se hace de los registros para que tengan mayor posibilidad de acceso, y por lo tanto mayor rapidez para acceder a los registros. La magnitud de actividad y el porcentaje son inversamente proporcionales. Cuanto mas pasadas se hagan, será menor el número de registros referenciados.

## Unidad de disco magnético

Así mismo, y también como introducción al tema, conviene recordar la composición y estructura física de las unidades de disco, ya que esta influye muy directamente en la definición del fichero como se vera en los siguientes apartados.

Una unidad de discos es un *armario* en el que se coloca uno o varios dispack, normalmente 2; estando cada dispack constituido por varios discos concéntricos.

Para leer o grabar información en un disco, se utilizan las cabezas de lectura-grabación. Estas cabezas se encuentran en los extremos de los brazos de acceso, y estos a su vez se encuentran anclados a un peine fijo, tal y como se muestra en la **figura 1**.

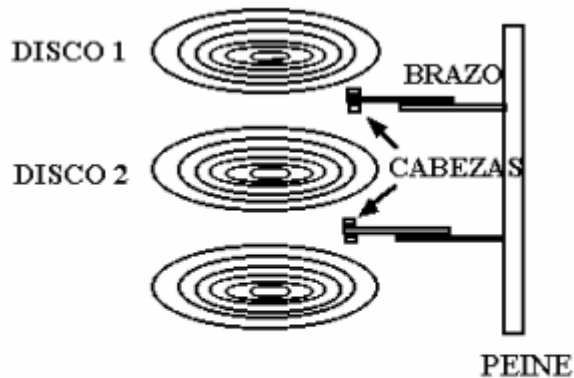
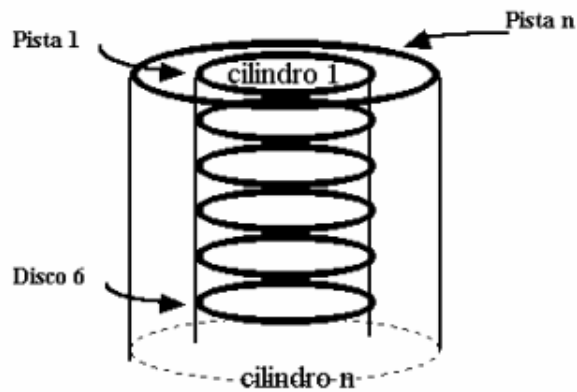


figura 1: DISPACK

Entre dos discos entra un brazo de acceso con dos cabezas (puede existir mas de una cabeza por pista). Todos los brazos de acceso se mueven longitudinalmente a la vez, con lo que todas las cabezas se encuentran en todo momento en la misma vertical. Este hecho nos lleva a definir el concepto de cilindro.

Cilindro: es el conjunto de todas las pistas a las que se puede acceder teniendo el mecanismo de acceso en una posición determinada; o también, conjunto de pistas a las que se puede acceder para un determinado movimiento de brazo. (**figura 2**)



**Figura 2: Cilindro y Pista**

Cada disco tiene un número determinado de pistas concéntricas, las cuales tienen la misma capacidad, por lo que varía la densidad en cada una de ellas.

La información se graba por cilindro, no por superficie, ganando de esta forma *tiempo de posicionamiento* de las cabezas. Y, dentro de cada pista, la información se graba en serie, bit a bit.

Existen, como el lector puede suponer, diferentes modelos de discos entre los que se pueden citar a modo de ejemplo los siguientes:

Unidad IBM 2311: cada dispack tiene 6 discos, de los cuales son útiles solo 10 superficies de las 12 de que dispone. No se graba información ni en la cara superior del conjunto ni en la inferior, porque es de suponer que en estas caras se concentran impurezas del ambiente, tales como polvo, rayas, etc.

Tiene 203 pistas, y en consecuencia 203 cilindros.

La capacidad de una pista es de 3625 octetos y por tanto la de un cilindro será  $3625 * 10$

Unidad IBM 3330: Cada dispack o pila de discos contiene 11 discos, de los cuales solo se utilizan 19 superficies.

Cada superficie tiene 411 pistas siendo la capacidad de una pista de 13030 octetos

Unidad IBM 3340: En esta unidad el dispack, que se compone de 6 discos, es hermético por lo que se utilizan las 12 superficies de los discos.

