

Arquitetura e Administração de Bases de Dados
Índices e Clusters

João Costa
jcosta@isec.pt

Agenda

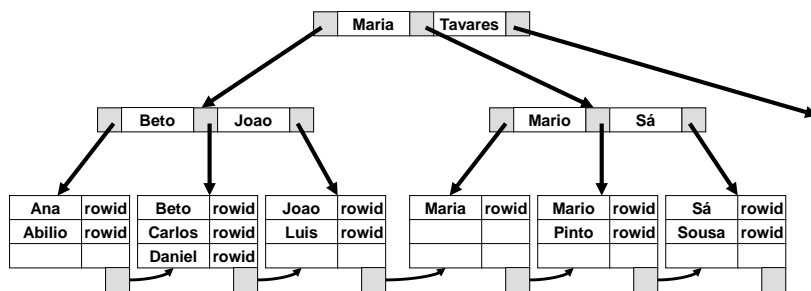
- ▶ **Indices**
 - ▶ B+tree
 - ▶ bitmap
 - ▶ function-based
- ▶ **IOT - Index organized table**
- ▶ **Clusters**
 - ▶ Index clusters
 - ▶ Hash clusters

Índices

- ▶ estruturas facultativas associadas a colunas de tabelas ou agrupamento de tabelas destinadas a acelerar o acesso aos dados
- ▶ são o equivalente a um índice remissivo de um livro pois permitem localizar rapidamente a informação pretendida

Índices

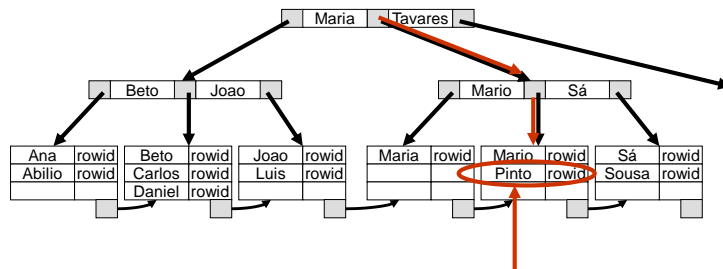
Estrutura interna de um índice B+Tree



Os índices são armazenados em B+Trees

Índices

Como funciona a pesquisa num índice B*Tree?



```
SELECT *  
FROM autores  
WHERE nome = 'Pinto';
```

Pesquisa no índice nome, para encontrar o ROWID do registo pretendido. Através do ROWID acede apenas ao bloco da tabela que contém o registo

Índices

- ▶ podem ser construídas sobre uma ou mais coluna
- ▶ só pode existir um índice por coluna
- ▶ podem ser realizados índices sobre todas as combinações das colunas de uma tabela
- ▶ mas não podem haver combinações de colunas repetidas

Índices

- ▶ estruturas lógica e fisicamente independentes dos dados da tabela a que estão associados
 - ▶ a criação / apagamento de índices não interferem com os dados da tabela nem com a estrutura da base de dados
- ▶ criados pelo utilizador através de comandos SQL
 - ▶ CREATE INDEX ...
 - ▶ DROP INDEX ...
 - ▶ ALTER INDEX ...

▶ 7

2022/23 - LEI - AABD - Índices e Clusters

Índices

Índices são geridos automaticamente pelo sistema

- ▶ em cada consulta (SELECT) a tabela com índices associados, o sistema decide se usa ou não o índice para aceder aos dados
- ▶ a inserção, atualização e/ou apagamento de dados na tabela leva a que o sistema atualize, de maneira transparente para o utilizador, os índices existente sobre essa tabela
- ▶ O facto de uma tabela ter índices, não altera em nada os comandos DML que se efetuam sobre os dados da tabela
- ▶ A única diferença para o utilizador é a obtenção de resultados às pesquisas mais rapidamente

▶ 8

2022/23 - LEI - AABD - Índices e Clusters

Índices

Criação de um Índice

```
CREATE INDEX titulos ON livros(titulo)
TABLESPACE users
STORAGE (
  INITIAL 20K
  NEXT 20K
  PCTINCREASE 75 )
PCTFREE 0;
```

Cria o índice títulos sobre a coluna título da tabela livros

O índice pode, e deve ficar, noutra tablespace diferente do tablespace da tabela, por questões de velocidade

Pode-se definir parâmetros de armazenamento que determinam como vai ser atribuído o espaço

▶ 9

2022/23 - LEI - AABD - Índices e Clusters

Índices

Vantagens da utilização de índices

- ▶ acesso mais rápido aos registos
- ▶ encontra mais rapidamente o(s) registo(s) procurado(s) quando o filtro (WHERE ...) utiliza atributos indexados

