# Classificação e Pesquisa de Dados

Aula 02

Organização de Arquivos: Discos magnéticos, Pesquisa de Dados em Arquivos

UFRGS

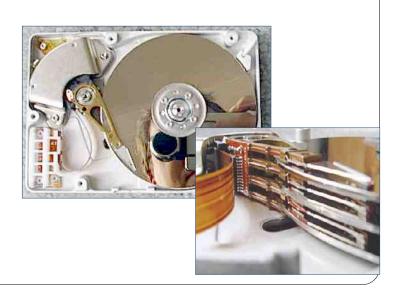
INF01124

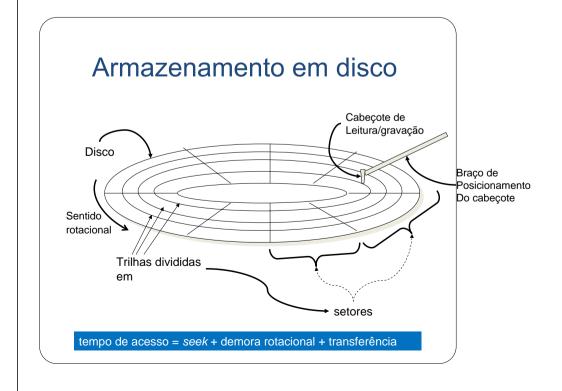
# Organização de Arquivos

- Propósito:
  - Estudo de técnicas de armazenamento e recuperação de dados em memória secundária
  - O que fazer quando os dados não cabem na memória principal?
  - Como manter os dados entre execuções dos programas?
- Operações importantes:
  - · Criar, destruir, abrir e fechar arquivos
  - Inserir, excluir, alterar e consultar registros e campos
- Principais técnicas: arquivos seqüenciais, indexados, invertidos, texto...

Instituto de Informática - UFRGS

### Estrutura do disco





## Trilhas, setores e *clusters*



# Características do armazenamento em disco

· Várias superfícies de gravação/leitura



efinido como sendo um conjunto enticalmente alinhadas e com netro

 Registros que pertencem a um mesmo cilindro n\u00e3o requerem deslocamento do mecanismo de acesso!

Instituto de Informática - UFRGS

## Sobre o acesso aos dados

- Antigamente, o acesso aos dados era feito indicando-se o cilindro, a cabeça e setor/bloco (CHS) onde encontravase o dado desejado
- Com o advento das tecnologias ATA e SCCI, a geometria do disco mudou e o acesso de baixo nível ficou mais distante do programador (abstração da geometria), mas conhecê-la é importante para determinar e desenvolver algoritmos Instituto de Informática - UFRGS

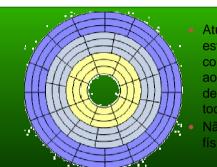
# Como aumentar a capacidade de armazenamento





### **Problema**

- Cada fabricante ou modelo tem uma geometria diferente, e gerenciar isso é muito complexo
- Solução ZBR (Zoned Bit Recording):



- Atualmente os discos utilizam estruturas internas complexas, como a exemplificada na figura ao lado, para manter uma alta densidade de gravação em toda a superfície
- Não há mais uma estrutura física visível pelo SO

### Sobre o acesso aos dados

- Atualmente, discos rígidos possuem microprocessadores inteligentes e que mapeiam sua geometria para um padrão conhecido
- Todos os discos são vistos como setores ou blocos seqüenciais (LBA – Logical Block Adressing)

#### Tarefa:

- Leia os artigos de leitura recomendada para esta aula procurando entender como a geometria atual funciona e qual o papel da controladora, drivers, SO e programador neste contexto
- Identifique o que afeta o desempenho e a transferência de dados.

Instituto de Informática - UFRGS

# Armazenamento em disco: características

- Custo de acesso ao disco:
  - Tempo de busca (seek time): tempo de deslocamento do mecanismo de acesso de uma trilha para outra
  - Latência rotacional: tempo para que o inicio do bloco que contenha o registro a ser lido passe pelo cabeçote de leitura/gravação
  - Tempo de transferência: quantidade de tempo necessário para que um bloco (ou setor) seja transferido para um buffer de memória

### Conceitos

- Arquivo
  - Seqüência de entradas ou registros
  - Registra informações sobre várias entidades de um mesmo tipo
- Entradas ou registros
  - Constituída por um ou mais campos (ou atributos), por exemplo:
  - <Número, nome, data-nasc, salário>

Instituto de Informática - UFRGS

## Terminologia empregada

- Arquivo
  - Coleção de registros lógicos
- Registro lógico
  - Seqüência de campos ou atributos da entidade ou objeto sendo modelado
- Campo
  - Corresponde a cada uma das informações que se deseja modelar a respeito da entidade ou objeto considerado
  - Ex.: Nome, salário, cargo, etc.

Instituto de Informática - UFRGS

## Tipos de acesso a arquivos

- · Acesso ao arquivo inteiro
  - · Computação gráfica
  - Engenharia de software
  - CAD
  - Documentos
- Acesso a registros
  - Banco de dados
  - · Processamento transacional

#### Arquivo de funcionário Registro lógico Campo: cargo 1012 Maria 01/10/1982 1250,00 analista Tipo: alfanumérico Comprimento: 30 bytes Campo: salário Tipo: numérico real Comprimento: 7 bytes Campo: data-nasc Tipo: data Comprimento: 8 bytes Campo: nome Tipo: alfabético Comprimento: 35 bytes Campo: número →Tipo: numérico inteiro Comprimento: 4 bytes Instituto de Informática - UFRGS

## Terminologia empregada

- Registro físico ou bloco
  - O armazenamento de um arquivo é feito, via de regra, por blocos de registros lógicos.
  - Um bloco corresponde a quantidade de dados transferidos em um acesso simples.
  - Múltiplos blocos contíguos podem ser transferidos em conjunto.
  - Um bloco de registros lógicos corresponde a um registro físico.

