INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

Processamento paralelo: SISD, SIMD, MISID, MIMD, SMP e NUMA.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

Processamento paralelo

Definição:

Utilizar múltiplos (dois ou mais) processadores, simultaneamente, para resolver um mesmo problema.

Objetivo:

Aumento de desempenho (i.e. redução do tempo necessário para resolver um problema)

Motivação:

(i) Problemas cada vez mais complexos;

(ii) O Clock dos processadores estão no limite da física;

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

Classificação dos computadores paralelos

SISD (*Single Instruction Stream, Single Data Stream*):
computadores sequenciais

SIMD (*Single Instruction Stream, Multiple Data Streams*):
computadores vetoriais e matriciais

MISD (*Multiple Instruction Streams, Single Data Stream*):
não existem

MIMD (*Multiple Instruction Streams, Multiple Data Streams*):
arquiteturas com múltiplos processadores independentes

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

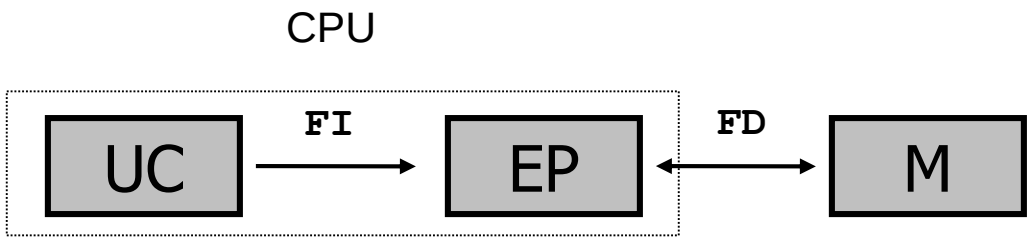
Sala: H204

AULA 15

SISD

Em arquiteturas **SISD** um único fluxo de instruções opera sobre um único fluxo de dados.

Exemplo: PCs com um único processador



INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

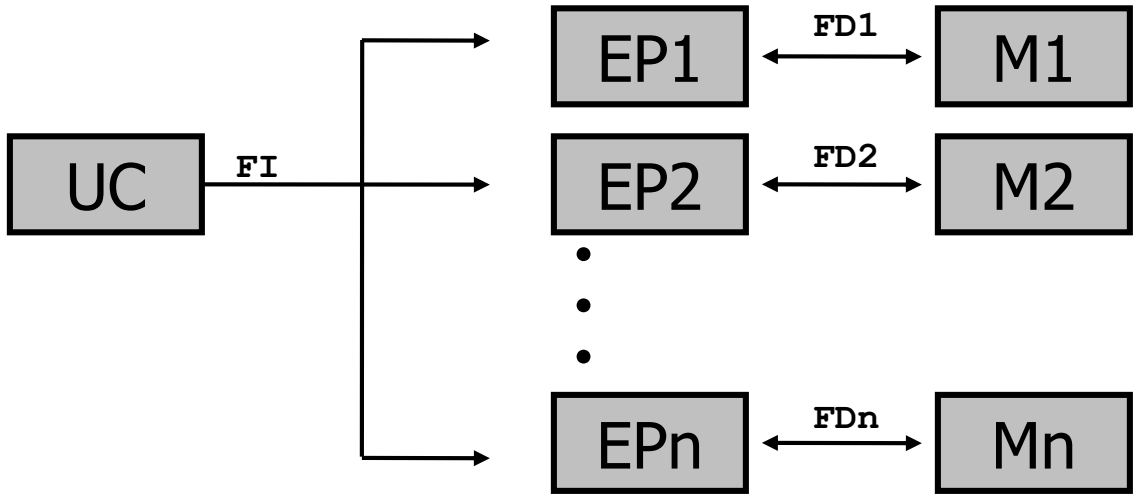
Sala: H204

AULA 15

SIMD

Em arquiteturas **SIMD** um único fluxo de instruções opera sobre múltiplos fluxos de dados. Existe uma única **UC** e múltiplos **EP**. Todos os **EP** executam de forma simultânea e sincronizada a mesma instrução sobre conjuntos de dados distintos.

Exemplo: Computadores vetoriais



INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

MISD

Em arquiteturas MISD múltiplos fluxos de instruções operam sobre um único fluxo de dados.

