

75.06 Organización de Datos

Trabajo Práctico Primer Cuatrimestre 2004 – Curso Servetto

Sean las siguientes especificaciones de estructuras de registro, con las cuales habrán de implementarse dos archivos principales:

Facturas

Atributo	Tipo	Restricción	Descripción
NroFac	Entero de 8 dígitos	No nulo	Número de Factura.
FechaEmisión	AAAAMMDD	No nulo	Fecha de Emisión de la Factura.
FechaVto	AAAAMMDD		Fecha de Vencimiento de la Factura, cuando es de crédito (1 a 3 meses respecto de la fecha de emisión).
NroRemito	Entero de 8 dígitos		Número de Remito.
Estado	String de 2 caracteres	No nulo	Estado de la factura: PN pagada normal, CD crédito al día, CM crédito en mora, SF cheque sin fondos, PM pagada con mora, NC no cobrada.
FP	2 caracteres	No nulo	Forma de pago: CO contado, CR crédito o CH cheque. 40% de pagos CO (todos con estado PN), 30% CR (60% PN, 32% CD y 8% CM) y el resto CH (50% PN, 30% NC y el resto SF).
PorcDoI	Fraccionario con 2 enteros y 2 decimales		Porcentaje de Descuento o de Interés anual Puede ser nulo, un interés si la factura es de crédito, o un descuento si el pago es al contado Valores absolutos entre 5 y 25 (negativos para descuentos y positivos para intereses).
NroCtaCte	Entero de 5 dígitos		Número de Cuenta Corriente, para facturas de crédito.
NroCheque	String de 18 caracteres		Número de cheque, para pago por cheque Los valores son 4 números exactamente (identificadores de banco, sucursal, cuenta y cheque) de 4, 3, 5 y 3 dígitos (puede haber ceros a izq.) separados por guiones.
Nota	Texto		Observaciones respecto a lugar y condiciones de entrega, a condiciones de pago, etc.

Artículos

Atributo	Tipo	Restricción	Descripción
NroArtículo	Entero de 8 dígitos	No nulo	Número de Artículo.
Descripción	String de hasta 50 caracteres	No nulo	Nombre y características del artículo.

Presentación	String de hasta 30 caracteres	No nulo	Presentación o empaque de los artículos.
Existencia	Entero de 8 dígitos	No nulo	Existencia en stock de unidades en la presentación correspondiente.
Ubicación	String de hasta 30 caracteres		Ubicación del producto en el almacenamiento.
PVU	Fraccionario con 5 enteros y 2 decimales	No nulo	Precio de Venta Unitario.
Emín	Entero de 8 dígitos	No nulo	Existencia mínima permitida, para pedidos de compra o producción.

ÍtemVenta

Atributo	Tipo	Restricción	Descripción
NroArtículo	Entero de 8 dígitos	No nulo	Número de Artículo.
CV	Entero de 8 dígitos	No nulo	Cantidad Vendida.
PVU	Fraccionario con 5 enteros y 2 decimales	No nulo	Precio de Venta Unitario.

Las facturas y sus ítem deben almacenarse en el mismo archivo.

Primera Parte

Cada grupo deberá implementar las siguientes versiones de organización de registros para cualesquiera de los archivos principales:

1. Con registros de longitud variable justificados a izquierda en bloques de tamaño parametrizado, e identificados a nivel físico por un atributo de control IdReg, de valor entero positivo mayor o igual que cero y generado automáticamente y en secuencia cuando se crea el registro. Para recuperar un registro con IdReg n se debe tener un archivo de números de bloque (tabla de direcciones de registro) tal que el número en la posición n sea el del bloque que contiene a ese registro; así, se lee el bloque que corresponde y se busca el registro secuencialmente por IdReg en él. Las altas y bajas en los archivos de registros de datos deben implementarse con recuperación de espacio libre, para lo cual cada archivo de datos debe contar con un archivo *de espacio libre por bloque*; la búsqueda de espacio para altas debe efectuarse en el archivo de espacio libre secuencialmente desde el registro 0, correspondiente al bloque con idéntica dirección, hasta encontrar un bloque en el que quepa el registro (política de primer ajuste¹) y quede al menos un porcentaje parametrizado de espacio libre para posibles modificaciones de registros del bloque que los hagan crecer en tamaño. Los números de IdReg que se den de baja también deben ser recuperables². En cualquier momento de la existencia de un archivo de registros de datos se debe poder consultar la relación entre el espacio libre y el

¹ Otras políticas podrían ser la del mejor ajuste (buscar el bloque con mayor cantidad de espacio libre) o la del peor (buscar el bloque con menor cantidad de espacio libre en el que quepa el registro), pero en estos casos los registros de espacio libre deberían contener el número de bloque de archivo de datos acompañando a la cantidad de espacio libre, y además deberían estar ordenados por cantidad de espacio en forma creciente o decreciente según corresponda. Para el primer ajuste, basta con que cada registro tenga únicamente la cantidad de espacio, ya que su posición corresponde al número de bloque.