El modelo M35 de esta Unidad dispone de 349 pistas por superficie, siendo la capacidad de una pista de 8368 octetos.

El modelo M70 contiene 701 pistas por superficie, y la capacidad por pista es de 8368 octetos.

## Organización de ficheros

Método de organización : Es el conjunto de técnicas establecidas para almacenar la información soportada en un volumen o soporte de información. Existen 5 métodos, tal y como se muestra en la figura 3, utilizándose uno u otro, por lo general, en función de cómo las aplicaciones van a tratar la información contenida en los distintos ficheros.

Método de acceso: Es la técnica utilizada para almacenar, localizar y recuperar de un soporte la información automáticamente. También hay 5 tipos, dependiendo de la organización del soporte.

El cuadro de la **figura 3** muestra la relación entre métodos de organización y métodos de acceso.

Metodo de organizacion	Metodo de acceso
<b>Secuencial</b>	SAM (Sequential Access Method)
<b>Secuencial indexado</b>	ISAM (Indexed Sequential Access Method)
<b>Directa</b>	DAM (Direct Access Method)
<b>Particionada</b>	PART (Particionado)
<b>VSAM</b>	VSAM (Virtual Storage Access Method)

Figura 3

## Formato de los registros

Los archivos en MVS pueden estar formados por registros de longitud FIJA (LF) o VARIABLE (LV), yendo en este caso precedido por 2 bytes, en binario puro, indicando la longitud del registro.

Así mismo, tanto unos como otros pueden agruparse o no en bloques, para grabarse en el volumen como BLOQUEADO (B) o NO BLOQUEADO (NB).

Por ultimo, existe otro formato, INDEFINIDO, que es usado normalmente para el almacenamiento de los módulos ejecutables.

Existen por tanto las 5 modalidades siguientes de formatos de registro: RLFNB, RLFB, RLVNB, RLVB y RLU.

## VTOC: (Volume Table Of Contain)

El VTOC es un archivo especial que obligatoriamente se encuentra en todos los volúmenes de la instalación, y que contiene los nombres de todos los ficheros que se encuentran en ese volumen. Este fichero recibe también la denominación de archivo de etiquetas y se encuentra en un área específica del volumen. Se podría comparar con la FAT (File Allocation Table) de los diskettes del MS-DOS.

El propio VTOC, por ser un fichero, también deberá estar referenciado dentro de sí mismo y tener por tanto su propia etiqueta, que precisamente se llama VTOC.

Esta etiqueta constituye el primer registro del fichero VTOC, el cual tiene un registro con formato especial, pues contiene información propia del VTOC. Este fichero puede estar en cualquier parte del volumen, y consta de un numero entero de pistas (dentro de un cilindro) llegando como máximo a un cilindro. Las pistas no son compartibles.

Este archivo se crea al inicializar cada volumen, pues el programa que realiza esta función, entre otras cosas se encarga de:

- Grabar la dirección del VTOC en la etiqueta de volumen
- Grabar la etiqueta de VTOC en la primera posición del VTOC
- Generar la segunda etiqueta del VTOC como etiqueta de espacio disponible.
- Reservar el espacio o área restante del VTOC.

Los registros del VTOC son registros con clave formados por los dos campos o áreas siguientes:

- área de clave de 44 octetos que contendrá el nombre del fichero.
- área de datos de 96 octetos, que contendrá Características, estadísticas y el puntero a comienzo del fichero.

Existen distintos formatos de registros de VTOC, identificados por los valores claves F1 a F6, pero que si bien contienen información importante del VTOC, no es relevante para el objetivo de este artículo.

## Organización SECUENCIAL

En esta organización, también llamada CONSECUTIVA, la información se almacena colocando cada registro uno a continuación de otro, como su propio nombre indica, y tiene la particularidad de que para recuperar una información se debe leer necesariamente toda la anterior.

Este tipo de organización es una organización totalmente compacta, es decir, sin huecos entre sus registros.

Los formatos admitidos son de longitud fija y variable, bloqueados y no bloqueados.

Procesos que se pueden realizar con estos ficheros:

Carga: La carga se debe realizar mediante procesos secuenciales.

