



# INSIGHT

DATA SCIENCE LABORATORY  
FEDERAL UNIVERSITY OF CEARÁ

# Capítulo 1 – Introdução

Introdução aos Sistemas Operacionais

# Capítulo 1: Introdução

---

- O que é um SO faz?
- Organização de um sistema de computação
- Arquitetura de um sistema de computação
- Estrutura de um Sistema Operacional
- Operações de um Sistema Operacional
- Gerenciamento de Processos
- Gerenciamento de Memória
- Gerenciamento de Armazenamento
- Proteção e Segurança
- Sistemas Distribuídos
- Sistemas de uso específico
- Ambientes de Computação

# Objetivos

---

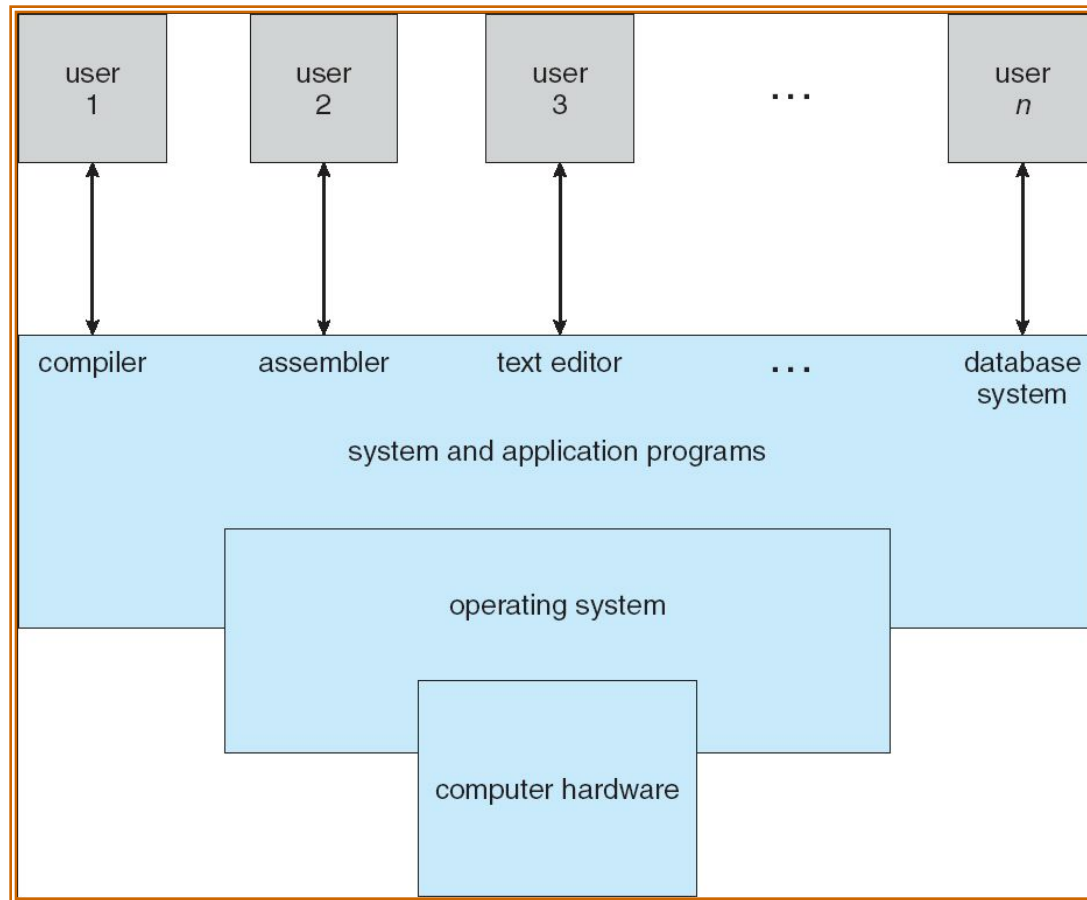
- Prover um tour sobre os mais importantes componentes dos sistemas operacionais
- Prover cobertura da organização básica de um sistema de computação

# O Que é um Sistema Operacional (SO)?

---

- Um programa que age como um intermediário entre um usuário do computador e o hardware do computador.
- Objetivos do Sistema Operacional:
  - Executar programas dos usuários e facilitar a solução de problemas dos usuários.
  - Tornar o sistema de computação mais conveniente para uso
- Usar o hardware do computador de maneira eficiente.

# 4 Componentes de um sistema de computação



# Estrutura de um sistema de computação

---

- Sistema de Computação pode ser dividido em 4 componentes
  - Hardware – prove os recursos básicos de computação
    - CPU, memória, dispositivos de E/S
  - Sistema Operacional
    - Controla e coordena o uso do hardware entre diversas aplicações e usuários
  - Programas de Aplicação – define as formas pelas quais os recursos do sistema são usados para resolver problemas computacionais dos usuários
    - Processadores de texto, compiladores, browsers para web, compiladores, SGBDs, Jogos por computador
  - Usuários
    - Pessoas, Máquinas, outros computadores

# Definição de Sistema Operacional

---

- SO é um alocador de recursos
  - Gerencia todos os recursos
  - Decide entre requisições conflitantes para o uso eficiente e adequado do sistema
  
- SO é um programa de controle
  - Controla a execução dos programas para prevenir erros e uso impróprio do computador



# Definição de Sistema Operacional (Cont.)

---

- Nenhuma definição é universalmente aceita
- “Cada fornecedor entrega um SO que é apenas uma boa aproximação da definição geral”
  - No entanto variam bastante
- “O único programa que roda o tempo inteiro no computador é o kernel (núcleo). Todo resto ou é um programa do sistema (entregue junto com o SO) ou é uma aplicação do usuário