² Para la recuperación se debe contar con otro archivo de números de IdReg liberados, y cuando se requiera uno para un alta, si este archivo existe (si no existe hay que generar un número nuevo consecutivo al último) se lee el último número y se trunca (recuperación en forma de pila).

tamaño del archivo, la relación entre el espacio ocupado por información de control³ y el tamaño del archivo, la cantidad promedio de espacio libre en bloques y las desviaciones extremas.

2. Con registros de longitud variable, identificados a nivel físico por un atributo de control IdReg, de valor entero positivo mayor o igual que cero y generado automáticamente y en secuencia cuando se crea el registro. Para recuperar un registro con IdReg n se debe tener un archivo de posiciones relativas de registro tal que el número en la posición n sea el desplazamiento relativo en bytes en que se encuentra el registro n ; así se lee el desplazamiento que corresponde y se va directamente al comienzo del registro en el archivo. Las altas y bajas en los archivos de datos deben implementarse con recuperación de espacio libre, para lo cual cada archivo debe contar con un archivo de espacios libres con posiciones relativas de comienzo del espacio, y dimensión en bytes del mismo. Si al modificarse un registro éste se expande, se debe liberar el espacio que ocupa generando un registro de espacio libre y buscar otro espacio capaz de albergarlo con su nueva longitud, o grabarlo al final del archivo si no se encuentra espacio. En cualquier momento de la existencia de un archivo de datos se debe poder consultar la relación entre el espacio libre y el tamaño del archivo, la relación entre el espacio ocupado por información de control y el tamaño del archivo, la cantidad promedio de espacio libre para registros y las desviaciones extremas.
3. Con registros de longitud fija justificados a izquierda en bloques de tamaño parametrizado⁴, e identificados a nivel físico por un atributo de control IdReg, de valor entero positivo mayor o igual que cero y generado automáticamente y en secuencia cuando se crea el registro. Para recuperar un registro con IdReg n se debe tener un archivo de números de bloque tal que el número en la posición n sea el del bloque que contiene a ese registro; así, se lee el bloque que corresponde y se busca el registro secuencialmente por IdReg en él. Las altas y bajas en los archivos de dato deben implementarse con recuperación de espacio libre, para lo cual cada archivo de datos debe contar con un archivo de espacio libre por bloque; la búsqueda de espacio para altas debe efectuarse en el archivo de espacio libre secuencialmente desde el registro 0, correspondiente al bloque con idéntica dirección, hasta encontrar un bloque en el que quepa el registro (política de primer ajuste). Los números de IdReg que se den de baja también deben ser recuperables⁵. En cualquier momento de la existencia de un archivo de datos se debe poder consultar la relación entre el espacio libre y el tamaño del archivo (en términos de registros y no de bytes), la relación entre el espacio ocupado por información de control y el tamaño del archivo, la cantidad promedio de espacio libre en bloques y las desviaciones extremas.

Todos los campos de datos deben almacenarse en formato ASCII, y en caso de que los registros sean de longitud fija, los valores deben estar debidamente justificados según el tipo del campo.

Para individualizar registros de longitud variable se pueden emplear separadores (un carácter especial), prefijos de longitud o, sólo cuando se disponen en bloques, mapas de extensión (los registros se disponen de izquierda a derecha del bloque, y al final del mismo, de derecha a izquierda se especifica la cantidad de registros en el bloque y una lista de longitudes; entre medio de los registros y el mapa queda el espacio libre del bloque).

Los registros de ítem deben considerarse componentes de un registro de factura, y por consiguiente no deben tener identificador. Para saber cuántos ítem tiene una factura se puede agregar al registro de factura un campo de control que los contabilice.

El campo de texto del archivo de facturas se debe almacenar en un archivo a parte, con longitud de bloque parametrizada, encadenamiento de bloques para un mismo campo y recuperación de bloques liberados.

³ Incluyendo al espacio ocupado por los archivos de direcciones de registro y de espacio libre, además del espacio ocupado por campos de control en los registros.

⁴ Los registros (bloques) de este archivo se deben definir como un arreglo de registros, más un atributo fantasma para completar los bytes del tamaño del bloque, si fuera necesario.

⁵ Para la recuperación se debe contar con otro archivo de números IdReg liberados, y cuando se requiera uno para un alta, si este archivo existe (si no existe hay que generar un número nuevo consecutivo al último) se lee el último número y se trunca (recuperación en forma de pila).