Ex. na tabela livros com 20 000 registos

```
SELECT TITULO FROM LIVROS WHERE ID_LIVRO = 9928;
```

▶ sem índices

- ▶ pesquisa sequencial, sendo necessário ler todos os registos até encontrar o desejado
média de blocos lidos = Total de blocos / 2

▶ com índice btree na coluna idlivro

- ▶ é uma pesquisa em árvore, sendo necessário ler do disco apenas os blocos necessários do índice e o bloco da tabela que contém o registo
média de blocos lidos = N níveis + 1

▶ 10

2022/23 - LEI - AABD - Índices e Clusters

Índices

Desvantagens da utilização de índices

- ▶ ocupam espaço em disco
 - ▶ se for criado um índice para cada coluna de uma tabela, o espaço ocupado por esses índices é superior ao espaço ocupado pela tabela
- ▶ os acessos apenas se tornam mais rápidos se forem efetuados sobre as colunas que tem índices
- ▶ sempre que existe alteração dos dados (INSERT, UPDATE, DELETE) é necessário reestruturar os índices que existem nessa tabela
 - ▶ Processamento de transações mais lento

Índices

Que tabelas se devem indexar?

- ▶ Tabelas grandes (com muitos registos)
- ▶ Tabelas em que as pesquisas sejam muito específicas (devolvam uma percentagem pequena de registos < 10%)
- ▶ Tabelas relativamente estáticas em que a principal operação é a pesquisa e as operações de escrita são pouco frequentes
 - ▶ operações de escrita ficam mais lentas, pois é necessário reestruturar os índices

Índices

Que colunas se devem indexar?

- ▶ colunas sobre as quais se fazem muitas pesquisas
- ▶ chaves forasteiras para acelerar as junções
 - ▶ as chaves primárias são implicitamente indexadas, mas as chaves estrangeiras não
- ▶ colunas em que a maioria dos valores são distintos
 - ▶ no caso de Btree, não faz sentido indexar uma coluna como sexo na tabela de autores, pois apenas tem 2 estados possíveis
- ▶ colunas com muitos valores nulos, mas que a maioria das pesquisas são efetuadas sobre os valores não-nulos
- ▶ não se podem indexar colunas do tipo LONG ou LONG RAW

▶ 13

2022/23 - LEI - AABD - Índices e Clusters

Índices

Quando é que se deve indexar uma tabela?

- ▶ pode-se indexar uma tabela em qualquer altura
- ▶ mas se a tabela vai receber uma grande quantidade de dados (por exemplo aquando de uma migração de dados), a indexação das colunas apenas deve ser realizada no fim da introdução dos dados
 - ▶ Senão, por cada registo migrado, terá que existir uma reestruturação de cada um dos índices que existir sobre a tabela

Quantos índices pode haver por cada tabela?

- ▶ pode haver qualquer número de índices por tabela
- ▶ não esquecer de que quantos mais índices mais lentas se tornam as operações DML

▶ 14

2022/23 - LEI - AABD - Índices e Clusters

Índices

Ordem das colunas na criação dos índices?

- ▶ Exemplo: prod_lojas(id_prod, id_loja, preço, quantidade)
- ▶ Supondo que :
 - ▶ existem algumas dezenas de lojas (idloja)
 - ▶ uma loja (idloja) pode ter milhares de produtos (idprod)
- ▶ Para consultas frequentes do tipo

```
SELECT * FROM prod_lojas
WHERE id_loja = 8
AND id_prod = 1243;
```

Qual o índice a criar?



A) CREATE INDEX Prod_loj_id ON prod_lojas(idloja, idprod);

B) CREATE INDEX log_prod_id ON prod_lojas(idprod, idloja);

Índice composto deve incluir primeiro a coluna com mais valores distintos (idprod)

▶ 15

2022/23 - LEI - AABD - Índices e Clusters

Índices

Eliminação de índices

```
DROP INDEX prod_loja_id;
```

- ▶ A eliminação de índices não altera a base de dados, só afecta a velocidade a que as operações passam a ser realizadas
- ▶ Quando se eliminam uma tabela todos os índices a ela associados também são eliminados

▶ 16

2022/23 - LEI - AABD - Índices e Clusters

Índices

Porque se eliminam índices?