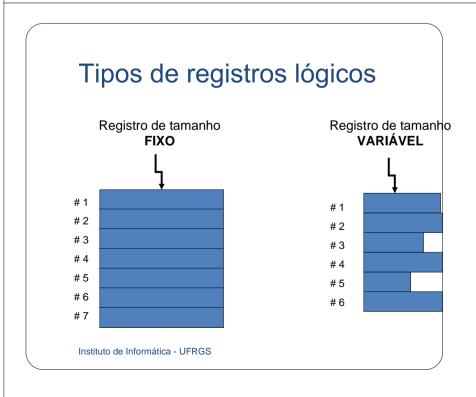
Instituto de Informática - UFRGS

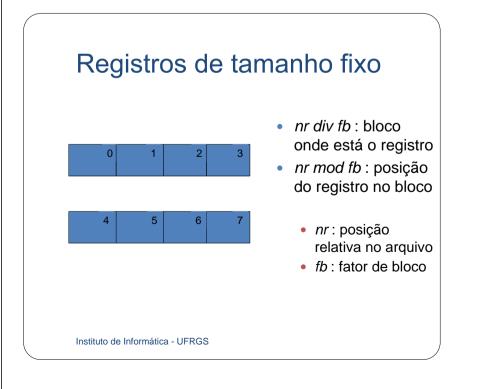
## Tecnologia

- Em cada leitura ou gravação é lido ou gravado um bloco e não apenas um registro lógico.
- Usualmente, o tamanho dos blocos é constante e múltiplo de uma unidade de armazenamento do meio físico utilizado.
- Fator de bloco: número inteiro de registros lógicos por bloco.

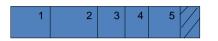
Instituto de Informática - UFRGS

### Estrutura de acesso Registro Um registro Programa **Buffers** Sistema operacional Um bloco em cada buffer Arquivo físico **Blocos** Arquivo físicos Instituto de Informática - UFRGS





# Registros de tamanho variável





Instituto de Informática - UFRGS

- Problemas de acesso, exceto o serial
- Problemas de atualização, podendo haver aumento de tamanho do registro
- As vezes é conveniente deixar alguns itens (campos) no formato fixo por serem usados com grande freqüência (dividir o arquivo).
- Assim, evita-se a expansão à cada vez que são manipulados. Os dados complementares são acessados por meio de

# Pesquisa de Dados em Arquivos

- Recuperação
  - · Pesquisa seqüencial em arquivos
  - Pesquisa binária em arquivos
  - Árvores
  - Transformação de chave (hashing)

Instituto de Informática - UFRGS

# Exemplo: arquivo de funcionários

- Contém informações homogêneas sobre os funcionários de uma empresa
- Uma entrada ou registro por funcionário
- Deve ser possível encontrar um determinado funcionário no arquivo

# Exemplo de arquivo de funcionário

Número	Nome	Data-Nasc	Salário
10130	José	250780	450
10850	Paulo	231281	1040
12700	Mário	100678	694
12800	José	200580	1200
13400	Maria	070177	823
14000	Ana	150379	250

Instituto de Informática - UFRGS

### Conceitos

 Chave é o atributo que permite a recuperação de um registro

Ex.: Nome

Número	Nome	Data-Nasc	Salário
10130	José	250780	450
10850	Paulo	231281	1040
12700	Mário	100678	694
12800	José	200580	1200
13400	Maria	070177	823
14000	Ana	150379	250

Instituto de Informática - UFRGS

## Conceitos (Cont.)

- Chave secundária permite a recuperação de registros, mas não apresenta a característica de unicidade.
  - Ex.: cargo, salário
- Chave de um registro é o valor da sua chave primária.
- Chave de pesquisa é o atributo utilizado para fazer a recuperação de um registro (chave primária) ou de múltiplos registros (chave secundária).

Conceitos (Cont.)

- Chave primária é um atributo que apresenta obrigatoriamente um valor diferente para cada registro, unicidade
  - Número do funcionário
  - Nome e Data-Nasc (?)
- Uma chave primária deve ser mínima, isto é, não deve conter campos supérfluos

Instituto de Informática - UFRGS

## Conceitos (Cont.)

- Arquivos estáticos
  - Não são alteradas por longos períodos de tempo.
  - Exemplos: CEP
- Arquivos dinâmicos
  - Alterações e consultas se alternam.
  - Exemplo: arquivo de reservas em v\u00f3os comerciais.

Instituto de Informática - UFRGS

## Leitura recomendada

- You don't know Jack about disks <a href="http://queue.acm.org/detail.cfm?id=864058">http://queue.acm.org/detail.cfm?id=864058</a>
- LBA (Wikipedia)
  <a href="http://en.wikipedia.org/wiki/Logical\_block\_ad-dressing">http://en.wikipedia.org/wiki/Logical\_block\_ad-dressing</a>