La implementación de las versiones debe estar integrada en un único sistema que permita escoger con qué versión crear un archivo o reconozca su versión al abrirlo, y una vez abierto un archivo detalle su versión y provea la posibilidad de inspeccionarlo físicamente, sea solicitando la visualización de un unidad lógica en particular (bloque o registro), o sea secuencialmente (mostrar la siguiente a la actual). La visualización de contenidos debe ser lineal, empleando todo el ancho de la pantalla o ventana de visualización. Si se pidiera la visualización de un registro en una versión con bloques, debe mostrarse todo el contenido del bloque que contenga al registro, y a éste en particular resaltado de alguna manera; y si la versión de organización física no es con bloques, igualmente exhibirlo contextualizado en el entorno de su unidad física de almacenamiento. A los espacios vacíos exhibirlos respetando su extensión.

El campo de texto del archivo de facturas debe exhibir el número de bloque inicial en el archivo de almacenamiento de textos. Para visualizar el contenido de este archivo se debe adoptar una convención coherente con lo estipulado antes.

Además de opciones para el acceso secuencial o directo a bloques y registros, la interfaz de visualización debe contemplar la posibilidad de hacer altas, bajas y modificaciones (sólo de atributos que no identifiquen registros) de registros. Los casos de bajas o modificaciones deben admitirse luego de seleccionar el registro para su visualización, y debe exhibirse el mismo bloque o contextualización del registro luego de efectuada la operación; para el caso de las modificaciones, abrir una interfaz que permita editar los campos con comodidad, y una vez aceptados los cambios pasar a la exhibición del nuevo contexto a continuación del previo. En el caso de que se solicite un alta, abrir una interfaz de edición como en las modificaciones, y luego de aceptada la operación mostrar el nuevo registro en su contexto, independientemente de lo que se hubiera visualizado antes (borrón y cuenta nueva...). También se debe poder consultar las estadísticas del archivo que se esté accediendo, así como reestructurarlo (recrearlo físicamente compactando espacios libres, y con la posibilidad de cambiar valores parametrizados como tamaño de bloque o porcentaje de espacio para extensiones, si existen) y reorganizarlo (cambiar la organización física de registros).

Las operaciones primitivas de acceso secuencial y relativo y de actualización de registros deben estar encapsuladas en un tipo de dato abstracto para cada versión de organización física de registros, y cada primitiva debe tener la misma interfaz, sea cual fuere la organización física de los registros del archivo.

El objetivo de esta primera parte del trabajo es hacer un análisis comparativo de las distintas versiones de organización física de registros. Por lo tanto, para que la comparación sea válida debe efectuarse un experimento de simulación de carga, reproduciendo las mismas condiciones para cada versión de organización (se puede hacer la carga de cualquier archivo en una cualquiera de las versiones, y luego reorganizarlo en otra).

Se deberá desarrollar un módulo de simulación que efectúe una carga de los archivos generando aleatoriamente valores de atributos no identificadores, dentro de rangos o conjuntos de valores razonables según el significado de cada atributo. Para los atributos de tipo string se puede definir arreglos con conjuntos de valores posibles (por ejemplo, de nombres de productos y marcas para descripciones de artículos, y de nombres de presentaciones y capacidades para presentaciones de artículos), y escoger aleatoriamente el valor dentro del arreglo. Para los campos de texto se puede proceder análogamente, con un subconjunto de textos preestablecidos de tamaños dispares.

El proceso de generación debe servir indistintamente para cualquiera de las versiones de organización física de los registros, y debe cargar inicialmente al archivo de artículos completo y luego el de facturas e ítem de venta en forma interactiva. Para generar las facturas y los ítem de cada una, primero se debe producir un registro de factura, luego determinar la cantidad de ítem que tendrá, y seguidamente generar cada ítem (escogiendo artículos distintos dentro de cada factura). El 50% de las facturas deben tener fechas de abril del 2004, y la otra mitad de mayo (la cantidad de facturas por día puede ser uniforme). El diseño de la forma de simulación de la carga de datos debe ser verificado y autorizado por la cátedra.

La presentación del trabajo debe efectuarse con una carga de cien artículos y mil facturas de entre uno y diez ítem de venta: el 30% de las facturas deben tener un sólo ítem, el 40% entre dos y cinco, el 20% entre seis y diez, y el resto entre diez y quince.

El cumplimiento de esta parte del trabajo se verificará mediante la comprobación del funcionamiento del sistema más un informe escrito con las estadísticas de cada archivo en cada versión de organización física y distintas alternativas de tamaño de bloque para cada organización, y las conclusiones fundadas acerca de cuál organización resultó mejor. También se deberá proponer en el informe, un criterio basado en las estadísticas de cada archivo para recomendar su reestructuración. Todos los análisis deben efectuarse con la carga especificada para la entrega.