- ▶ O índice não levou à melhoria de velocidade que se esperava
- ▶ As pesquisas efectuadas não permitem ao sistema a utilização dos índices
- ▶ A tabela vai passar por uma fase de inserção / actualização intensiva, pelo que é melhor eliminar o índice para o voltar a reconstruir mais tarde
- ▶ O índice está muito fragmentado, pelo que é melhor eliminá-lo e depois reconstruí-lo

▶ 17

2022/23 - LEI - AABD - Índices e Clusters

Índices

Tipos de índices

- ▶ **B* Tree**
 - ▶ índices mais habituais, informação guardada numa estrutura em árvore
- ▶ **BITMAP**
 - ▶ índices muito utilizados para atributos com pouca cardinalidade (poucos valores distintos para esse atributo)

▶ 18

2022/23 - LEI - AABD - Índices e Clusters

Índices : B*Tree

Características

- ▶ Dados encontram-se apenas nas folhas, sempre á mesma profundidade
- ▶ Pesquisa de qualquer registo demora sempre aproximadamente o mesmo tempo
- ▶ Os índices estão sempre ordenados para permitir acessos mais rápidos
- ▶ A B*Tree está sempre equilibrada pelo que após a inserção, alteração ou apagamentos de dados o sistema pode reestruturar a B*Tree
- ▶ Para minimizar o tempo na reestruturação da B* Tree, os blocos estão em média ocupados até 3/4 do seu espaço

Índices : B*Tree

Criação implícita de índices B*Tree

- ▶ ao definir-se uma chave primária a coluna (ou colunas) que constituem a chave são automaticamente indexadas
 - ▶ (Joins entre tabelas normalmente são efetuadas através das chaves)
- ▶ o mesmo se aplica para as CONSTRAINTS com restrição UNIQUE (chaves candidatas)
- ▶ também é possível construir índices compostos de várias colunas

```
CREATE INDEX aut_liv  
ON livros(idautor, idlivro);
```

Índices : BITMAP

- ▶ são uma alternativa muito poderosa aos índices BTree para determinadas situações muito específicas
- ▶ realizado sobre colunas com muitos valores repetidos (baixa cardinalidade) mas com poucos valores distintos.
 - ▶ Por exemplo:
sexo : Masculino, Feminino
Estado civil : Solteiro, Casado, Separado, Divorciado, Viúvo
- ▶ Nestes casos torna-se impraticável a construção dum B-Tree sobre estas colunas pois ficariam com muitos ROWID para cada um dos valores distintos

▶ 21

2022/23 - LEI - AABD - Índices e Clusters

Índices : BITMAP

Como funcionam os índices BITMAP

- ▶ Para cada ROWID distinto é adicionado N bits que representam cada um dos estados possíveis.
 - ▶ Por exemplo:
 - ▶ Na tabela de autores com 10 000 registos
 - ▶ um índice na coluna sexo contendo os ROWID e 2 bits
 - ▶ um índice na coluna estado_civil contendo os ROWID e 5 bits

▶ 22

2022/23 - LEI - AABD - Índices e Clusters

Índices : BITMAP

Vantagens dos índices BITMAP?

- ▶ Para colunas com pouca cardinalidade, permite pesquisas mais rápidas e ocupam menos espaço que os índices BTree
- ▶ Mecanismo que além de permitir uma maior rapidez na procura de registros de um determinado valor,
- ▶ permite uma maior rapidez na combinação de valores pois a sua estrutura facilita o OR e o AND lógico
- ▶ Impraticável em colunas com muita cardinalidade

Function-Based Index

Índices baseados em funções, permite :

- ▶ Eficaz para pesquisas com condições complexas
- ▶ Pré-computa expressões complexas e guarda o resultado no índice
- ▶ Aumente o número de situações em que o otimizador pode fazer *range-scan* ao invés de *full-table scan*

Requisitos para criar um function-based index

Deve incluir os seguintes parâmetros de inicialização:

- `QUERY_REWRITE_ENABLED = TRUE`
- `COMPATIBLE` set to 8.1.0.0.0 or a greater value

Function-Based Index

Exemplo:

- ▶ Obter os empregados com nome começado por JOH:
`SELECT * FROM emp WHERE UPPER(ename) LIKE 'JOH%';`
- ▶ Cria-se um Function-Based Index para pesquisas em maiúsculas
`CREATE INDEX idx ON emp (UPPER(ename));`

Exemplo:

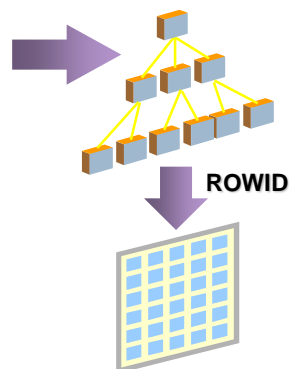
- ▶ Pesquisas com expressões nas condições
`SELECT a FROM t WHERE a + b * (c - 1) < 100;`
- ▶ Cria-se um function-based index para a expressão
`CREATE INDEX idx ON t (a + b * (c - 1), a);`

