

Algoritmos e Estruturas de Dados II

Elton M. Cardoso

24 de maio de 2015

Sumário

1	Introdução	1
1.1	Operações fundamentais de processamento de dados para arquivos	2
1.1.1	Arquivos lógicos e Arquivos físicos.	2
1.2	Tecnologias de armazenamento	3
1.2.1	Fita magnética	4
1.2.2	Disco Rígido (HD)	4
1.2.3	Disco Ótico	4
1.2.4	Memórias Flash	5
2	Fundamentos de Estruturas de dados para Arquivos	7
3	Campos e Registros	9
4	Ordenação de memória secundária	11

Capítulo 1

Introdução

Em sua concepção, os primeiros computadores possuíam um desenho muito mais focado no cálculo (computação) do que no armazenamento de dados propriamente ditos. Foi em 1945 que surgiu uma arquitetura de computação que separava a unidade de processamento da memória, a arquitetura *Von Neumann*. Nessa nova arquitetura dados são separados do programa, podendo potencialmente serem armazenados em um módulo de memória fora da unidade de processamento.

Computadores modernos são equipados com níveis de memória que podem ser:

- Registradores: São unidades de memória que armazenam resultados intermediário das operações realizadas pelo processador. Em geral são limitados a poucas unidades e são situados dentro do processador.
- Memória *cache* : É uma memória intermediária entre os registradores e a memória principal para melhor eficiência. A memória *cache* é situada dentro do próprio processador.
- Memória principal: A memória principal, diferentemente das duas primeiras, é situada fora do processador e é conectada a ele por meio de barramentos de dados.

Tantos registradores, memória *cache* e memória principal são memórias voláteis e não capazes de reter os dados após o sistema ser desligado. Há no entanto uma memória não volátil, também chamada de memória secundária, que preserva seu conteúdo após ser desligada.

Para acessar dados tanto na memória principal quanto na memória secundária a unidade de processamento necessita se comunicar com essas

memórias, por meio de um barramento, informando quais dados deve ser transferidos para a memória *cache* ou para os registradores. O processo de transferência dos dados em si pode ser uma operação mais complexa que pode requerer uma certa quantidade de tempo para se concretizar.

Idealmente processador e memórias deveriam operar de modo que o processador não tivesse que esperar até que um dado requisitado fosse transferido para a memória interna do processador. Na prática isso não acontece e o processo de transferência dos dados pode levar vários ciclos de operação do processador.

O problema é agravado quando o dispositivo de memória externa é lento e requiere muito tempo para acessar e recuperar os dados, como acontece por exemplo com discos rígidos. Em geral memórias secundárias são mais lentas que a memória principal, a ponto de ser necessário levar esse tempo em consideração para evitar que o que nos leva ao tópico central desse texto. Técnicas para recuperação eficiente de dados em memória secundária.

Antes que o assunto seja apresentada de modo mais aprofundado, vamos apresentar os conceitos preliminares necessários.

1.1 Operações fundamentais de processamento de dados para arquivos

Dados são armazenados em memória secundária na forma de arquivos, que são basicamente sequencia de bytes gravados na memória secundária. No entanto a aplicação e o hardware tem visões distintas do arquivo.

1.1.1 Arquivos lógicos e Arquivos físicos.

Quando fala-se de um arquivo armazenado em um disco, por exemplo, nos referimos a um coleção de bytes gravados nesse disco. Diz-se então que um arquivo existe fisicamente, e nos referimos a esse conjunto de bytes como um arquivo físico.

No entanto aplicações tem uma visão bastante distinta do que é um arquivo. Um aplicação acessa arquivos por meio do sistema operacional e pode apenas ler e escrever dados para o arquivo, mas a aplicação não sabe nada sobre onde está o arquivo. Para a aplicação o arquivo é apenas um canal por onde ela pode ler e escrever dados. Desse modo esse canal pode estar conectado a um arquivo físico, ou pode estar conectado a um teclado, um socket de rede

etc. Esse canal é utilizado pela aplicação é comumente referenciado como *arquivo lógico*.

Antes que um programa possa utilizar um arquivo, ele deve solicitar ao sistema operacional para associar um dado arquivo físico com o canal de comunicação utilizado pelo programa. Em várias linguagens de programação isso é feito por meio de um sistema

1.2 Tecnologias de armazenamento

Existem, atualmente, diversas alternativas de tecnologias de memória secundária. Cada uma dessas alternativas tem suas próprias peculiaridades, se mostrando mais ou menos eficiências para um determinado propósito específico, ao passo em que também podem ser mais ou menos acessíveis financeiramente.