Exemplo: Nunca foi implementada

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

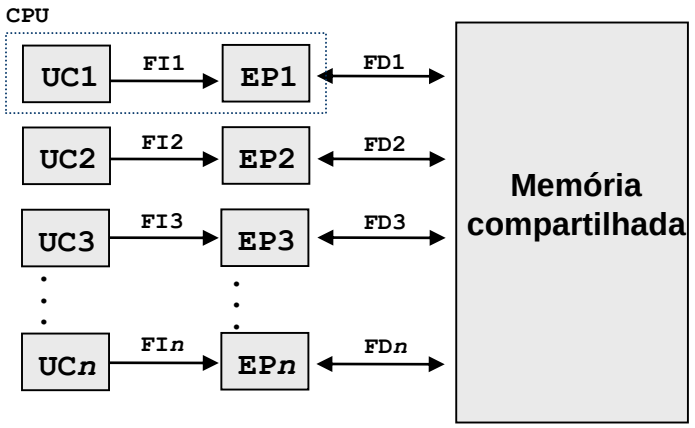
Sala: H204

AULA 15

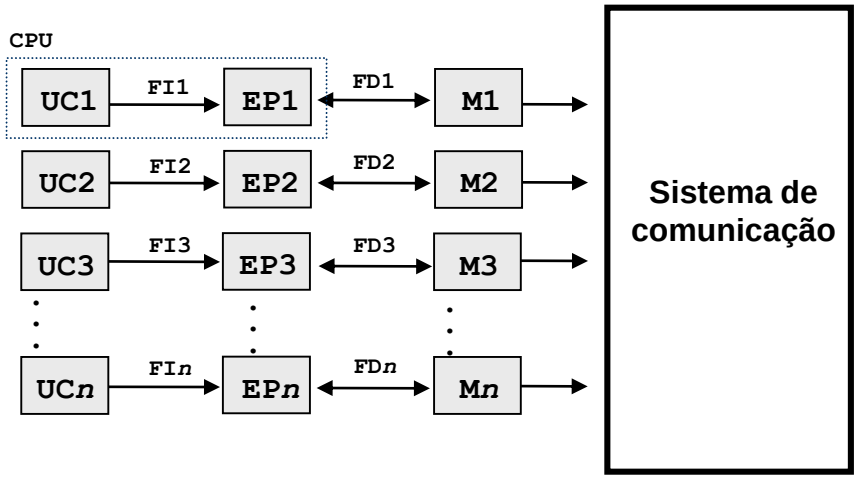
MIMD

Na arquitetura **MIMD** está a maioria das arquiteturas paralelas da atualidade. De acordo com o método de comunicação entre os processadores, a classe MIMD pode ser divida em:

MIMD compartilhada



MIMD distribuída



INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

MIMD

Três modelos de arquiteturas MIMD são bastante utilizadas:

Multiprocessador Simétrico (SMP)

arquitetura MIMD com memória compartilhada

Acesso Não-Uniforme à Memória (NUMA)

arquitetura MIMD com memória compartilhada

Agregado de Computadores (Cluster)

arquitetura MIMD com memória distribuída

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

Multiprocessadores simétricos

Consiste de múltiplos processadores similares conectados entre si a memória por um barramento ou alguma outra forma de circuito de conexão interno.

