

Organización de Datos

Curso Servetto

Acceso Directo a Registros por Clave

Metodos De Dispersion (Hashing)

1. Especifique un algoritmo de Hashing basado en el resto de la división por "m" para una clave de tipo string[12].
2. Considere la organización de un archivo cuya clave de acceso es NOMBRE, de tipo string[15] (nombres propios de personas, como Juan, María, etc.) y su espacio de direcciones 2047. Defina un algoritmo de Hashing basado en el análisis de dígitos, previa realización de un muestreo de tamaño 20.
3. Especifique un algoritmo de Hashing basado en la operación x-or: $(arg1 \mid arg2) \& (-(arg1 \& arg2))$
Considere que la clave a aplicar el algoritmo es de tipo string[25] y el espacio de direcciones 4095.
4. Halle las direcciones correspondientes a Christian, María, Sebastián, Ximena y Estefanía utilizando un método
 - a) de análisis de dígitos
 - b) de plegado y suma
 - c) de corrimiento y suma
 - d) de x-or
 - e) del centro del cuadrado
 - f) de transformación a bases 8, 11 y 13

Considere que el tipo de la clave es string[10] y el espacio de direcciones 7.

Analisis de Rendimientos

1. Suponer que hay 10000 direcciones para almacenar 8000 registros en un archivo con dispersión aleatoria (método de hashing), y que cada dirección puede almacenar un registro. Calcular los siguientes valores.
 - a) La densidad del archivo.
 - b) El número esperado de direcciones a los que la transformación de hashing no asignará registros.
 - c) El número esperado de direcciones con un único registro asignado (sin sinónimos).
 - d) El número esperado de direcciones con más de un registro asignado.
 - e) El número esperado de colisiones (registros de overflow o que causan saturación).
 - f) El porcentaje esperado de colisiones.
2. Considerando el archivo descrito en el ejercicio anterior, decir cuál es el número esperado de colisiones si las 10000 localidades se reorganizan como
 - a) 5000 compartimientos (buckets) de dos registros, y
 - b) 1000 compartimientos de diez registros.

- Hacer una tabla que muestre valores de la función de Poisson para densidades 0,1; 0,5; 0,8; 1; 2; 5 y 10. Examinar la tabla y analizar características y patrones que proporcionen información de utilidad con respecto a la dispersión o transformación de hashing.

Tratamiento De Colisiones

- Considere un archivo directo con espacio de direcciones de veinte slots y direccionamiento abierto con búsqueda o sondeo lineal (dispersión progresiva).
Detalle los cambios producidos en el archivo con el movimiento de registros cuyo orden, claves, movimiento y direcciones relativas se detallan a continuación (+ alta, - baja), y a partir del estado de vacuidad:

+Pérez 9	+Salas 13	-Escabo 15
+Balzac 19	+Escalante 19	+Molinari 9
+Gutiérrez 19	+Berraux 20	+Berra 2
+Escabo 15	-Maciel 19	-Berraux 20
+Maciel 19	+Pellegrini 5	+Pacino 20
+Gianini 13	+Páez 14	-Basualdo 9
+Basualdo 9	-Balzac 19	

Indique para cada alta, la dirección que realmente le corresponde al registro que se incorpora, si es una colisión con un sinónimo o con un registro de overflow, y la cantidad de accesos de lectura de slots que requirió. Realice lo solicitado listando para cada slot la sucesión de todos los estados que tuvo.

- Resuelva el ejercicio anterior con encadenamiento de sinónimos.
- Resuelva el ejercicio 1 pero para un archivo de siete compartimientos (buckets) con capacidad para tres registros (recalcular las direcciones).
- Suponga que deberá automatizar la administración de una empresa de servicios que cuenta con 840 clientes cuyos datos guardará en un archivo directo. El movimiento de clientes de la empresa es lento (el archivo será más bien estático) y con una leve tendencia de expansión. El archivo estará organizado de la siguiente manera:
 - acceso a slots (compartimientos individuales)
 - direccionamiento abierto con búsqueda lineal (saturación progresiva)
 - organización de sinónimos a continuación del espacio de direcciones en el mismo archivo (encadenamiento en área de saturación separada)
 - Defina el espacio de direcciones del archivo (la cantidad de slots que habrá de tener) y describa el procedimiento de creación del mismo.
 - Describa puntualmente la incorporación de un registro detallando todas las condiciones posibles. Grafique las estructuras de datos necesarias.
- Sea un archivo directo diseñado para almacenar información sobre los miembros de un club. El mismo se manejará con una tabla de dispersión estática (Index Random) de direccionamiento abierto con búsqueda lineal. La dimensión de la tabla es de veinticuatro entradas con dirección relativa a partir de 0.
 - Grafique el contenido de la tabla y del archivo sabiendo que se incorporaron en el siguiente orden:

ORDEN	CLAVE	DIR	ORDEN	CLAVE	DIR
1	Pablo	19	9	María	19
2	Roxana	8	10	Daniel	11
3	Marcos	8	11	Esther	15
4	Andrea	19	12	Carlota	23
5	Fabián	15	13	Esteban	16
6	Laura	17	14	Ignacio	23
7	Alejandro	18	15	Lorena	1
8	Rodolfo	15	16	Sebastián	19

- b) Refleje el nuevo estado de la tabla y del archivo luego de las bajas de Andrea, Carlota, Esther y Lorena. Indique la cantidad y tipo de operaciones (de lectura y de escritura) de registros que hubiera requerido cada una.
- c) ¿Dónde y cómo se almacenarían físicamente los componentes de este archivo (la tabla y los registros) y cómo y cuándo se accedería a ellos? Especifique las estructuras de datos.
- d) ¿Qué haría si se terminara el espacio libre de la tabla? Detalle puntualmente el procedimiento.
- e) ¿Cómo implementaría un acceso alternativo a los registros de este archivo, como por ejemplo la edad de los miembros del club? Especifique las estructuras de datos que agregaría y sus organizaciones. Asuma, desde luego, que la edad de los asociados ya figura como un campo de los registros.
6. Resuelva el ejercicio 5 con acceso a buckets (sin tabla primaria).
7. Resuelva el ejercicio 5 con acceso a buckets y localización de overflows simplemente encadenados en un archivo específico para tal fin.
8. Sea un archivo directo cuyo manejo ha de efectuarse con acceso a buckets de tamaño 3, y con hashing extensible. La clave de acceso a este archivo es CÓDIGO y habrá también una clave alternativa secundaria NOMBRE. El resultado de la transformación que se aplica a la clave secundaria se apunta entre paréntesis.
- a) Defina las estructuras de implementación de este archivo.
- b) Refleje el estado de las tablas y del archivo tras las siguientes operaciones (especificadas por columnas, primero NOMBRE y luego CÓDIGO):
- | | |
|----------------------|----------------------|
| + BUSTILLO (5), 369 | + GOIN (7), 592 |
| + DULIO (3), 854 | + CICCHERO (8), 689 |
| + GROSSO (5), 482 | + PEREZ (11), 094 |
| + POVEDA (7), 330 | - GOIN |
| - BUSTILLO | + VAZQUEZ (9), 393 |
| + HOPSON (6), 877 | ----- |
| - GROSSO | - MONTERO |
| + MONTERO (9), 811 | + FARINA (13), 390 |
- c) Detalle el costo en operaciones de lectura de cada una de las dos últimas operaciones asumiendo que las tablas se mantienen siempre actualizadas en memoria magnética.
- d) Explique en qué consiste la creación de este archivo.