# Inicialização do Computador

---

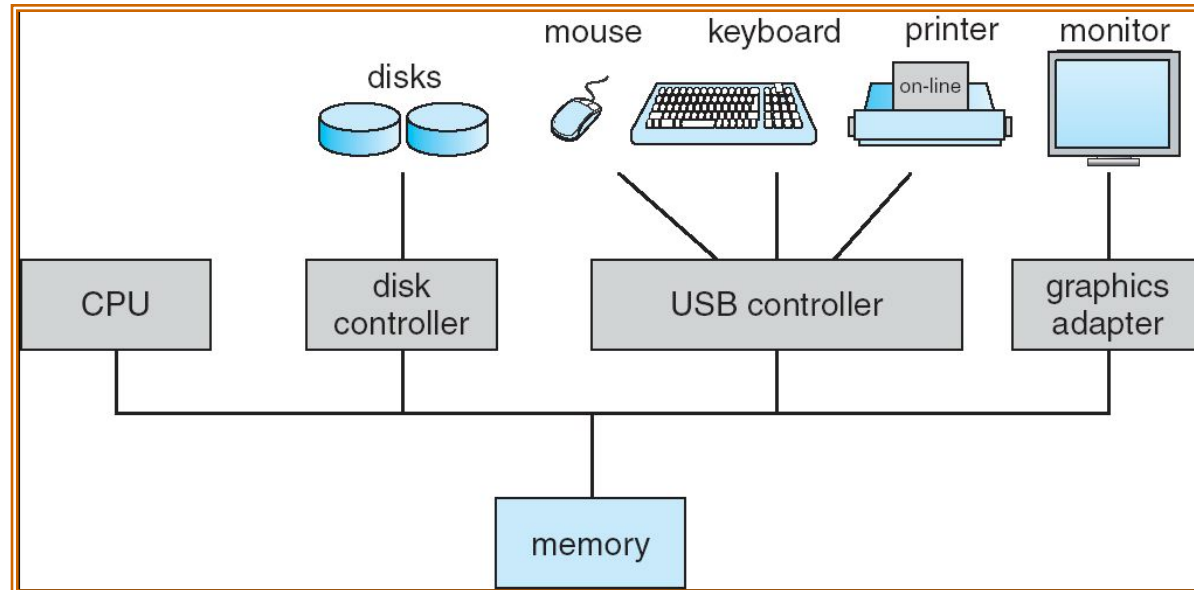
- **Programa de bootstrap** é carregado no momento da início ou reinício do computador
  - Tipicamente armazenado em ROM ou EPROM, geralmente conhecido como **firmware**
  - Inicia todos os aspectos do sistema
  - Carrega o kernel do sistema operacional e inicia sua execução

ROM - Read Only Memory

EPROM - Erasable Programmable Read Only Memory

# Organização de um sistema de computação

- Operação de um sistema de computação
  - Um ou mais CPUs, dispositivos de controle conectados através de um barramento (bus) provendo acesso a memória compartilhada)
  - Execução concorrente das CPUs e dos dispositivos competem pelos ciclos de memória



# Operação de um Sistema de Computação

---

- ❑ Dispositivos de E/S e a CPU podem executar concorrentemente
- ❑ Cada dispositivo controlador é responsável por um tipo particular de dispositivo.
- ❑ Cada dispositivo controlador tem um buffer local.
- ❑ A CPU move dados de/para memória principal para/dos buffers locais
- ❑ E/S é realizada do dispositivo para o buffer local do controlador.
- ❑ Dispositivo controlador informa CPU que ele terminou sua operação gerando uma interrupção.

# Funções comuns de Interrupção

---

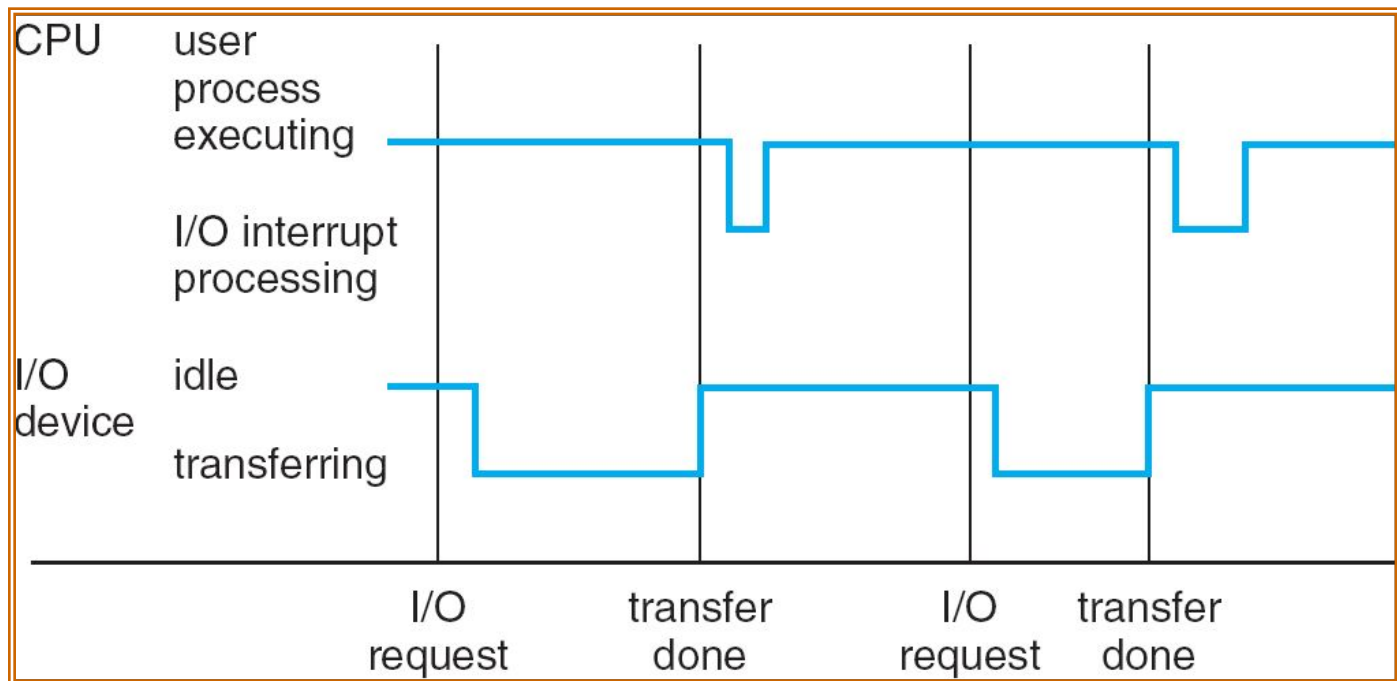
- Um sistema operacional é dirigido a interrupção.
- Interrupção transfere controle para a rotina do serviço de interrupção, através do vetor de interrupções, o qual contém os endereços de todas as rotinas de serviços.
- A arquitetura de interrupção deve salvar o endereço da instrução interrompida.
- Interrupções que chegam são desabilitadas enquanto outra interrupção está sendo processada para prevenir perda de interrupção.
- Uma *trap* é uma interrupção gerada por software causadas por um erro ou por uma requisição do usuário.