▶ 25

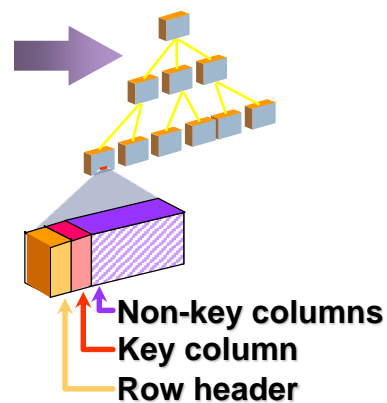
2022/23 - LEI - AABD - Índices e Clusters

IOT - Index-Organized Tables

Regular table access



IOT access



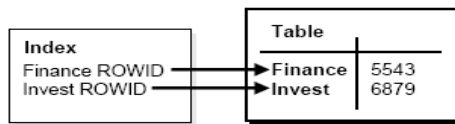
▶ 26

2022/23 - LEI - AABD - Índices e Clusters

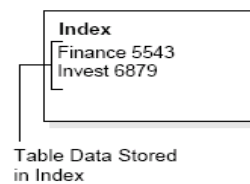
IOT - Index-Organized Tables

- ▶ Variante da B-tree, em termos de organização
- ▶ Dados guardados na estrutura do índice, de forma ordenada de acordo com a chave
 - ▶ Ao contrário das tabelas normais, cujos dados não são ordenados
 - ▶ Além do valor da chave, cada entrada inclui o valor de cada uma das colunas da tabela

Regular Table and Index



Index-Organized Table



IOT - Index-Organized Tables

- ▶ acesso mais rápido aos dados da tabela, quando as pesquisas são efetuadas pela chave
- ▶ menores requisitos de espaço
 - ▶ Apenas uma estrutura, ao invés de tabela + índice
- ▶ Limitações
 - ▶ Tabela tem que ter uma chave primária
 - ▶ Não pode incluir restrições de UNIQUE
 - ▶ Não pode ser utilizada num cluster

IOT - Index-Organized Tables

```
SQL> create table sales (  
        office_cd    number(3),  
        qtr_end       date,  
        revenue       number(10,2),  
        constraint sales_pk  
                PRIMARY KEY (office_cd,qtr_end)  
    )  
    ORGANIZATION INDEX  
    tablespace indx  
    STORAGE ( ... );
```

▶ 30

2022/23 - LEI - AABD - Índices e Clusters

Que índices ??

```
SELECT titulo  
From Livros  
Where genero='Informática';
```

```
SELECT titulo  
From Livros  
Where genero like '%Informática%';
```

```
SELECT titulo  
From Livros  
Where upper(genero) = 'INFORMÁTICA';
```

```
SELECT titulo  
From Livros  
Where upper(genero) = upper('INFORMÁTICA');
```

▶ 31

2022/23 - LEI - AABD - Índices e Clusters

Que índices ??

```
SELECT titulo
FROM Livros
WHERE preco_tabela *1.23 < 100; -- preco c/IVA < 100
```

```
SELECT titulo
FROM Livros
WHERE quant_stock - 100 > 0; -- mais que 100 em stock
```

```
SELECT *
FROM vendas
WHERE quantidade*preco_unitario> 200; -- vendas > 200
```

```
SELECT titulo
FROM Livros
WHERE SYDATE - data_edicao < 10; -- ultimos 10 dias
```

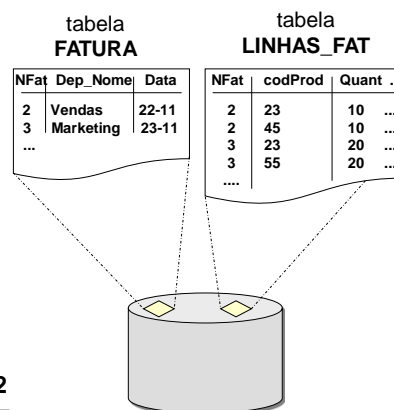
► 32

2022/23 - LEI - AABD - Índices e Clusters

Clusters

► Ver os dados da fatura nº 2

```
SELECT F.NFAT, DEP_NOME, COD_PROD, QUANT
FROM FATURA F, LINHAS_FAT L
WHERE F.NFAT = L.NFAT
AND F.NFAT = 2
```