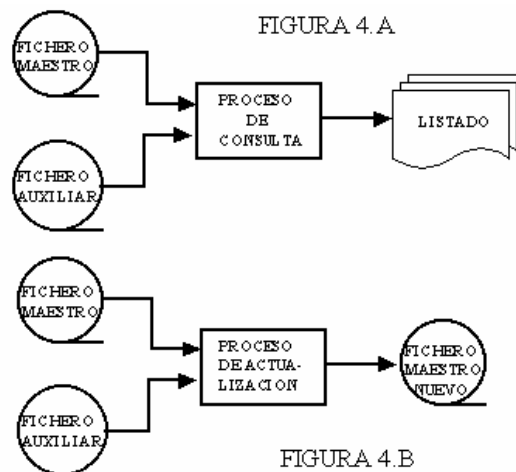
Recuperación: La recuperación de un registro supone la lectura consecutiva de todos los registros que preceden al registro a recuperar, debiendo realizarse mediante procesos secuenciales. Lo normal es que los registros lleven secuencia de clave.

Existe un índice que sirve para medir el tiempo medio de acceso a un registro. Es el tiempo de registro, y se define como  $t=T/n$

donde  $t$  representa el tiempo medio de proceso de registro  
 $T$  representa el tiempo total de proceso de fichero  
 $n$  representa el número total de registros.

Actualización: Los procesos de Alta - Baja y Modificación requieren el uso de un fichero auxiliar sobre el que se irán reflejando las actualizaciones que se requieran, y que deberá ser consultado en cada uno de los accesos que se realicen al fichero maestro a fin de determinar si los datos del registro principal han sido modificados (FIGURA 4-A). Esta consulta será necesario realizarla hasta que se realice el proceso de actualización, mediante el cual se fusionara el fichero maestro y el fichero auxiliar, generándose un fichero nuevo actualizado, para posteriormente borrar o pasar a histórico el maestro anterior (FIGURA 4-B).

Cuando se da secuencialidad a un fichero mediante el uso de una clave, se pasa de una organización consecutiva a una organización secuencial.



## Organización SECUENCIAL-INDEXADA

Este tipo de organización tiene las siguientes particularidades:

- Todos los registros se encuentran clasificados en orden ascendente de clave.
- Para poder recuperar un registro determinado mediante un proceso al azar, es preciso realizar un estudio jerárquico de los índices que se definan. Estos índices son unos registros especiales que contienen la información necesaria para permitir el acceso directo al registro sin necesidad de recuperar todos los registros previos.

Antes de tratar esta organizacion, y con objeto de facilitar la explicación correspondiente, el autor propone el siguiente ejemplo:

Imaginemos un diccionario que por ser demasiado voluminoso, se encuentra dividido en tres tomos, figurando en el lomo de cada tomo la primera y ultima palabra descrita en el mismo (**figura 5**)

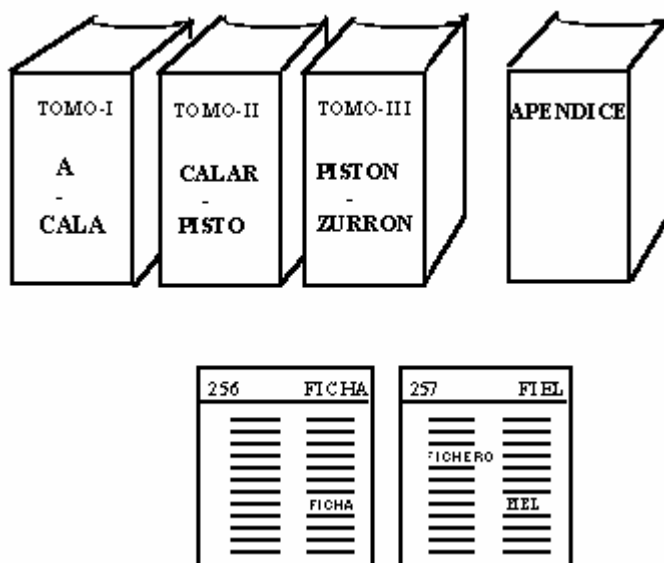


Figura 5

Por su parte, cada pagina de ese diccionario dispone en su parte superior de la ultima palabra que se encuentran en esa pagina.