# Tratamento de Interrupção

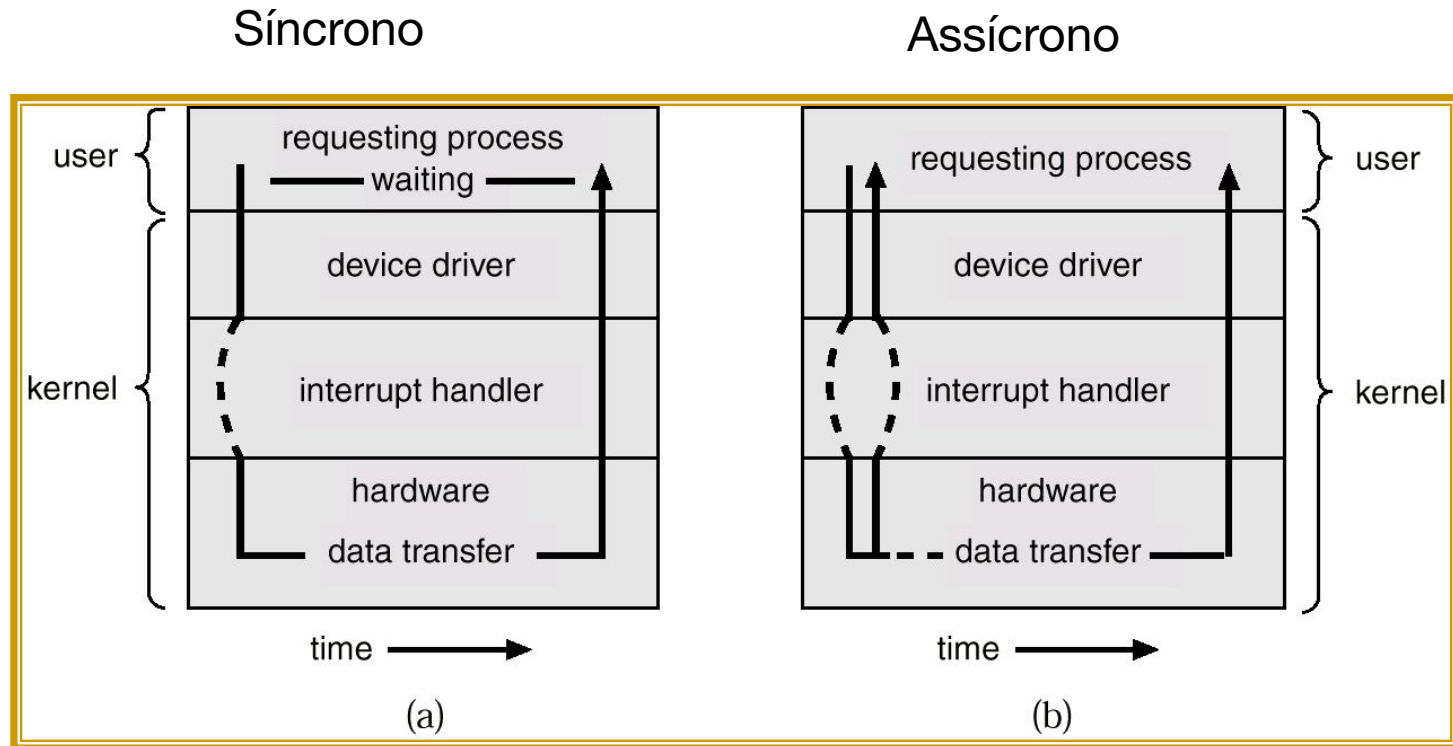
---

- O sistema operacional preserva o estado da CPU armazenando registradores e contadores de programa.
- Determina qual o tipo de interrupção ocorreu:
  - *Polling*
  - *Sistema de interrupção vetorizado*
- Pedacos de códigos separadas determinam qual ação deve ser tomada para cada tipo de interrupção

