

Organización de Datos – Curso Servetto

Organización de Archivos

Artículo(Código, Descripción, (Disciplina)?, Marca, Precio de Venta, (Color, Existencia)*)

Factura(Número, Fecha, ((“CO”)|(“DA”, Tarjeta, Número)|(“TC”, Tarjeta, Número, Vencimiento, Código, Cuotas)|(“CH”, Número)), Factor de Descuento o Interés, (Código, Color, Cantidad, Precio Unitario)+)

Tema 1

- Archivo de artículos: directo con bloques de 1 kb, saturación lineal y cantidad estimada de 1500 registros de longitud variable – primitivas – primitiva de creación
- Archivo de facturas: secuencial indexado por fecha con registros de longitud variable – primitivas – función para totalizar lo facturado en una fecha

Estructura física de archivos

Artículo(Espacio libre, Indicador de desborde, Código, Longitud descripción, Descripción, Longitud Disciplina, (Disciplina)?, Longitud Marca, Marca, Precio de Venta, Cantidad colores, (Longitud color, Color, Existencia)*)

Factura(Número, Fecha, ((“CO”)|(“DA”, Tarjeta, Número)|(“TC”, Tarjeta, Número, Vencimiento, Código, Cuotas)|(“CH”, Número)), Factor de Descuento o Interés, Cantidad artículos distintos, (Código, Longitud color, Color, Cantidad, Precio Unitario)+)

IndiceFechaFactura((Nivel no terminal, Espacio libre, Nodo fechas menores, (Fecha, Nodo fechas igual y/o mayores)+)(“0”, Espacio libre, (Fecha, Desplazamiento de primer factura)+))

La estructura de Artículo es la estructura de un bloque, la unidad de organización, y los campos de control son los que figuran en gris. El campo **Indicador de desborde** sirve para terminar búsquedas infructuosas de registros en bloques que nunca se hayan desbordado. El indicador de espacio libre puede reemplazarse o complementarse con un campo contador de registros dentro del bloque.

La estructura de Factura es la estructura de un registro, ya que este archivo NO se organiza en bloques.

La estructura del índice de facturas por fecha es la estructura de un nodo no terminal o de uno terminal u hoja de nivel 0 para un árbol B+. Como en el archivo de facturas no hay bloques ni altas aleatorias de registros que rompan el orden de la fecha, en los nodos terminales del conjunto de secuencia no hace falta el campo de encadenamiento hacia adelante; por esto mismo, también puede estructurarse el índice como árbol B sin perder la organización secuencial indexada:

IndiceFechaFactura((Nivel no terminal, Espacio libre, Nodo fechas menores, (Fecha, Desplazamiento de primer factura, Nodo fechas mayores)+)(“0”, Espacio libre, (Fecha, Desplazamiento de primer factura)+))

Primitivas

	<i>Creación</i>	<i>Actualización</i>	<i>Recuperación</i>	<i>Mantenimiento</i>
Artículos (Acceso Directo)	Requiere la inicialización del espacio de almacenamiento y la previsión de más espacio del necesario para el manejo de colisiones: se debe prever espacio para aproximadamente $1500 \times 1,5$ registros y determinar la cantidad de bloques $(1500 \times 1,5 \times t) \div 1024$, donde t es el tamaño medio estimado de un registro; luego se debe inicializar cada bloque poniendo en 1024 y en falso el tamaño libre y el indicador de desborde respectivamente.	Inserción de un registro completo en el bloque que determine la clave de identificación o en el primero en el que se encuentre espacio a partir de aquél. Supresión de un registro proveyendo su clave de identificación. Modificación de un registro proveyendo el registro modificado completo y la diferencia de tamaño respecto a la versión previa.	Permite sólo recuperación selectiva de registros individuales por clave primaria. Para recuperar rangos de registros según clave primaria se debe mantener un índice o lista de claves. Permite recuperación por otras claves por mantenimiento de índices alternativos con referencia a claves primarias (indirección simple a registros de datos).	Requiere reestructuraciones periódicas, y si la saturación supera la proporción prevista en la creación se debe redefinir el espacio de almacenamiento. Las copias de respaldo se pueden tomar extrayendo en un archivo secuencial los registros almacenados (reorganizando al archivo como secuencial sin ordenación de registros).
Facturas (Acceso Secuencial Indexado)	Las cargas iniciales requieren una primitiva de alta secuencial (append) de registros, que deben estar ordenados por clave. Los archivos también se pueden crear vacíos.	Inserción de un registro completo en el último bloque del archivo (append) y actualización del índice también con primitiva de append (carga del índice al 100%) sólo si la fecha de la factura es distinta a la de la última. No hay bajas ni actualizaciones debido a la naturaleza del sistema de información.	Permite recuperación selectiva de registros o rangos de registros por clave primaria: se busca en el índice la primera factura con la fecha de la factura que se desea recuperar y se accede al archivo de datos iniciando una búsqueda secuencial. Para totalizar lo facturado en una fecha se debe procesar a todas las facturas ubicando la primera a través del índice. Para cada factura se debe calcular su total TF sumando para cada artículo vendido $TF := TF + Cantidad \times Precio$ Unitario y luego multiplicando por el Factor de Descuento o Interés. Se acumulan los totales de cada factura hasta llegar a la primera que tenga fecha distinta a la fecha objetivo o hasta que se termine el archivo.	Este archivo no requiere mantenimiento. Las copias de respaldo se toman sólo del archivo de datos organizado secuencial puro. Puede requerirse depuración de facturas anteriores a una fecha, para lo cual se puede efectuar una reestructuración selectiva.

