

# Organización de Datos – Lic. Servetto

## CASOS DE ARCHIVOS SECUENCIALES INDICADOS

### SCOPE

Los archivos **SIS (SCOPE Index Sequential)** de Control Data Corporation se constituyen de bloques de almacenamiento secuencial de registros, indicados por pares clave/dirección. La clave es la **menor** del grupo de registros almacenados en un bloque. La dirección es el número relativo del bloque en el archivo. Estos indicadores de bloques de datos, también organizados en bloques, a su vez pueden estar indicados por otro índice análogo, y este último por otro, y así siguiendo, de acuerdo a la dimensión de los bloques de datos y de los bloques del índice hasta formar una pirámide de indicación o índice en cascada en la que cada nivel es un índice al nivel siguiente. Un archivo SIS puede tener hasta 63 niveles.

Esta organización de archivos no prevee áreas de desborde para la inserción de registros. Los registros se incorporan en el área primaria de almacenamiento (bloque), en espacio que se deja a tal efecto en el momento de la carga o reorganización. Cuando un bloque se llena y no puede aceptar más registros, se divide.

### Ejercicio

a) Cargue al 60% de su capacidad un archivo tipo SIS con bloques de almacenamiento de registros de factor de ablocamiento 5 y bloques de indicación de factor de indicabilidad 2, con los registros cuyas claves se consignan a continuación:

**173, 195, 234, 275, 290, 379, 401, 490, 503, 579, 620, 683, 704, 780, 826, 904, 950, 983, 1013, 1066, 1074, 1215, 1340, 1390**

Grafique el estado del archivo de datos y de su índice, designando a los registros por sus claves, y utilizando como punteros a bloques de datos los símbolos  $D_i$  ( $i=0, 1, 2, \dots$ ) y como punteros a bloques de indicación los símbolos  $I_j$  ( $j=0, 1, 2, \dots$ ).

b) ¿Cómo quedaría el archivo de datos tras la incorporación de los registros con claves 181, 204 y 175? ¿Y el índice? Discuta alternativas a este respecto.

### VSAM

En los archivos VSAM (Virtual Storage Access Method, método de acceso de almacenamiento virtual), de IBM, los registros de datos se agrupan y almacenan en unidades lógicas denominadas intervalos de control o trenes, compuestas por un número fijo de bloques. Los registros pueden ser tanto de longitud fija como de longitud variable y pueden extenderse entre bloques de un mismo intervalo.

La asignación de espacio de almacenamiento para estos archivos se efectúa por áreas de control o porciones, consistentes de un número fijo de intervalos de control. Cuando se requiere específicamente, el ablocamiento de registros de datos puede ser extendido entre intervalos de una misma área de control.

Los archivos VSAM pueden definirse como ESDS (Entry Sequenced Data Set, conjunto de datos con entrada secuencial), para acceder a los registros por su RBA (Relative Byte Address, byte de dirección relativa) que es el desplazamiento del registro desde el inicio del espacio asignado al archivo; como RRDS (Relative Record Data Set, conjunto de datos de registro relativo), para acceder a los registros por su posición relativa, cuya obtención es responsabilidad del usuario, o como KSDS (Key Sequenced Data set, conjunto de datos por secuencia de claves), para acceder a los registros por su clave y para los que el VSAM lleva un índice de control.

### Organización VSAM KSDS

En los archivos definidos como KSDS, cada área de control está indicada por un intervalo de control del índice que contiene un par clave/dirección para cada intervalo del área. La clave es la mayor del intervalo y la dirección es la posición del primer bloque del mismo.

Las áreas de control pueden tener intervalos vacíos, cuyos punteros también se mantienen en su respectivo intervalo del índice.

Los intervalos no vacíos de un área de control contienen un número variable de registros. Los registros efectivos de un intervalo se disponen secuencialmente en su área inicial y se distinguen en forma precisa a partir de marcadores de tres bytes almacenados en orden reverso desde el final del intervalo. Los registros adyacentes de igual tamaño se controlan mediante un sólo marcador doble. Cada intervalo dispone además de un campo de control general que determina la posición y extensión de su espacio libre.