1º ir ao índice **pk_fat** verificar se existe o **NFAT= 2**

2º se existir ir a tabela **FATURA** obter **DEP_NOME**

3º ir ao índice **pk_lfat** verificar se existe o **NFAT= 2**

4º se existir ir a tabela **LINHAS_FATURA** obter **COD_PROD** e **QUANT**

► 33

2022/23 - LEI - AABD - Índices e Clusters

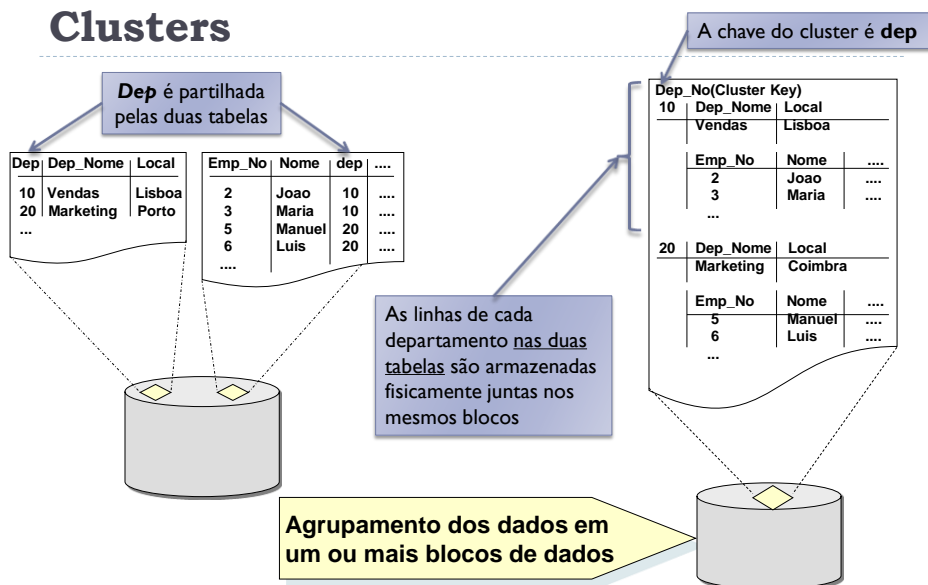
Clusters

- ▶ Forma alternativa de armazenar os dados de 2 ou mais tabelas dentro do mesmo segmento e consequentemente dentro dos mesmos blocos de dados
- ▶ É um modo eficaz para armazenar dados, quando as tabelas que constituem o cluster são acedidas em conjunto
- ▶ As tabelas necessitam de possuir um ou mais colunas em comum
- ▶ Essa chave comum (cluster key) é partilhada pelas várias tabelas, diminuindo o número de colunas
- ▶ Os dados das tabelas estão agrupadas pelo cluster Key
- ▶ Todas as tabelas que constituem o cluster partilham os mesmos blocos, os mesmos extends e o mesmo segmento

▶ 34

2022/23 - LEI - AABD - Índices e Clusters

Clusters



▶ 35

2022/23 - LEI - AABD - Índices e Clusters

Clusters

Vantagens dos Clusters?

- ▶ Rapidez - sempre que for necessário retirar dados de mais do uma tabela que constituem o cluster, a informação já se encontra no mesmo bloco não sendo necessário ir a outro bloco buscar os dados da outra tabela
- ▶ menos informação a registar, pois apenas é guardada uma das chaves comuns

Desvantagens dos Clusters?

- ▶ Decidido e construído pelo programador
- ▶ Alteração do modo como a informação é guardada

Funcionamento

▶ Depois de criado um cluster

- ▶ Podem ser criadas tabelas no cluster
- ▶ Só podem ser inseridas linhas nas tabelas
 - ▶ Depois de ser criado um índice no cluster (num cluster indexado)
- ▶ Podem ser criados índices adicionais sobre as tabelas
 - ▶ Da forma habitual
- ▶ Não devem ser criados clusters para tabelas que sejam acedidas individualmente com frequência

Gerir clusters

▶ Algumas linhas orientadoras para a gestão dos clusters

- ▶ Escolher as tabelas a integrar no cluster
- ▶ Escolher as colunas para a chave do cluster
- ▶ Espaço necessário para uma chave e linhas associadas
- ▶ Localização do cluster



2022/23 - LEI - AABD - Índices e Clusters

Clusters

Que tabelas agrupar?

- ▶ As tabelas usadas essencialmente para pesquisa
- ▶ As tabelas acedidas maioritariamente através de junções
- ▶ Se as tabelas forem acedidas individualmente com muita frequência então não se devem agrupar
 - ▶ pois necessita de percorrer mais blocos para aceder á mesma informação

Que colunas agrupar?

