Apresentação do Grid File para o Seminário

Como mencionado no início, o Grid é um método de repositório em que o espaço é dividido em grades com os divisores fixos. Já o Grid File, os divisores não precisam estar em posições regulares. Cada célula da grade corresponde a um único bucket que, por sua vez, pode corresponder a mais de uma célula, porém a região, definida por todas elas, precisa ser convexa.

Isso fornece uma grid de n-dimensões em que n representa quantas chaves podem ser usadas para referencias um único ponto.

Grid Files não contém nenhum dado em si, mas, ao invés, contém referência ao bucket correto.

Um Grid File normalmente é usado em casos em que um único valor pode ser referenciado por múltiplas chaves.

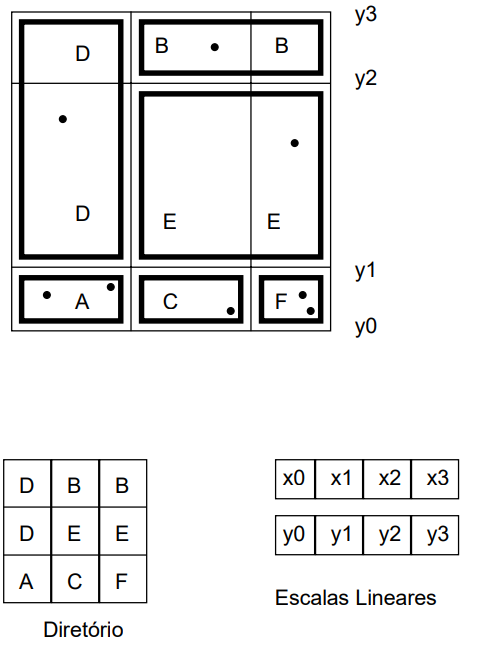
Essas diferenças tratam as desvantagens e dificuldades do Grid mencionado anteriormente, como a questão de memória para salvar os dados, dificuldade para realizar pesquisas de múltiplas chaves e o problema do pior caso de quando um bucket está vazio.

A estrutura de acesso é composta de dois elementos

- Um array que mapeia células em buckets (armazenado em disco)

- Escalas lineares (uma para cada dimensão) onde as cotas dos hiperplanos de partição são armazenadas (em memória interna)

A principal vantagem é que acesso A cada ponto é dado por apenas 2 acessos a disco.



Exemplo de uso de um grid File:

Considere o banco de dados sendo dados de um censo. Um único registro refere-se a uma família e todos os dados são agrupados em buckets. Esses registros podem ser agrupados pela cidade (desse modo o nome será o mesmo para todos os registros de um mesmo bucket) ou pelas ruas daquela cidade em que os nomes começam com a mesma letra do alfabeto.

A estrutura do Grid File, portanto, irá fornecer um index eficiente para essa estrutura, em que os registros serão agrupados em 26 grupos, cada um relacionado aos nomes da rua que começam com a mesma letra do alfabeto na cidade. A estrutura será vista como um array de duas dimensões, eixo x e y, sendo x o nome da cidade e y uma letra do alfabeto.

Cada registro dessa estrutura é uma célula. Cada célula contem um ponteiro para o bucket adequado na base de dados onde os dados em si estão armazenados. Uma célula extra será necessária para guardar o nome da cidade, as demais células agrupadas com essa extra apenas apontarão para o respectivo bucket, já que a primeira célula corresponde às ruas com letra A, a segunda com letra B e assim sucessivamente.

Ainda há a possibilidade de futuramente estender o para conter um campo de continente, caso queira levar esse censo para outros continentes. Os registros, portanto, passariam a corresponder às famílias de uma rua com a mesma letra do alfabeto, de uma mesma cidade em um mesmo continente.

Assim, as células no GridFile corresponderiam à um cabeçario da cidade e, diferentemente de como era antes que continham apenas grupos de 26, agora passará a conter 6 diferentes grupos de 26. Passando a ser então um array tridimensional, correspondendo a cidade, ao continente, e a letra do alfabeto.

Vantagens:

Nenhuma computação especial é necessária

Apenas os registros certos são recuperados

Também pode ser usado para pesquisa de chave única

Facilmente expansível para pesquisa de n chaves

Melhora significativa no tempo de processamento para pesquisa de múltiplas chaves

Para acessar os dados precisa de acessar apenas 2 discos

Desvantagens:

TEXTO BASE

Em Ciência da Computação, um grid file ou bucket grid é um método de ponto de acesso que separa o espaço em grid não periódica em que uma ou mais células da grid refere-se a um pequeno conjunto de pontos. Grid Files (uma estrutura de dados simétricos) proporciona um método eficiente de armazenar esses indexes em disco para realizar pesquisa complexa de dados.

Isso fornece uma grid de n-dimensões em que n representa quantas chaves podem ser usadas para referencias um único ponto.

Grid Files não contém nenhum dado em si, mas, ao invés, contém referência ao bucket correto.

Um Grid File normalmente é usado em casos em que um único valor pode ser referenciado por múltiplas chaves.

Um Grid File começou a ser usado porque "arquivos de estruturas tradicionais que fornecem múltiplas chaves de acesso para registros, por exemplo, arquivos invertidos, são extensões de estrutura de arquivo originalmente desenvolvido para acesso de chave única. Eles manifestam várias deficiências, em particular para acesso de múltiplas chaves para arquivos altamente dinâmicos."

Em uma tradicional estrutura de dados monodimensional (exemplo hash), uma pesquisa sob um único critério é normalmente muito simples, mas pesquisar por um segundo critério pode ser muito mais complexo.

Grid Files representa um tipo especial de hashing, em que o hash tradicional é substituído por um diretório grid.

Considere a data base contento dados de um censo. Um único registro referente uma casa, e todos os registros estão agrupados em buckets. Todos os registros em um bucket podem ser indexados tanto por sua cidade (em que o nome é o mesmo para todos os registros daquele bucket), quanto pela rua daquela cidade em que os nomes começam com a mesma letra.

Um Grid File pode ser usado para prover um index eficiente para essa estrutura, em que os registros são agrupados em grupos de 26, cada um deles relacionados com o nome da rua de uma cidade começando com uma das letras do alfabeto. Essa estrutura pode ser pensada como um vetor, tabela, ou grid com duas dimensões, que chamaremos eixos de x e y.

Podemos considerar o eixo x sendo a cidade e o eixo y sendo cada letra do alfabeto, ou vice-e-versa.

Cada registro nessa estrutura é conhecido como célula. Cada célula contém um ponteiro para o bucket apropriado no arquivo onde os dados estão armazenados. Uma célula extra, ou cabeçario de registro, talvez tenha que armazenar o nome da cidade. Outras células agrupadas terão que ter apenas o ponteiro para seu respectivo bucket, uma vez que a primeira a célula corresponde ao nome das ruas começando com "A", a segunda começando com "B", e assim sucessivamente.

A data-base, futuramente, poderá se estender para guardar um campo de continente para expandir o censo para outros continentes. Isso faria com que os registros do mesmo bucket correspondessem a casas de uma rua que começa com a mesma letra, na mesma cidade e no mesmo continente.

As células nesse Grid File consistiriam então em um cabeçario da cidade, e seis (um para cada continente, antártica não incluso) grupos de 26 células relacionadas às ruas que começam com a mesma letra, da mesma cidade, de uma mesmo continente e poderia agora ser pensado como um vetor tridimensional.