Algumas dessas tecnologias já não são empregadas em larga escala hoje em dia, com exceção de alguns nichos mais específicos, devido a existências de tecnologias melhores. Ainda sim tais tecnologias são apresentadas e discutidas pela sua contribuição histórica.

Outras tecnologias de memória secundária são bem mais recentes, como é o caso das memórias *flash*. Tais dispositivos tem ganhado cada vez mais atenção e espaço no mercado devido a seu baixo custo e grande velocidade de operação. Essas tecnologias são discutidas de maneira mais modesta.

Independentemente do tipo de tecnologia empregada para armazenamento e recuperação de informações, é importante ter em mente que o uso de estruturas de dados eficientes é de grande importância para que um sistema de processamento de dados apresente um bom desempenho.

Quando utiliza-se um sistema que tenha que realizar consultas em dados armazenados, espera-se que tal sistema seja capaz de recuperar as informações necessárias em um curto intervalo de tempo, também espera-se que tais sistemas sejam capazes de sumarizar as informações armazenadas de maneira compacta e compreensível. Para que se possa atingir esse objetivo é necessária analisar quais operações são mais custosas ao processo de recuperação de informação.

A seguir discutimos sucintamente algumas das tecnologias de armazenamento de dados.

1.2.1 Fita magnética

Subtilizadas, mas ainda presentes em algumas empresas e utilizadas como meio de backup. A tecnologia é de armazenamento magnético do dado impresso em bandas impregnadas com material ferromagnético.

Esses dispositivos foram explorados para armazenamento (e comercialização) de áudio, vídeo (fitas VHS) e como meios de backup, devido a grande capacidade de armazenamento e baixo custo.

Não existe um sistema de endereçamento na fita magnética e desse modo

No entanto o acesso a informação armazenada tem de ser feito de modo sequencial, i.e., para acessar a uma informação armazenada em uma dada posição na fita, é necessário passar por todos os dados armazenados fisicamente antes dessa posição. Isso torna fitas magnéticas menos eficientes para utilização em aplicações mais dinâmicas, onde se necessário acessar dados de maneira aleatória.

1.2.2 Disco Rígido (HD)

É dispositivo de armazenamento que utiliza vários discos, empilhados, recobertos com material magnetizável. A leitura e escrita é feita por meio de um atuador que se situa próximo a superfície de cada disco. Esse atuador é preso a um braço mecânico móvel de modo que o atuador possa ser mover entre a borda externa e o centro do disco. Os discos giram a uma determinada velocidade permitindo que o atuador acesse qualquer região na superfície do disco. Ainda é um dos dispositivos de memória secundária principal da maioria dos computadores populares modernos.

1.2.3 Disco Ótico

Tais dispositivos utilizam um laser para imprimir marcas em um substrato abaixo da superfície transparente de um disco. Essas marcas podem, posteriormente, serem detectadas pelo laser do dispositivo permitindo assim a leitura dos dados. O dispositivo normalmente possui uma plataforma móvel que hospeda um laser. Essa plataforma descreve uma trajetória entre o centro e a borda do disco. O disco é encaixado em um motor que o faz girar, permitindo assim que o dispositivo possa acessar qualquer ponto na superfície do disco.

1.2.4 Memórias Flash

Existem várias tipos de memórias *flash* e esses dispositivos geralmente empregam algum mecanismo elétrico para armazenamento dos dados. Tais dispositivos tem se tornado muito atrativos pelo seu baixo tempo de acesso, além disso a ausência de dispositivos mecânicos, como no disco rígido e discos ópticos o torna menos suscetível a falhas mecânicas. Memórias flash ainda são, no entanto, ligeiramente mais caras que as demais tecnologias para serem utilizadas em larga escala. Além disso várias dessas memórias.

Capítulo 2

Fundamentos de Estruturas de dados para Arquivos

```
1  // Calsse campo
2  public abstract class Field {
3      public abstract void read(ReadStream i);
4      public abstract void write(WriteStream i);
5      public abstract byte[] pack();
6      public abstract void read(byte b[]);
7  }
```

Código 2.1: Calsse java para campo

Capítulo 3

Campos e Registros

Capítulo 4

Ordenação de memória secundária