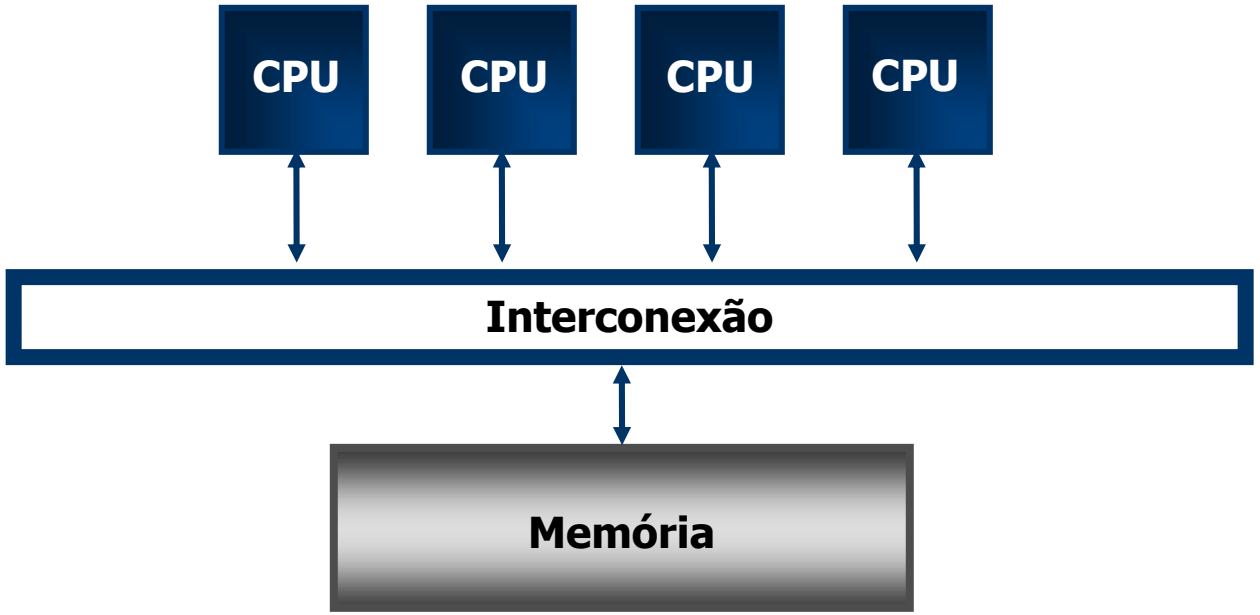


Diagrama de blocos de um SMP

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

Multiprocessadores simétricos

Características:

Possuem dois ou mais processadores similares, com capacidade de processamento próximas;

Todos os processadores compartilham uma mesma memória (i.e. existe um único espaço de endereçamento);

Todos os processadores compartilham acesso aos mesmos dispositivos de E/S, através de canais comuns, ou não;

O sistema inteiro é controlado por um único sistema operacional que torna transparente ao usuário a existência de vários processadores. Windows e Linux estão preparados para rodar em sistemas SMP.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

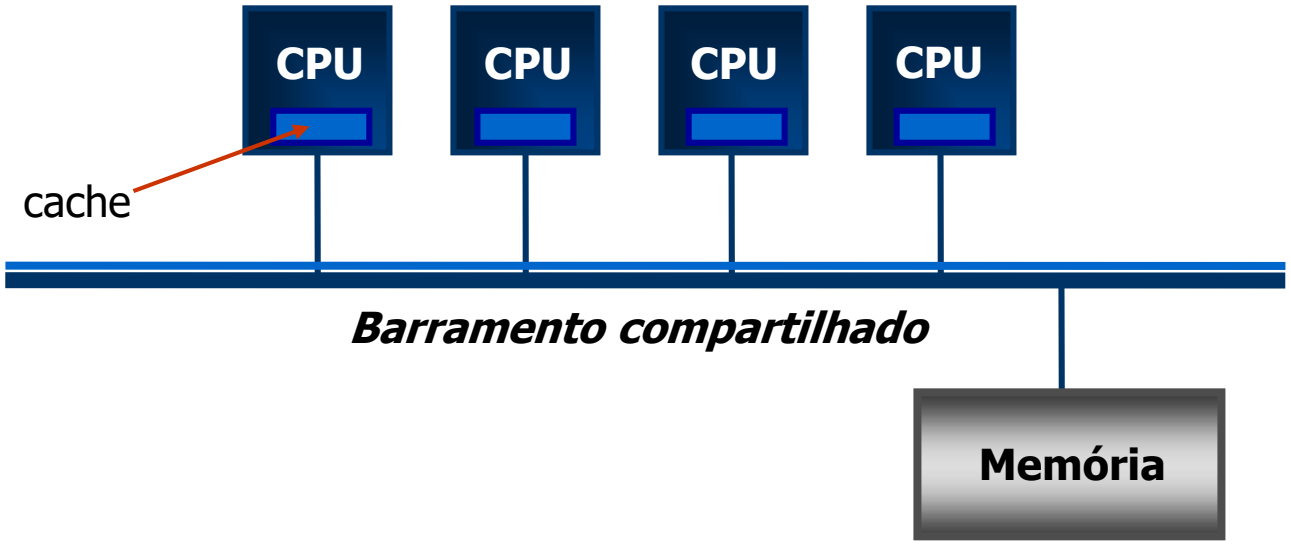
07h40-09h20

Sala: H204

AULA 15

Barramento de Tempo Compartilhado

Uso de cache melhora o tráfego no barramento e possibilita o uso de mais processadores.



SMP com barramento de tempo compartilhado e CPUs com cache

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

Coerência de Cache

Problema de coerência de cache:

Acontece porque podem existir várias cópias de um mesmo dado nas caches de diferentes processadores.