La estructura formal de un intervalo de control es entonces:

**$R_1, R_2, \dots, R_h, EL, RhDF, R(h-1)DF, \dots, R1DF, C1DF$**

suponiendo registros de longitud variable, y en donde  $R_j$  representa al registro  $j$ -ésimo dentro del intervalo ( $j=1, 2, \dots, h$ ),  $EL$  al espacio libre,  $R_jDF$  al campo para la definición del  $j$ -ésimo registro (Record Definition Field) y  $C1DF$  al campo para la definición del intervalo de control (Control Interval Definition Field).

El CIFI se compone de dos pares de bytes cuyos valores binarios representan respectivamente el desplazamiento hasta el espacio libre en el intervalo de control y la cantidad de bytes no utilizados en el mismo.

Un RDF se compone de tres bytes, de los cuales

- del primero (byte 0), sus bits, numerados de 1 a 8 significan:

BIT	VALOR	SIGNIFICACION
1	0	nada
2	1	emparejamiento de dos RDF para contener juntos la información
	0	caso contrario
3 y 4	00	el registro no ocupa dos o más intervalos
	01	es el primer segmento de un registro expandido
	10	es el último segmento de un registro expandido
	11	es un segmento intermedio de un registro expandido
5	0	los bytes 1 y 2 especifican la longitud de un registro
	1	los bytes 1 y 2 especifican el número de registros consecutivos con la misma longitud
6	0	slot vacío (sólo para los RRD)
	1	caso contrario
7 y 8	00	nada

- los bytes 1 y 2 contienen la longitud del registro de datos o el número de registros consecutivos con la misma longitud

La inserción de registros se realiza en los intervalos de control que les correspondan según el ordenamiento por clave. Para ello se explota el espacio libre fluctuante en los mismos intervalos, destinado a aceptar no sólo a los registros sino también a sus marcadores. En caso de que el espacio libre en un intervalo no alcance para albergar un nuevo registro, la inserción se soluciona capturando un nuevo intervalo vacío del área subyacente para disponer la mitad del desbordado en él. Esta operación de excepción se realiza de la siguiente manera, en la mayor atención a la integridad de los datos:

- Se copian los registros correspondientes a la primer mitad del intervalo lleno en el intervalo nuevo, tras lo cual, previo establecimiento de los marcadores y actualización del campo de control, este último se graba en disco.
- Se actualiza el índice apareando al puntero del intervalo recientemente capturado con la clave del último registro que pasó a residir en él.
- Se corren los registros de la segunda mitad del intervalo original hacia el principio eliminándose a los registros ya copiados en el nuevo intervalo junto con los respectivos marcadores, y luego el intervalo particionado se reescribe.

A todo esto el registro de maras se agrega en el momento oportuno, según le correspondiere luego de la escisión.

Si ante la necesidad de particionar un intervalo no existiere ningún otro vacío en la misma área de control, este problema se arreglaría procediendo a la definición de una nueva área de control y pasándole a ella la mitad de los intervalos de la vieja (aquellos indicados por la primera mitad de las entradas del registro de índice).

Para evitar que tras la creación o reorganización del un archivo KSDS las inserciones produzcan inmediatamente particionamientos de intervalos y áreas de control, se lo carga con densidad menor a 1. Si, por ejemplo, se cargan los intervalos al 60% y las áreas al 90%, la utilización inicial de espacio rondará el  $0.60 \times 0.90 = 54\%$ ; y tras una cantidad de actualizaciones equivalente al 40% de la carga inicial, suponiendo inserciones aleatorias, la utilización esperada del espacio ascendería a  $1.40 \times 0.54 = 75\%$ .