Por ultimo, también existe un cuarto volumen, el APÉNDICE, en el cual debemos buscar aquella palabras que se han definido con posterioridad a la edición del diccionario.

Pues bien, si el lector quisiera conocer el significado de la palabra *FICHERO*, debería buscar en el lomo del Tomo-I y al comprobar que CALA, alfabéticamente, es anterior a FICHERO, desecharía dicho volumen y pasaría a examinar el Tomo-II, que como contiene las palabras desde CALAR a PISTO debe incluir la que se busca.

A continuación esa misma selección se haría con las paginas, no fijándose en el contenido de las mismas sino en su índice superior, hasta que llegue a la pagina 257, en la cual resulta que FIEL es la primera palabra del índice que es posterior a FICHERO, y como a su vez FICHA es anterior, resulta que la definición buscada está en esa pagina 257.

Una vez seleccionada la pagina, volvemos a hacer la misma operación, pero esta vez con las palabras de la pagina, hasta encontrar FICHERO.

Pues bien, con la organización secuencial indexada, el propio sistema en tiempo de generación del fichero, cuando va colocando la información en el área de datos reservada al efecto, va generando en el área de índices unos índices jerarquizados de acuerdo con el campo clave de esa información y que, al igual que en el ejemplo visto, consta de los siguientes índices:

- Índice MAESTRO : representaría el índice de la biblioteca.
- Índice de CILINDRO: representado por la información contenida en el lomo de cada tomo
- Índice de PISTAS: Estaría representado por las palabras de cabecera de cada pagina.

Y por ultimo, el dato buscado se encontraría examinando secuencialmente la clave de todos los registros de información contenidos en ese bloque o pagina seleccionada.

Si al tiempo que la Real Academia va creando una palabra, pudiera poner una marca en el hueco donde debiera ir para remitirnos al apéndice, nos habría resuelto el tener que explicar el hecho de que el índice de pistas tenga dos entradas, :

- la entrada normal; y
- la entrada de excedentes que se vera mas adelante.

## Índices

Los índices son ficheros especiales, compuestos por unos registros llamados entradas y que siempre estarán en formato desbloqueado aunque el área de datos este bloqueado.

Estas *entradas* tienen dos campos o áreas: Área de clave y área de datos.

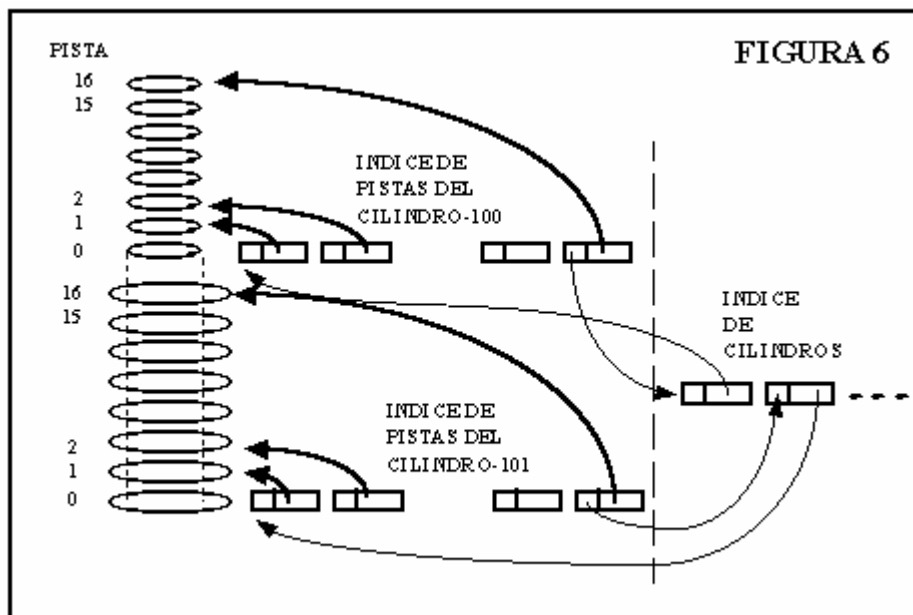
En el área de índices, la clave tendrá la misma longitud que en el área de datos; y en el área de datos del registro de índices figurará un puntero con una longitud de 10 octetos.