- ▶ A coluna ou colunas usadas nas junções
- ▶ Colunas em que a maioria dos valores são distintos
 - ▶ se a coluna ou colunas usadas nas junções não tiverem estas características então é melhor não usar clusters
- ▶ Colunas que raramente são atualizadas

▶ 39

2022/23 - LEI - AABD - Índices e Clusters

Escolher as tabelas

- ▶ Faz sentido incluir tabelas que:
 - ▶ Sejam principalmente consultadas
 - ▶ Não sejam principalmente objeto de inserções ou alterações
 - ▶ Em que os registos
 - ▶ São frequentemente consultados em conjunto
 - ▶ Ou são objeto de junções



2022/23 - LEI - AABD - Indices e Clusters

Escolher as colunas para a chave

- ▶ Se as pesquisas sobre as tabelas usam várias colunas na junção
 - ▶ Criar uma chave de cluster composta
- ▶ Uma boa chave tem valores únicos suficientes para:
 - ▶ Um grupo de linhas correspondente a um valor da chave
 - ▶ Uma chave ocupa aproximadamente um bloco de dados
 - ▶ Poucas linhas por cada valor da chave desperdiçam espaço
 - ▶ Muitas linhas por cada chave implicam mais tempo de pesquisa
- ▶ O índice do cluster não pode ser UNIQUE nem ter colunas do tipo LONG



2022/23 - LEI - AABD - Indices e Clusters

Criar um cluster

```
CREATE CLUSTER emp_dept (id_department NUMBER(3))  
  SIZE 600  
  TABLESPACE users  
  STORAGE (INITIAL 200K  
    NEXT 300K  
    MINEXTENTS 2  
    MAXEXTENTS 20  
    PCTINCREASE 33);
```

Nº médio de bytes
ocupados por um valor da
chave do cluster e
respectivas linhas

Clustered Key department_id			
20	department_name	location_id	
	marketing	1800	
	employee_id	last_name	...
201			
	202	Hartstein	...
		Fay	...
110	department_name	location_id	
	accounting	1700	
	employee_id	last_name	...
205			
	206	Higgins	...
		Gietz	...

Chave do cluster

2022/23 - LEI - AABD - Índices e Clusters

Espaço necessário para uma chave

- ▶ comando CREATE CLUSTER tem uma cláusula SIZE
 - ▶ Nº estimado de bytes necessários para um valor da chave e respectivas linhas
 - ▶ Usado para :
 - ▶ Estimar o nº de valores da chave (e respectivas linhas) que cabem num bloco de dados
 - ▶ Limitar o número de valores da chave colocados num bloco
 - Maximiza a eficiência de armazenamento
 - ▶ Por defeito, é colocado apenas um valor da chave em cada bloco

43

2022/23 - LEI - AABD - Índices e Clusters

Localização do cluster

- ▶ O cluster e o índice associado podem ser criados em qualquer TABLESPACE que esteja online
 - ▶ Desde que se tenham permissões e quota suficientes
- ▶ O cluster e o índice podem ser criados em TABLESPACES diferentes
 - ▶ Se estiverem armazenados em dispositivos distintos
 - ▶ Os dados das tabelas e do índice podem ser acedidos em simultâneo
 - ▶ Menos conflitos nos acessos ao disco



2022/23 - LEI - AABD - Índices e Clusters

Criar tabelas no cluster

```
CREATE TABLE dept (  
  id_department NUMBER(3) PRIMARY KEY,  
  ... )  
CLUSTER emp_dept (id_department);
```

```
CREATE TABLE emp (  
  empno NUMBER(5) PRIMARY KEY,  
  ename VARCHAR2(15) NOT NULL,  
  id_department NUMBER(3),  
  FOREIGN KEY(id_department)  
    REFERENCES dept(id_department)  
CLUSTER emp_dept (id_department);
```

Clustered Key department_id		
20	department_name	location_id
	marketing	1800
	employee_id	last_name
110	department_name	location_id
	accounting	1700
	employee_id	last_name
201	Hartstein	...
202	Fay	...
205	Higgins	...
206	Gietz	...

Cluster + Chave do cluster



2022/23 - LEI - AABD - Índices e Clusters

Apagar clusters

- ▶ Quando um cluster é apagado
 - ▶ São apagadas as tabelas que contenha
 - ▶ É apagado o índice correspondente
- ▶ Apagar um cluster sem tabelas e respetivo índice

```
DROP CLUSTER emp_dept;
```



2022/23 - LEI - AAEED - Índices e Clusters

Apagar clusters (cont.)