# Linha de tempo de uma interrupção



# Dois métodos de E/S





# Estrutura de E/S (dois métodos)

---

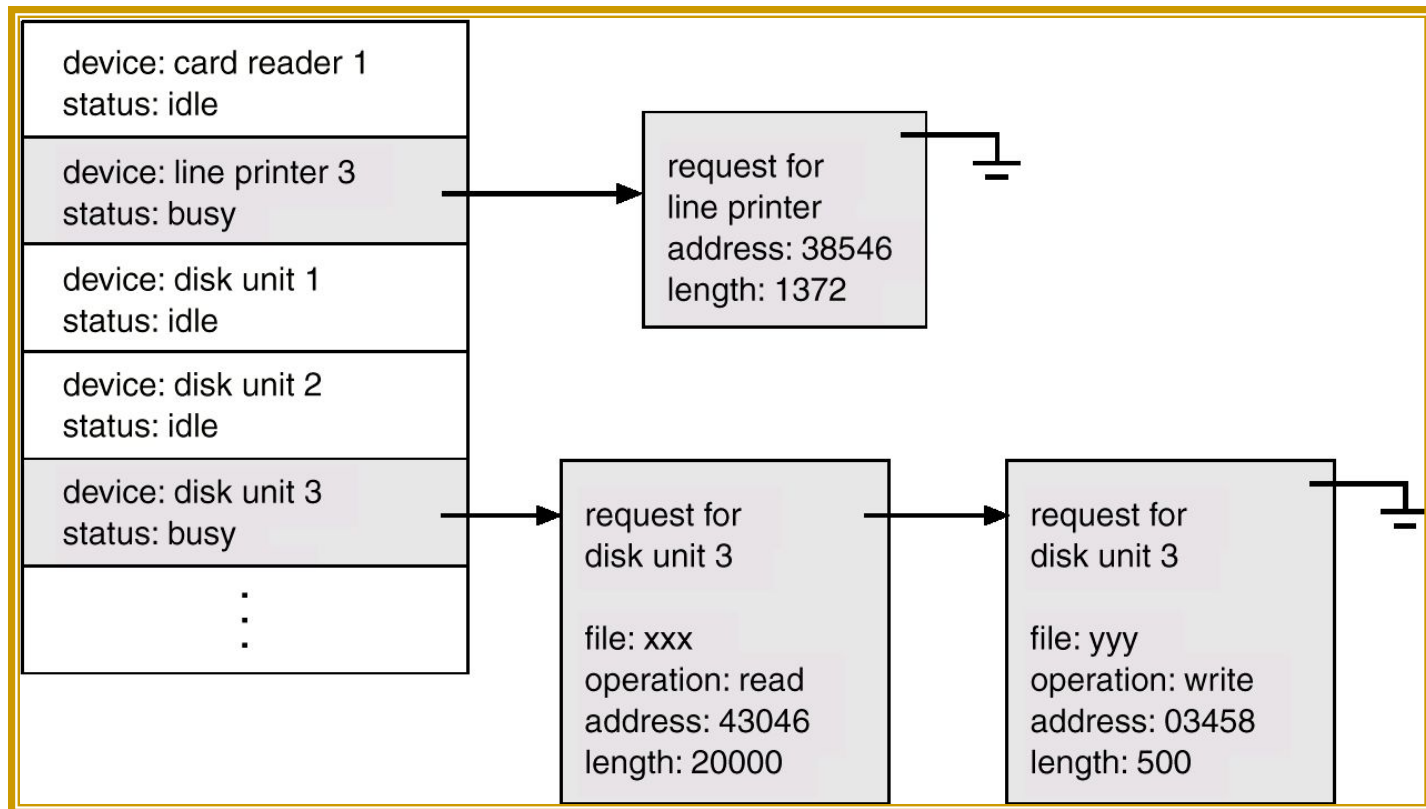
- SÍNCRONO: Depois que a E/S iniciou, o controle retorna para o programa do usuário somente após a finalização da operação de E/S.
  - Instrução de espera (wait) inativa a CPU até a próxima interrupção
  - Loop de espera (contenção para acesso à memória).
  - No máximo uma requisição de E/S é atendida por vez, nenhum processamento de E/S simultâneo ocorre .

# Estrutura de E/S (dois métodos)

---

- ASSÍNCRONO: Depois que a E/S iniciou, o controle retorna para o programa do usuário sem aguardar pela a finalização da operação de E/S.
  - *System call* – requisição ao sistema operacional para permitir o usuário esperar pela finalização da E/S.
  - Tabela de estado dos dispositivos contém entrada para cada dispositivo de E/S indicando seu tipo, endereço, e estado.
  - Sistema operacional indexa tabela de E/S para determinar os estado e modificar a entrada da tabela para incluir uma interrupção.