Tema 2

- Archivo de artículos: secuencial indexado con registros de longitud variable – primitivas
- Archivo de facturas: secuencial con registros de longitud variable – primitivas – función para totalizar lo facturado entre dos fechas para una forma de pago determinada y versión alternativa con estructuras de control (índice/s)

Estructura física de archivos

Artículo(Espacio libre, Código, Longitud descripción, Descripción, Longitud Disciplina, (Disciplina)?, Longitud Marca, Marca, Precio de Venta, Cantidad colores, (Longitud color, Color, Existencia)*)

ÍndiceCódigoArtículo((Nivel no terminal, Espacio libre, Nodo códigos menores, (Código, Nodo códigos igual y/o mayores)+)(("0", Espacio libre, Siguiente nodo hoja, (Código, Desplazamiento de bloque)+))

Factura(Tamaño de registro, Número, Fecha, ((("CO"))(("DA", Tarjeta, Número))("TC", Tarjeta, Número, Vencimiento, Código, Cuotas))("CH", Número)), Factor de Descuento o Interés, Cantidad artículos distintos, (Código, Longitud color, Color, Cantidad, Precio Unitario)+)

La estructura de Artículo es la estructura de un bloque, la unidad de organización, ya que no se concibe otra organización para archivos con registros de longitud variable que no sea en bloques, y los campos de control son los que figuran en gris. El campo de control Espacio libre puede acompañarse también de un contador de registros.

La estructura del índice de artículos por código es la estructura de un nodo no terminal o de uno terminal u hoja de nivel 0 para un árbol B+.

Primitivas

	Creación	Actualización	Recuperación	Mantenimiento
Artículos (Acceso Secuencial Indexado)	Las cargas iniciales requieren una primitiva de alta secuencial (append) de registros, que deben estar ordenados por clave. También se puede crear vacío.	Puede haber inserciones secuenciales e inserciones aleatorias. Las inserciones y supresiones degradan la organización del archivo disminuyendo la densidad de carga de los bloques y empeorando las condiciones de acceso secuencial por alteración del orden físico de los bloques.	Permite recuperación selectiva de registros o rangos de registros por clave primaria. Permite recuperación por otras claves por acceso secuencial o por mantenimiento de índices alternativos al índice primario (indirección doble a registros de datos).	Requiere reestructuraciones frecuentes, cuya necesidad se puede monitorear contabilizando la cantidad de bloques que se han desbordado o partido. Las copias de respaldo se toman sólo del archivo de datos organizado secuencial puro.
Facturas (Acceso secuencial)	Se crea vacío, y las cargas iniciales se hacen con inserciones sucesivas de registros ordenados por clave de identificación.	Los registros no se actualizan dada la naturaleza de la información. Sólo se pueden agregar registros al final, con la primitiva de append.	Permite recuperación selectiva de registros o rango de registros por acceso y búsqueda secuencial.	No requiere reestructuración. Las copias de respaldo se hacen copiando el archivo completo. Puede requerir depuración de facturas anteriores a una fecha.

Totalización lo facturado entre dos fechas para una forma de pago determinada.

Sin estructuras de control de apoyo:

Se lee secuencialmente el archivo de facturas desde su registro hasta encontrar la primer factura con fecha igual a la inicial, y a partir de ésta, según su forma de pago sea la requerida se totaliza y acumula el total hasta encontrar la primer factura con fecha siguiente a la fecha final. La búsqueda de la primer factura sólo se realiza procesando los campos Tamaño de registro, Número y Fecha de cada factura y realizando el fseek al comienzo de la próxima factura por desplazamiento relativo. Para cada factura se debe calcular su total TF sumando para cada artículo vendido $TF := TF + \text{Cantidad} * \text{Precio Unitario}$ y luego multiplicando por el Factor de Descuento o Interés.

Con estructuras de control de apoyo se puede considerar un índice por forma de pago más fecha de la factura, que proveería una clasificación de registros razonable. También

se podría pensar en un conjunto de índices selectivos por número de factura, uno para cada forma de pago (así en cada índice estarían identificadas las facturas de una forma de pago en particular, y para resolver la consulta se debería usar el índice que correspondiera). Indexar las facturas sólo por forma de pago no es una opción válida, ya que la clasificación resulta demasiado gruesa (para cada valor de clave habrá una cantidad de referencias muy grande, y difícil de organizar).

La estructura del índice por forma de pago y fecha (de clasificación) sería:

Indice((Nivel no terminal, Espacio libre, Nodo fechas menores, ("CO"|"DA"|"TC"|"CH", Fecha, Nodo fechas igual y/o mayores)+)|("0", Espacio libre, Siguiete hoja, ("CO"|"DA"|"TC"|"CH", Fecha, Cantidad de facturas (Posición factura)+)+))

Posición de factura es el desplazamiento relativo de una factura en el archivo de datos. Se podría limitar la cantidad de posiciones a mantener en una entrada de índice organizándolas análogamente a los punteros a bloques en los inodes de Unix: por ejemplo, se podrían aceptar hasta diez posiciones, y si cantidad de facturas fuere menor o igual a las posiciones, éstas serían posiciones efectivas de facturas; en cambio si la cantidad de facturas fuere mayor, entonces la última posición sería la de un bloque de organización de posiciones en otro archivo de control.

La estructura tipo de un índice selectivo de forma de pago por número de factura (de identificación) sería:

IndiceFP((Nivel no terminal, Espacio libre, Nodo números menores, (Número, Nodo números mayores)+)|("0", Espacio libre, Siguiete hoja, (Número, Posición factura)+))