Um processador pode alterar um dado na sua cache tornando o conteúdo das memórias caches dos outros processadores incoerente/inválido;

Solução:

Implementadas em software (ex: sistema operacional e/ou compilador) e em hardware (= mais comuns).

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

Protocolo MESI

Esse protocolo determina 4 estados possíveis para um dado em cache. Para isso utiliza 2 bits adicionais por linha de cache para armazenar o estado. Os 4 estados são:

Modificado (M): O dado na cache não é igual ao dado na memória principal e encontra-se somente nesta cache.

Exclusivo (E): O dado na cache é igual ao dado na memória principal e encontra-se somente nesta cache.

Compartilhado (S): O dado na cache é igual ao dado na memória, porém encontra-se em outra(s) cache(s).

Inválido (I): O dado na cache não é um dado válido.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

NUMA

Assim como em SMPs, neste tipo de arquitetura existem dois ou mais processadores, que compartilham uma memória global (= um único espaço de endereçamento). Em um sistema NUMA os processadores são organizados em nós. Cada nó possui 1 ou mais processadores, com sua(s) própria(s) memória(s) cache (um, dois, ou mais níveis) e alguma memória principal conectados por um barramento ou outro sistema de interconexão.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

NUMA

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

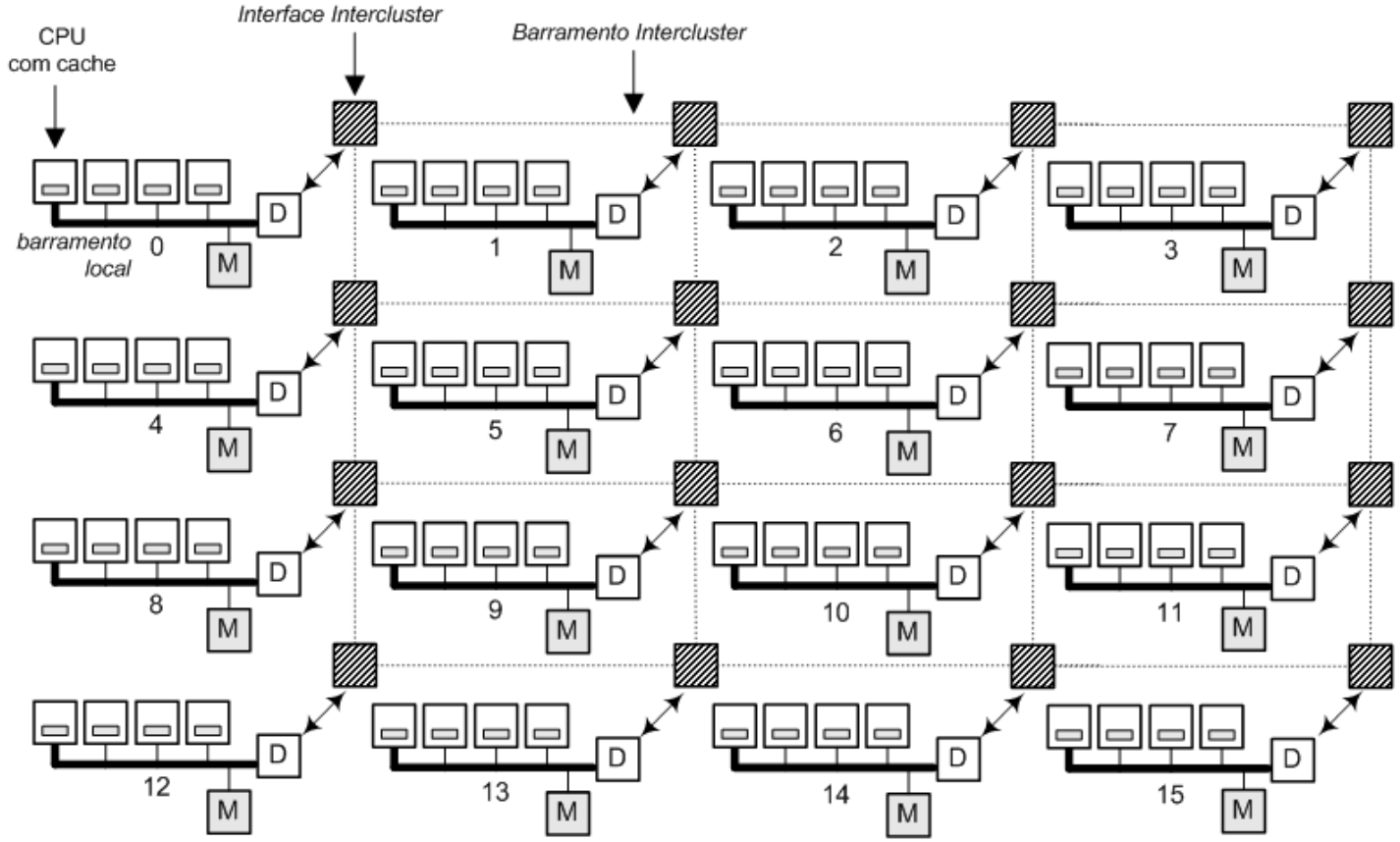


Diagrama de uma arquitetura NUMA

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

NUMA

Características:

A **principal característica** de uma arquitetura NUMA é o acesso não uniforme à memória, ou seja, embora todos os processadores possam acessar todas as posições de memória, os tempos de acesso variam de acordo com o endereço acessado.

O acesso a uma posição de memória local (memória no mesmo nó do processador que está realizando o acesso) é mais rápido do que o acesso a uma posição de uma memória remota.

Assim, o sistema operacional de uma máquina NUMA deveria, sempre que possível, escalonar as *threads* de um processo entre os processadores do nó da memória usada pelo processo.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

NUMA

Existem dois tipos de arquiteturas NUMA:

NC-NUMA

NUMA que não utiliza cache

CC-NUMA (Cache Coherent NUMA)

NUMA com cache (e coerência de cache)

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

NC-NUMA

Observações:

A coerência dos dados em máquinas NC-NUMA é garantida porque não existe *caching*;

Porém, um dado no lugar "errado" gera muitas penalidades. Se forem feitas três referências seguidas a uma posição de memória remota, são necessárias três buscas através do barramento de sistema.

Para amenizar este problema, usam-se esquemas implementados em software. Por exemplo: um daemon pode fazer estatísticas de uso do sistema. Se for verificado que uma página parece estar no lugar "errado", ele remove esta página da memória para que da próxima vez ocorra uma page fault ela possa ser alocada em um novo nó.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

CC-NUMA

Um sistema NUMA com coerência de cache é chamado CC-NUMA. A existência de *cache* requer algum protocolo de coerência. O método mais popular para a coerência de cache em sistema CC-NUMA é baseado no conceito de diretório. O diretório funciona como uma espécie de banco de dados que indica a localização das várias porções de memória, assim como o *status* das caches.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

Agregação de computadores (Clusters)

É um conjunto de computadores completos/autônomos (usualmente chamados nós) interconectados, que podem trabalhar juntos, como um recurso de computação unificado, criando a ilusão de uma máquina única.

Os nós são geralmente conectados através de uma porta de E/S (geralmente interfaces de rede) de alto desempenho.

Atualmente eles são utilizados com sistemas gerenciadores de bancos de dados, com servidores WEB e, principalmente, para proc. paralelo.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

Agregação de computadores (Clusters)

Características:

Configuração dos nós: Os nós podem ser PCs ou mesmo SMPs (neste caso ele é chamado um cluster de SMPs);

Escalabilidade absoluta: É possível construir clusters muito grandes, cuja capacidade de computação ultrapassa várias vezes a capacidade da maior máquina individual;

Escalabilidade incremental: Um cluster pode ser configurado de maneira que seja possível adicionar novos nós, expandindo-o de forma incremental;

Disponibilidade: Como cada nó de um cluster é um computador independente, uma falha em um nó não implica necessariamente na perda total de serviço.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15

Diferença entre Grid e Cluster

Ao contrário dos Grids, que geralmente interligam grandes distâncias, países, universidades, empresas e organizações, os clusters são mais locais.

Os membros de um cluster são chamados de nós e geralmente ficam em um datacenter (ambiente apropriado), prédio ou sala.

A administração dos recursos, tanto de hardware como de processamento, também é de ordem local (geralmente de uma empresa), diferente do Grid, onde cada um contribui e administra 'um pouco'.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