Los índices son generados por el método de gestión de los datos de forma automática, es decir, el método ISAM va generando los índices al tiempo que se crea el fichero.

Como ya se ha dicho anteriormente, este método cuenta con los TRES ÍNDICES siguientes:

El índice de pistas es obligatorio e ira siempre en la pista 1 de cada cilindro asignado al fichero, reservándose el resto del cilindro para datos del fichero, y dejando normalmente al ultima pista como área de overflow.

La información que lleva este índice de pistas hace referencia a cada una de las pistas de ese cilindro, tal y como se puede apreciar en la **figura 6**.



El índice de pistas es el de mas baja jerarquía, pues remite directamente a la pista donde se encuentra la información. Un fichero puede tener asignados una o varias pistas para ser utilizadas como índices de pistas de cada cilindro, aunque lo normal es que haya solo una.

Cada pista del área primaria genera dos entradas en el índice de pistas:

- una entrada normal que contendrá la dirección de la pista y la clave mas alta de esa pista; y
- una entrada de excedentes, que contendrá la entrada de excedentes, es decir en el área de clave figurará la clave mas alta de la cadena asociada, y en el área de datos, un puntero con la dirección de la pista(no del registro) donde se encuentra el registro de clave mas baja de la cadena de excedentes asociada a la pista de entrada normal.

Índice de cilindros: Este índice también es obligatorio. Contiene una entrada por cilindro, y puede grabarse en cualquier pista del dispack, o incluso en otro dispack.

La información que lleva es:

- en el área de clave la clave mas alta del cilindro; y
- en el área de datos, la dirección del cilindro.

Índice maestro: Este índice es optativo. Es un índice de pistas de un índice de cilindros. Se utiliza cuando el índice de cilindros ocupa mas de 4 pistas.

Las entradas de este índice contienen la siguiente información:

En el área de clave: la clave mas alta de la pista del índice maestro

En el área de datos: la dirección de esa pista.

### **Cadena de excedentes**

Los registros en el área de overflow se encuentran formando cadenas de excedentes.

Una cadena de excedentes es un conjunto de registros que deberían pertenecer a una pista determinada de datos, pero que por haberse dado de alta con posterioridad a la carga inicial del fichero se han encontrado sin espacio físico donde ubicar esa información. Esto quiere decir que una cadena de excedentes esta asociada a una pista de datos.

Los registros en el área de overflow se encuentran en formato desbloqueado, y el propio sistema operativo añade a cada registro un campo de enlace o puntero de 5 octetos, en el cual coloca la dirección del siguiente registro lógico, o 'FF' hexadecimales si es el ultimo registro de la cadena.

En tiempo de creación, las áreas de overflow están vacías. Con el uso, al irse llenando, la flexibilidad del fichero disminuye. En el caso de que se omita en tiempo de creación la definición de este área, el sistema reserva un determinado numero de pistas del cilindro.

En el área de overflow se colocan, no los registros nuevos, sino los que sobran de las pistas del área primaria al colocar en su sitio al registro nuevo.

Si por casualidad el área de overflow de un cilindro quedara llena, y hubiera que colocar un registro excedentario, este se colocaría en el área de overflow independiente.

Hay que tener en cuenta que en el caso de que una pista de datos cuente por ejemplo con tres registros: de clave 19, 45 y 87 en el índice de pistas viene la clave del registro mas alto y la dirección del registro mas bajo, es decir: 87-d(19)

También hay que tener en cuenta que en el área de overflow, los registros cuentan con un campo llamado campo de enlace, en el cual figura la dirección del siguiente registro lógico.

La información contenida en el campo de enlace la genera y la trata el propio sistema.

### **Próxima entrega**

Evidentemente el tema es largo por lo que debe quedarse para la próxima entrega las Organizaciones Directa y VSAM, así como unos ficheros especiales que juegan un papel muy importante en el acceso a la información de los ficheros: son los Catálogos. Por ultimo, también se comentarán los ficheros GDG y si es posible, ejemplos de definición de cada uno de los tipos descritos.