# Tabela de estado de dispositivo

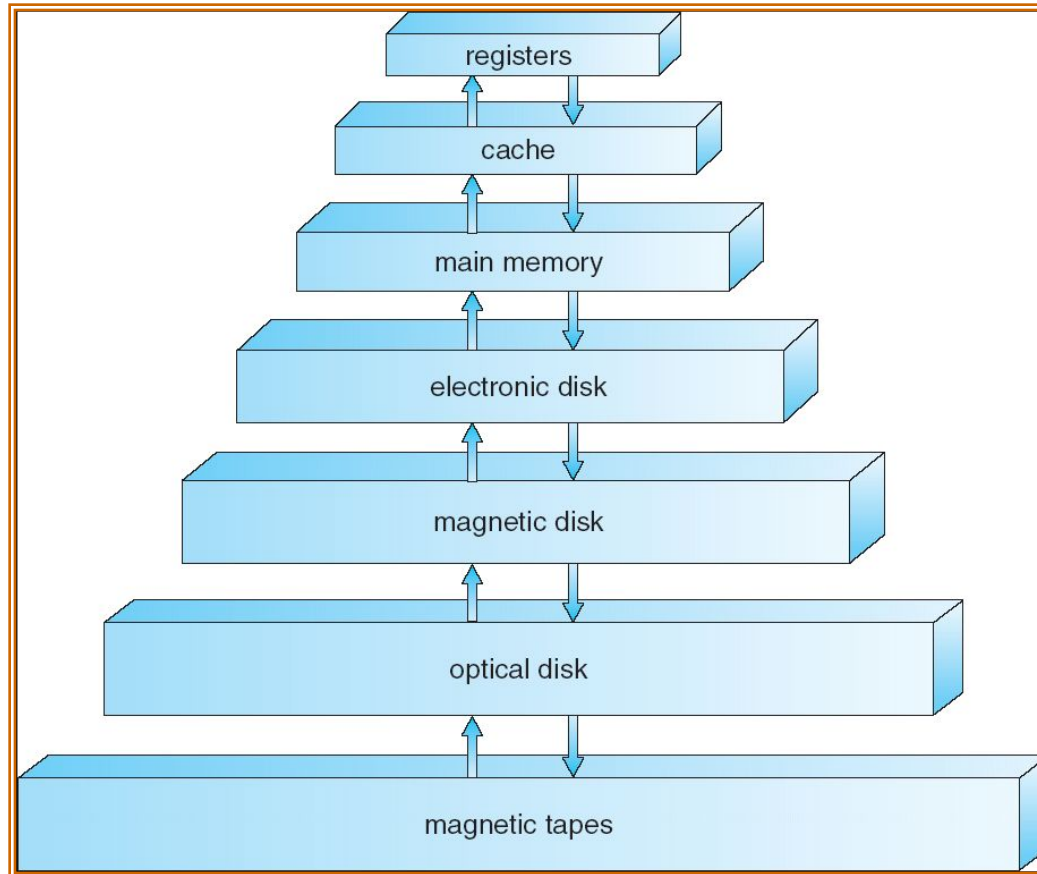


# Estrutura de Acesso Direto a Memória (DMA)

---

- Permite que certos dispositivos de hardware acessem a memória do sistema para leitura e escrita independentemente da CPU
- Usada por dispositivos de alta-velocidade habilitados a transmitir informação perto da velocidade da memória.
- Dispositivo de controle que transfere blocos de dados do buffer de armazenamento diretamente para memória principal sem intervenção da CPU.
- Somente uma interrupção é gerada por bloco, ao invés de uma interrupção por byte.

# Hierarquia de dispositivo de armazenamento



# Estrutura de Armazenamento

---

- Memória principal – a única mídia de longo armazenamento que a CPU pode acessar diretamente.
- Armazenamento secundário – extensão da memória principal que provê capacidade de armazenamento não volátil de larga escala.
- Discos magnéticos – metal rígido ou pratos de vidro cobertos com material de gravação magnético
  - A superfície do disco é logicamente dividida em trilhas, as quais são subdivididas em setores.
  - O controlador de disco determina a interação lógica entre um dispositivo e o computador

# Hierarquia de Armazenamento

---

- Sistemas de armazenamento são organizados em hierarquia.
  - Velocidade
  - Custo
  - Volatilidade
- *Caching* –copiando informação em um sistema de armazenamento rápido; memória principal pode ser vista com um último cache para armazenamento secundário.

# Caching

---

- Princípio importante, executado em vários níveis em um computador (no hardware, sistema operacional, software)
- Informação em uso copiada do armazenamento mais lento para o mais rápido temporariamente
- Armazenamento mais rápido (cache) checado primeiro para determinar se a informação está lá
  - Se ela está, a informação é usada diretamente do cache
  - Se ela não está lá, dados copiados para o cache e usados lá
- Cache menor do que armazenamento sendo cacheado
  - Gerenciamento de cache é um problema importante de projeto
  - Tamanho do cache e política de reposição



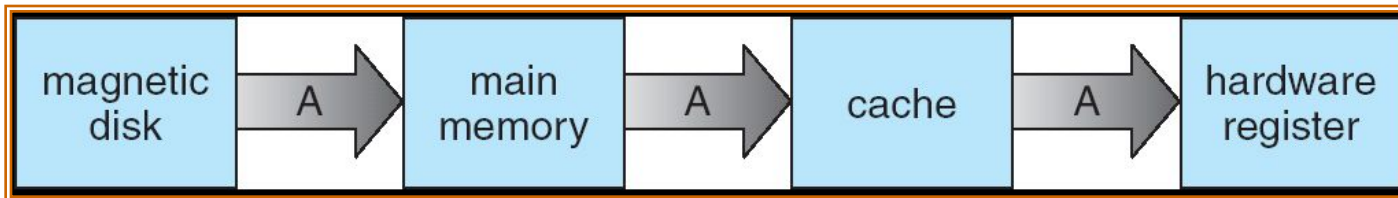
# Performance de vários níveis de armazenamento

- Movimento entre níveis da hierarquia de armazenamento pode ser explícita ou implícita

Level	1	2	3	4
Name	registers	cache	main memory	disk storage
Typical size	< 1 KB	> 16 MB	> 16 GB	> 100 GB
Implementation technology	custom memory with multiple ports, CMOS	on-chip or off-chip CMOS SRAM	CMOS DRAM	magnetic disk
Access time (ns)	0.25 – 0.5	0.5 – 25	80 – 250	5,000.000
Bandwidth (MB/sec)	20,000 – 100,000	5000 – 10,000	1000 – 5000	20 – 150
Managed by	compiler	hardware	operating system	operating system
Backed by	cache	main memory	disk	CD or tape

# Migração de inteiro A do disco para registrador

- Ambientes multitarefas devem ser cuidadosos para usar o valor mais recente independente do local onde ele está armazenado na hierarquia



- Ambientes com multiprocessador devem prover coerência entre o cache no hardware de tal forma que todas as CPUs tenham o valor mais recente em seu cache
- Situação em ambientes distribuídos é mais complicada ainda
  - Diversas cópias de um dado podem existir