- ▶ Se o cluster tiver tabelas
- ▶ Se o cluster tiver tabelas e não for incluída a cláusula INCLUDING TABLES
 - ▶ Retorna um erro
- ▶ Se alguma tabela do cluster tiver colunas referenciadas por restrições FOREIGN KEY de tabelas fora do cluster

```
DROP CLUSTER emp_dept  
INCLUDING TABLES  
CASCADE CONSTRAINTS;
```



2022/23 - LEI - AAEED - Índices e Clusters

Clusters

```
CREATE CLUSTER Alug_socio (cod_socio number(3))
PCTFREE 15
PCTUSED 70
SIZE 1000
TABLESPACE users
STORAGE (
    INITIAL 100k
    NEXT 100K
    MINEXTENDS 2
    MAXEXTENDS 20
    PCTINCREASE 30);
```

Cria o cluster alug_socio, sendo a cluster key o cod_socio

Estimativa do espaço necessário (em média) para guardar uma chave do agrupamento e os registos a ela associados

Usado pelo ORACLE para limitar o número de chaves do agrupamento que podem ser atribuídas a um dado bloco

Após esta definição, não tem efeito a definição de parâmetros similares nas tabelas

Clusters

Agrupamentos de dados em um ou mais blocos de dados

Tipos de Clusters

▶ Cluster Indexados

- ▶ Cluster habituais
- ▶ Agrupamentos em que o acesso aos dados é efetuado por índices

▶ Hash Cluster

- ▶ Agrupamentos em que o acesso aos dados é efetuado por uma função de Hashing

Clusters Indexados

- ▶ O acesso dentro cluster é feito através de um índice construído com base nos valores da cluster key
- ▶ Necessidade de existir um índice sobre o cluster senão os dados não se encontram acessíveis (IMPORTANTE)

```
CREATE CLUSTER nome  
(coluna_que_compo_e_a_chave tipo)  
PCTFREE valor1  
PCTUSED valor2  
SIZE valor  
TABLESPACE nom_tab;
```

```
CREATE CLUSTER socio  
(cod_socio number(3))  
PCTFREE 15  
SIZE 1000  
TABLESPACE biblio;
```

▶ 50

2022/23 - LEI - AABD - Índices e Clusters

Clusters Indexados

Criação de um Índice de um cluster

```
CREATE INDEX ind_alug_soc ON CLUSTER alug_socio  
TABLESPACE index  
STORAGE (  
    INITIAL 20K  
    NEXT 20K  
    PCTINCREASE 75 )  
PCTFREE 0;
```

Cria o índice do cluster
(sobre a cluster key)

O índice pode e deve ficar
noutro tablespace
diferente do da tabela, por
questões de velocidade

Só depois do índice criado, é que podem
ser introduzidos dados no cluster

▶ 51

2022/23 - LEI - AABD - Índices e Clusters

Clusters Indexados

Considerações sobre os índices de cluster

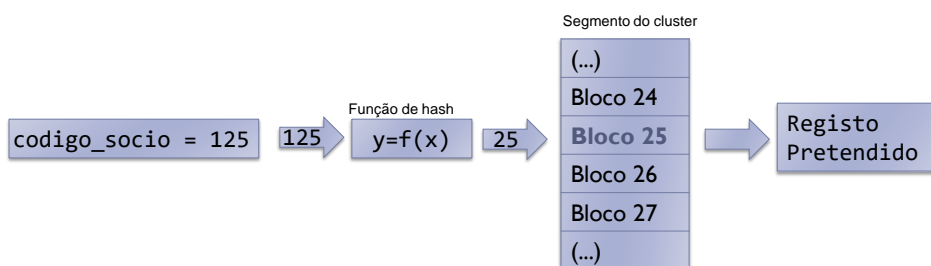
- ▶ criado e armazenado num segmento de índice
 - ▶ é gerido pelo sistema
-
- ▶ É obrigatório para se poder aceder aos dados
 - ▶ Tem que ser criados antes de introduzir os dados
 - ▶ Valores nulos da chave do agrupamento também tem uma entrada no índice de agrupamento no índice no agrupamento
 - ▶ quando o índice do agrupamento é eliminado os dados deixam de estar disponíveis até que o novo índice seja criado

▶ 52

2022/23 - LEI - AABD - Índices e Clusters

Hash clusters

- ▶ Função de hash
 - ▶ Gera uma distribuição de valores numéricos com base em valores concretos da chave do cluster



▶

54

2022/23 - LEI - AABD - Índices e Clusters

Hash clusters

- ▶ Pesquisar ou gravar um registo
 - ▶ Numa tabela ou cluster indexado
 - ▶ Um ou mais acessos ao disco para procurar ou inserir o valor chave no índice
 - ▶ Outro acesso ao disco para escrever o registo na tabela ou cluster
 - ▶ Num hash cluster
 - ▶ Uso da função de hash para localizar o registo (sem acesso ao disco)
 - ▶ Apenas um acesso para ler ou gravar o registo



2022/23 - LEI - AAM - Índices e Clusters

Quando usar hash clusters?