Diferença Grid e Computação distribuída

ECM 245

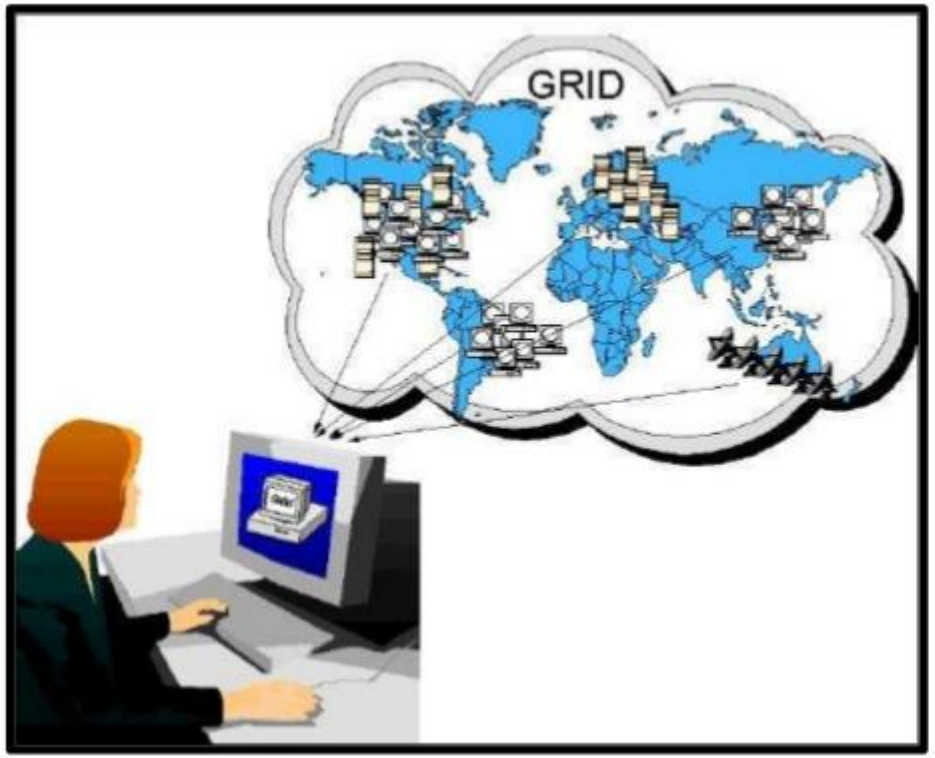
Arquitetura e Organização de Computadores

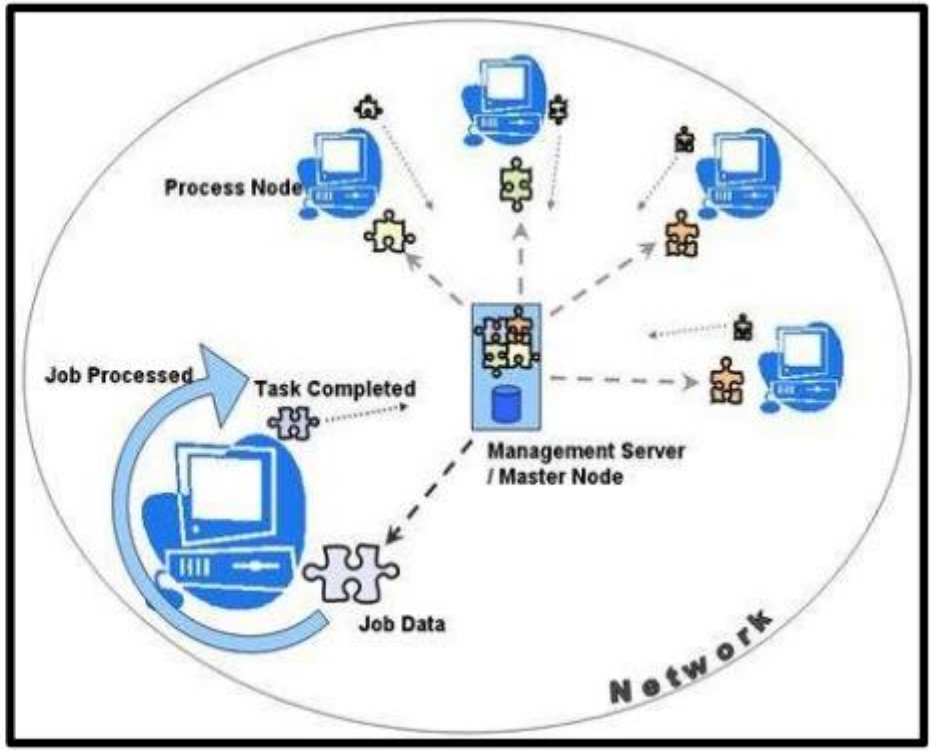
5ª-feira

07h40-09h20

Sala: H204

AULA 15





INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

ECM 245

Arquitetura e Organização de Computadores

5ª-feira

07h40-09h20

Sala: H204

AULA 15