# Estrutura do Sistema Operacional

---

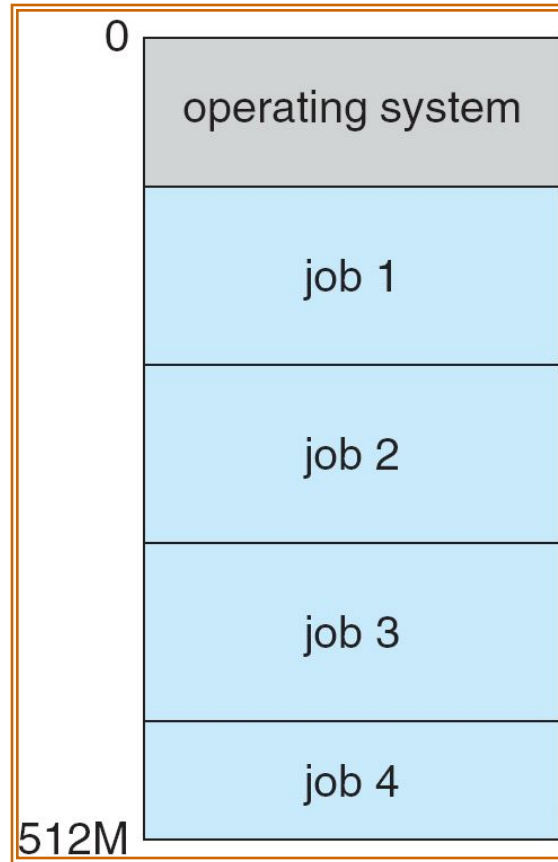
- **Multiprogramação** necessária para eficiência
  - Um único usuário não pode manter os dispositivos ocupados por todo tempo
  - Multiprogramação organiza jobs (código e dados) de tal forma que a CPU sempre tem algo para executar
  - Um subconjunto do total de jobs no sistema é mantido em memória
  - Um job selecionado e executado via escalonamento de job
  - Quando um job tem que aguardar, SO troca para outro job

# Estrutura de um Sistema Operacional (Cont.)

---

- **Timesharing (multitarefa)** é extensão lógica da multiprogramação na qual a CPU troca de job tão frequentemente que usuários podem interagir com cada job enquanto ele está executando, criando uma computação interativa
  - **Tempo de resposta** deve ser  $< 1$  segundo
  - Cada usuário tem no mínimo um programa executando em memória □ **processo**
  - Se existem diversos jobs prontos para executar ao mesmo tempo □ **CPU scheduling**
  - Se processos não cabem em memória, mecanismo de **swapping** é usado para movê-los para dentro e para fora da CPU para executá-los
  - **Memória Virtual** permite a execução de processos não armazenados completamente em memória

# Layout da Memória de um sistema de Multiprogramação



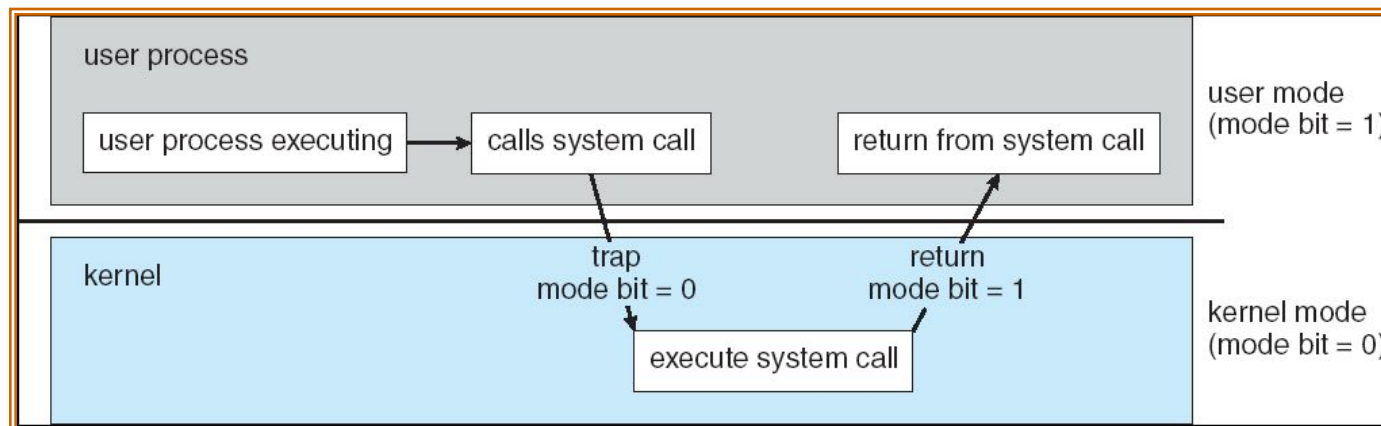
# Operações de um sistema operacional

---

- Dirigido por interrupção implementado por hardware
- Erro de software ou requisições criam exceções ou traps
  - Divisão por zero, requisição por serviço do sistema operacional
- Outros problemas de processo incluem loop infinito, processos modificando uns aos outros, ou o próprio sistema operacional
- **Operação Dual-mode** permite o SO proteger a si próprio e outros componentes do sistema
  - **User mode e kernel mode**
  - **Mode bit** provida por hardware
    - Provê habilidade para distinguir quando o sistema está executando em modo user code ou kernel code
    - Algumas instruções projetadas como privilegiadas, somente executáveis em kernel mode
    - Chamadas ao sistema modifica modo para kernel, retornando da chamada redefine para modo user

# Transição de Mode User para Kernel

- Temporizador (timer) para prevenir loop infinito / processos monopolizando recursos
  - Disparar interrupção após período específico
  - Sistema operacional decrementa contador
  - Quando contador zera gera uma interrupção
  - Definido antes do escalonamento de processos para retomar controle ou terminar programa que excedeu tempo alocado



# Gerenciamento de Processos

---

- Um processo é um programa em execução. É uma unidade de trabalho dentro do sistema. Programa é uma entidade passiva, processo é uma entidade ativa.
- Processos precisam de recursos para executarem suas tarefas
  - CPU, memória, E/S, arquivos
  - Inicialização de dados
- Finalização de processos requer a recuperação de qualquer recurso reusável
- Processo single-threaded tem um program counter (PC) especificando localização da próxima instrução a ser executada
  - Processos executam instruções sequencialmente, uma por vez, até a sua finalização
- Processos Multi-threaded possuem um program counter por thread
- Tipicamente um sistema possui vários processos, alguns usuários, algum sistema operacional executando com uma ou mais CPUs
  - Concorrência através da multiplexação de CPUs entre processos/threads