- ▶ Situações em que é vantajoso
 - ▶ A maioria das pesquisas têm restrições de igualdade na chave do cluster

```
SELECT ... WHERE cluster_key = ...;
```

- ▶ As tabelas são eminentemente estáticas em tamanho
 - ▶ O número de registos e espaço necessário pode ser determinado à partida



2022/23 - LEI - AAM - Índices e Clusters

Quando usar hash clusters? (cont.)

- ▶ Situações em que **não** é vantajoso
 - ▶ A maioria das pesquisas devolve registos que correspondem a uma gama de valores da chave de cluster

```
SELECT ... WHERE cluster_key < ...;
```
 - ▶ A tabela não é estática, mas cresce continuamente
 - ▶ Não é possível predeterminar o espaço necessário para o cluster
 - ▶ Pesquisas frequentes que façam *scan* completo à tabela e a tabela está preenchida de forma esparsa
 - ▶ Não é possível pré-alocar o espaço necessário ao cluster

▶ 57

57

2022/23 - LEI - AABD - Índices e Clusters

Criar hash clusters

```
CREATE CLUSTER socio
(cod_socio number(3))
HASH IS mod(cod_socio,100)
HASHKEYS 100
PCTFREE 15
SIZE 1000
TABLESPACE biblio;
```

Chave do cluster.

Função de hash.
Omitindo HASH IS é usada uma função interna.

Nº máximo de valores de hash gerados pela função.

Nº médio de bytes ocupados por um valor da chave do cluster e respectivas linhas.

```
CREATE TABLE socios(
  cod_socio number(3),
  nome varchar(30),
  idade number(4)
  CLUSTER socio(cod_socio);
```

▶ 58

2022/23 - LEI - AABD - Índices e Clusters

Clusters

Como criar tabelas usando os clusters?

- ▶ As tabelas são criadas normalmente.
- ▶ A maioria dos parâmetros não podem ser definidos, tais como:
 - ▶ os parâmetros de STORAGE
 - ▶ o tablespace onde colocar a tabela
- ▶ Com a definição do cluster já se está a definir o segmento que conterà as tabelas

```
CREATE TABLE nome (  
    coluna1 tipo,  
    coluna2 tipo,  
    ...)  
CLUSTER nome_cluster(coluna1)
```

```
CREATE TABLE socios (  
    cod_socio number(3),  
    nome varchar(30),  
    idade number(4))  
CLUSTER socio(cod_socio);
```

▶ 59

2022/23 - LEI - AABD - Índices e Clusters

Clusters

Remoção dos clusters?

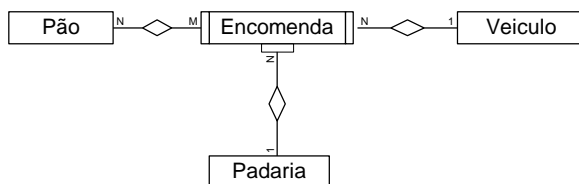
```
DROP CLUSTER nome  
[INCLUDING TABLES [CASCADE CONSTRAINTS]]
```

- ▶ É necessário apagar as tabelas que fazem parte do cluster, ou então deve-se incluir a cláusula INCLUDING TABLES para que estas sejam apagadas antes do cluster
- ▶ Se outras tabelas estiverem a referenciar estas tabelas deve-se incluir a cláusula CASCADE CONSTRAINTS para eliminar as constraints existentes

▶ 60

2022/23 - LEI - AABD - Índices e Clusters

Que tipo de otimizações faria nesta base de dados?



Pão < id_pão, tipo, peso, >

Encomenda < id_encomenda, data, hora, local_entrega, ..., ***id_padaria, matricula*** >

Enc_pao < id encomenda, id pao, quantidade, >

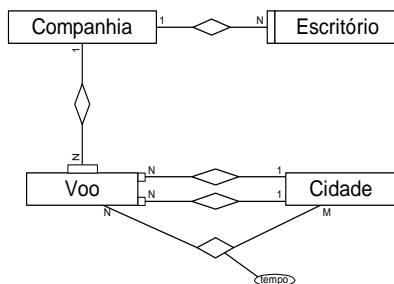
Padaria < id_padaria, nome, morada, ... >

Veiculo < matricula, marca, modelo, tara,>

61

2022/23 - LEI - AABD - Indices e Clusters

Que tipo de otimizações faria nesta base de dados?

**Companhia** < sigla_comp, nome, n_avioes, >

Escritório <id_escritório, cidade, país, morada,, **sigla_comp**>

Cidade < id_cidade, nome, pais, >

Voo < id_voo, tempo_estimado, **sigla_comp**, **id_cidade_origem**, **id_cidade_destino**.... >

Escala < id_voo, id_cidade, tempo_permanencia >

62

2022/23 - LEI - AABD - Indices e Clusters