Respecto a los índices de control, éstos pueden tener a lo sumo tres niveles, y la estructura de sus intervalos es:

#### **Cabecera, punteros a intervalos libres, EI, Sn, ..., S2, S1, RDF, CIFI**

En donde el campo Cabecera describe el resto del formato del intervalo, EI representa al espacio inutilizado en el intervalo, Si a la i-ésima sección de indicación ( $i=1, 2, \dots, n$ ), RDF al campo de definición del registro de índice (se considera que todo intervalo contiene un único registro de índice de longitud variable), y el CIFI es igual que en los intervalos del archivo de datos.

El campo Cabecera incluye:

- la longitud del registro
- el tamaño de los punteros a intervalos (puede variar según sean a intervalos de datos o a intervalos de índice)
- el número de nivel de indicación (1, para el nivel de indicación a registros de datos, 2 o 3)
- un puntero de encadenamiento secuencial al próximo intervalo de indicación (sólo válido para el nivel 1)

- el desplazamiento hasta el espacio inutilizado en el registro
- el desplazamiento hasta la última entrada de índice de la última sección (Sn)
- el desplazamiento hasta la última entrada de índice de la primera sección (S1)

Las secciones de indicación no son más que grupos de entradas de índice, localizados al efecto de acelerar las búsquedas dentro de los intervalos. El número n de secciones en un intervalo resulta de la raíz cuadrada del número total de entradas iniciales del intervalo.

Las estructuras de secciones y entradas son:

**Sección i: di, Er, ..., E2, E1**

en donde di representa la distancia entre la última entrada de índice de la sección (entrada r), y la última de la sección siguiente (sección i+1).

**Entrada j: Clave Abreviada, F, L, Puntero a Intervalo**

en donde la clave abreviada es de longitud variable y los campos F y L, ambos de un byte de longitud, describen la amplitud de la abreviatura. El campo F indica el número de caracteres omitidos de la parte inicial de la clave, y el campo L indica el número de caracteres restantes luego de la abreviación (longitud de la clave abreviada). Por ejemplo, supóngase que las claves del último registro de los intervalos de control de datos apuntados por una sección de índice son:

**Berringer, Bigley, Breslow, Bretshnev, Brody, Bruckner, Buhler, Calhoun, Call, Cross**

luego las correspondientes entradas serían:

Clave Abreviada	F	L	Cotas del intervalo
Be	0	2	hasta Bez
i	1	1	Bfa... a Biz
res	1	3	Bja... a Bresz
t	3	1	Breta... a Bretz
o	2	1	Breua... a Broz
	2	0	Brpa... a Brz
u	1	1	Bsa... a Buz
Calh	0	4	Bva... a Calhz
	3	0	Calia... a Calz
	1	0	Cama... a Caz

Cuando F es 0, la parte inicial de las claves se reconstruye a partir de las entradas previas, por lo que la clave correspondiente a la última entrada de una sección no debe estar abreviada en esa forma: la búsqueda de una entrada dentro de un registro de índice se efectúa en un primer paso saltando secciones, por medio del indicador de longitud de sección di, y cuando se halla la sección adecuada, en un segundo y último paso se busca dentro de la sección exhaustivamente.

Los archivos KSDS pueden tener también accesos alternativos, por cualquier otro atributo distinto de la clave primaria que se especifique. Para esto se utiliza una estructura de inversión del archivo KSDS original, organizada mediante el empleo de un archivo KSDS adicional para cada índice alternativo, en el que los registros de datos, de longitud variable, son las llamadas listas invertidas. Estas listas se componen del valor de la clave alternativa, la cantidad de registros en el archivo de datos original con ese valor para la clave alternativa, y una lista con las claves primarias, ordenadas ascendentemente, de todos esos registros (para el caso en que la clave alternativa sea de valores únicos, los registros de índice alternativo en lugar de tener una lista de muchas claves primarias tendrían una sola clave, es decir una lista unitaria ...).

## Ejercicios

I) Sea un archivo VSAM KSDS con intervalos de control de 1024 bytes.