# Atividades de Gerenciamento de Processos

---

- O Sistema operacional é responsável pelas seguintes atividades em relação ao gerenciamento de processos:
- ❑ Criação e eliminação tanto de usuários como de processos do sistema
  - ❑ Suspensão e finalização de processos
  - ❑ Disponibilização de mecanismos para sincronização de processos
  - ❑ Disponibilização de mecanismos para comunicação de processos
  - ❑ Disponibilização de mecanismos para tratamento de deadlock

# Gerenciamento de Memória

---

- Todos os dados em memória antes e depois do processamento
- Todas as instruções em memória a serem executadas
- Gerenciamento de memória determina o que está na memória quando
  - Se está otimizando a utilização da CPU e o tempo de resposta para usuários
- Atividades do gerenciamento de memória
  - Rastrear quais partes da memória estão sendo usadas no momento e por quem
  - Decidir quais processos e dados (ou partes desses) para mover para dentro e fora da memória
  - Alocar e desalocar espaço de quando necessário

# Gerenciamento de Armazenamento

---

- SO provê uma visão uniforme e lógica da informação armazenada
  - Abstrai propriedades físicas para unidades de armazenamento lógica - arquivo
  - Cada mídia é controlada por um dispositivo (i.e., disk drive, tape drive)
    - Propriedade incluem velocidade de acesso, capacidade, taxa de transferência, método de acesso (sequencial ou randômico)
- Gerenciamento de sistema de arquivos
  - Arquivos usualmente organizados em diretórios
  - Controle de acesso na maioria dos sistemas para determinar quem acessa o que
  - Atividades do SO incluem
    - Criação e eliminação de arquivos e diretórios
    - Primitivas para manipular arquivos e diretórios
    - Mapear arquivos para armazenamento secundário
    - Copiar arquivos em mídia de armazenamento não volátil

# Gerenciamento de armazenamento massivo

---

- Usualmente discos usados para armazenar dados que não cabem em memória ou dados que devem ser mantidos por longo período de tempo.
- Gerenciamento próprio é de suma importância
- A velocidade do computador é limitada pelo subsistemas de disco e seus algoritmos
- Atividades do SO
  - Gerenciamento de espaço livre
  - Alocação de armazenamento
  - Escalonamento de disco
- Alguns armazenamento não precisam ser rápidos
  - Armazenamento terciário inclui fita magnética, armazenamento óptico
  - Ainda precisa ser gerenciado
  - Varia entre WORM (write-once, read-many-times) e RW (read-write)

# Subsistema de E/S

---

- Uma proposta do SO é esconder peculiaridades dos dispositivos de hardware do usuário
- Subsistema de E/S é responsável por
  - Gerenciamento de memória de E/S incluindo bufferização (armazenamento de dados temporários enquanto dados estão sendo transferidos), caching (armazenamento de partes de dado em armazenamento rápido para melhorar performance), spooling (a conexão entre a saída de um job e a entrada de outros jobs)
  - Interface geral para device-driver
  - Drivers para dispositivos de hardware específicos

# Proteção e Segurança

---

- **Proteção** – qualquer mecanismos para controlar acesso de processos ou usuários a recursos definidos pelo SO
- **Segurança** – defesa do sistema contra ataques internos e externos
  - Uma grande diversidade de ataques, incluindo denial-of-service, worms, vírus, roubo de identidade, roubo de serviço
- Sistemas geralmente primeiro distinguem o que os usuários podem fazer
  - Identidades dos usuários (IDS dos usuários, IDS de segurança) incluem nome e número associado, um por usuário
  - ID do usuário é então associado com todos os arquivos, processos deste usuário para determinar seu controle de acesso
  - Identificador de grupo (**group ID**) permite que um conjunto de usuários serem definidos e controles de gestão, então também associado com cada processo, arquivo
  - **Escalonamento de privilégio** permite usuário alterar para um ID efetivo com mais direitos

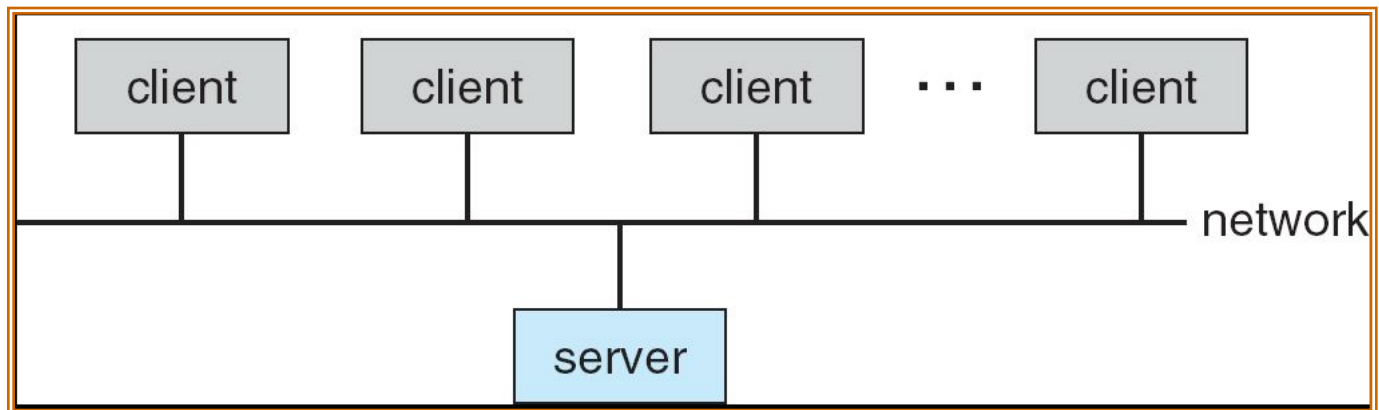
# Ambientes de Computação

---

- Computador tradicional
  - Desfocado ao longo do tempo
  - Ambiente de escritório
    - PCs conectados a uma rede, terminais anexados a mainframe ou minicomputadores provendo processamento batch ou time sharing
    - Servidores permitindo rede computadores e sistemas remotos acessarem os mesmos recursos
  - Redes em casa
    - Usada por um único sistema com modems
    - Agora usando firewall, e em rede

# Ambientes de Computação (Cont.)

- Computação Cliente-Servidor
  - Terminais burros substituídos por PCs inteligentes
  - Vários sistemas agora são servidores, respondendo as requisições geradas pelos clientes
    - Servidor provê uma interface para cliente solicitar serviços (i.e. SGBD)
    - Servidor de arquivos provê uma interface para clientes armazenar e recuperar arquivos





# Computação Peer-to-Peer (P2P)

---

- Outro modelo de sistema distribuído
- P2P não distingue clientes de servidores
  - Todos os nós são considerados peers
  - Cada peer pode agir como cliente, servidor, ou ambos
  - Um nó deve se juntar a rede P2P
    - Registra seus serviços com serviço descentralizado de descoberta sobre a rede, ou
    - Broadcast requisição por serviço e responder as requisições por serviço via protocolo de descoberta
  - Exemplos incluem *Napster*, *Gnutella*, *Emule*

# Computação baseada na Web

---

- Web tornou-se obíqua (presente em todo lugar)
- PCs são os dispositivos que prevalecem
- Mais dispositivos tornam-se acessíveis via rede permitindo acesso via web
- Nova categoria de dispositivos para gerenciar tráfego web entre servidores similares:  
**balanceadores de carga**
- Sistemas do tipo Windows 95, clientes, evoluíram para SO's do tipo Linux e Windows XP, os quais podem ser clientes e servidores

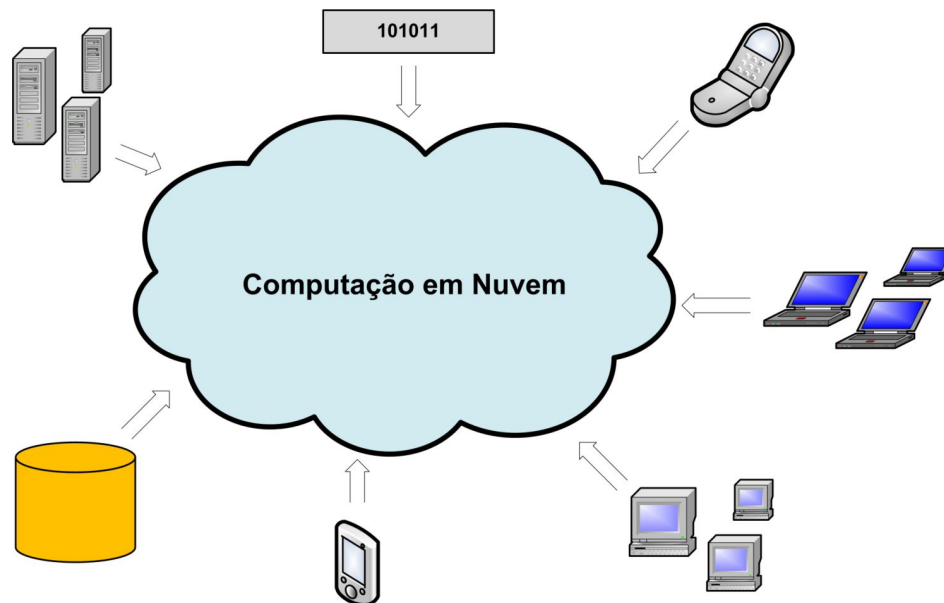
# Computação em Nuvem

---

- **Serviços básicos e essenciais** que são entregues de uma forma completamente **transparente**
  - Serviços de utilidade pública
    - Água, gás, eletricidade e telefone
- **Modelo de pagamento** baseado no uso
  - Cobrança de acordo com as diferentes políticas para o usuário final

# Computação em nuvem

- A mesma idéia de utilidade tem sido aplicada no contexto da informática
  - *Cloud Computing* ou Computação em Nuvem



# Computação em nuvem

---

- Uma tendência recente de tecnologia
  - Proporcionar serviços de TI sob demanda com pagamento baseado no uso
- Tendências anteriores à computação em nuvem foram **limitadas**:
  - A uma determinada classe de usuários
  - Focadas em tornar disponível uma demanda específica de recursos de TI, principalmente de informática

# Computação em nuvem

---

Para utilizarem os serviços, os usuários necessitam:

- Um navegador e acesso a Internet

Os recursos estão disponíveis na Internet

- As máquinas dos usuários não necessitam ter altos recursos computacionais
- Todo hardware pode ser utilizado para realizar alguma tarefa que seja adequada ao seu poder de processamento
- Novos recursos podem ser adicionados a fim de aumentar o poder de processamento e cooperar com os recursos existentes

# Computação em nuvem

---

Pretende ser global e prover serviços para as massas

- Usuário final que hospeda seus documentos pessoais na Internet
- Empresas podem terceirizar toda a parte de TI para outras empresas

Nunca uma abordagem para a utilização real foi tão global e completa

- Não apenas recursos de computação e armazenamento são entregues sob demanda
- Mas toda a pilha de computação pode ser aproveitada na nuvem

# Computação em nuvem

---

## □ Nuvem

- *É uma metáfora para a Internet ou infra-estrutura de comunicação entre os componentes arquiteturais, baseada em uma abstração que oculta a complexidade de infra-estrutura*

## □ Cada parte desta infra-estrutura é provida como um serviço

- Estes serviços são normalmente alocados em *datacenters*, utilizando hardware compartilhado para computação e armazenamento.



# Computação em nuvem



SLAs



Web Services



Virtualization

# Fim do Capítulo 1