(a) Grafique el estado de un intervalo de control conteniendo los registros cuyas claves y longitudes se consignan a continuación:

Abel	150	Abelardo	150	Adelina	180
Adolfo	150	Alcira	155	Agapito	63

Denote a los registros escribiendo únicamente sus claves, y a los RDF por el valor decimal de sus dos últimos bytes (el primer byte es 0).

(b) Ilustre la división del intervalo anterior merced a la incorporación del registro con clave Aldana, de 250 bytes.

II) Sea un archivo secuencial indicado de registros de longitud fija estilo VSAM, con intervalos de control con capacidad para cuatro registros (intervalos de un sólo bloque de almacenamiento), y áreas de control de cuatro intervalos.

(a) Mostrar el estado del archivo y su índice primario tras cargarlo, al 75% de densidad tanto de intervalos como de áreas, con los registros cuyas claves se listan a continuación:

**1, 15, 24, 38, 43, 55, 62, 79, 88, 92, 99, 108, 113, 117, 122, 130, 144, 149, 151 y 168**

(b) Luego graficar la evolución del archivo e índice resultantes según se le incorporen los registros de claves 110, 90, 116 y 105 en el mismo orden en que se enuncian.

A los efectos de graficar los registros, sólo emplear la clave primaria como referencia...

Numerar los intervalos, tanto de datos como de índices, con números enteros a partir del 0 (cero), denotándolos claramente, y considerar los niveles superiores del índice con factor de ablocamiento 2 (ésto último, lo del factor de ablocamiento 2, es un subterfugio sin ningún sentido práctico que tiene como objetivo determinar un árbol de indicación con varios niveles y así lograr una mayor ejercitación).

III) Sea un archivo secuencial indicado estilo VSAM con registros de longitud fija, intervalos de control con capacidad para cuatro registros, áreas de control de tres intervalos y claves primaria y alternativas NUMERO, NOMBRE y EDAD, respectivamente.

(a) Mostrar el estado del archivo y sus índices tras cargarlo con densidades de intervalos y de áreas del 75% y 66%, respectivamente, con los registros cuyas claves son:

<u>NUMERO</u>	<u>NOMBRE</u>	<u>EDAD</u>	<u>NUMERO</u>	<u>NOMBRE</u>	<u>EDAD</u>
17	Paco	29	67	Fito	24
28	Chona	32	74	Lito	31
35	Zuny	29	76	Tita	25
44	Tito	30	80	Ñeco	26
49	Pepe	25	83	Pato	33
52	Paty	24	89	Truly	29
58	Teto	32	95	Ñata	30
65	Coca	29	99	Queco	35

Emplear las mismas convenciones que en el ejercicio anterior y construir los índices alternativos almacenando tres registros por intervalo de control, y con áreas de tres intervalos. Los índices alternativos se crean simultáneamente con el primario durante la carga del archivo de datos.

(b) Mostrar la evolución del archivo e índices a partir del estado anterior según se le incorporen , en el orden en que se listan, los registros de claves (50, Franchi, 29), (41, Quico, 35), (12, Flo, 18) y (32, Chicha, 24).

(c) Idem anterior pero según se supriman los registros de Zuny y Pato.

IV) Sea un archivo VSAM con áreas de control de 12 intervalos, cargadas originalmente con una densidad del 75%. Las claves del último registro de cada intervalo de control ocupado en una de sus áreas, junto con el número de intervalo correspondiente son:

**Agata, 0; Beatriz, 1; Cyntia, 2; Doris, 3; Elina, 4; Fedora, 5; Gloria, 6; Irma, 7; Laura, 8 y Alma, 9**

(a) Reconstruir el registro de índice correspondiente a esta área, omitiendo el encabezamiento del mismo y sin abreviar las claves.

(b) Mostrar el nuevo estado del registro luego de que se produjeren en el archivo el particionamiento de un intervalo, siendo Gilda la clave del último registro del nuevo intervalo, y el fusiónamiento de otros dos, con la consiguiente liberación del que tiene como último registro el de clave Irma.