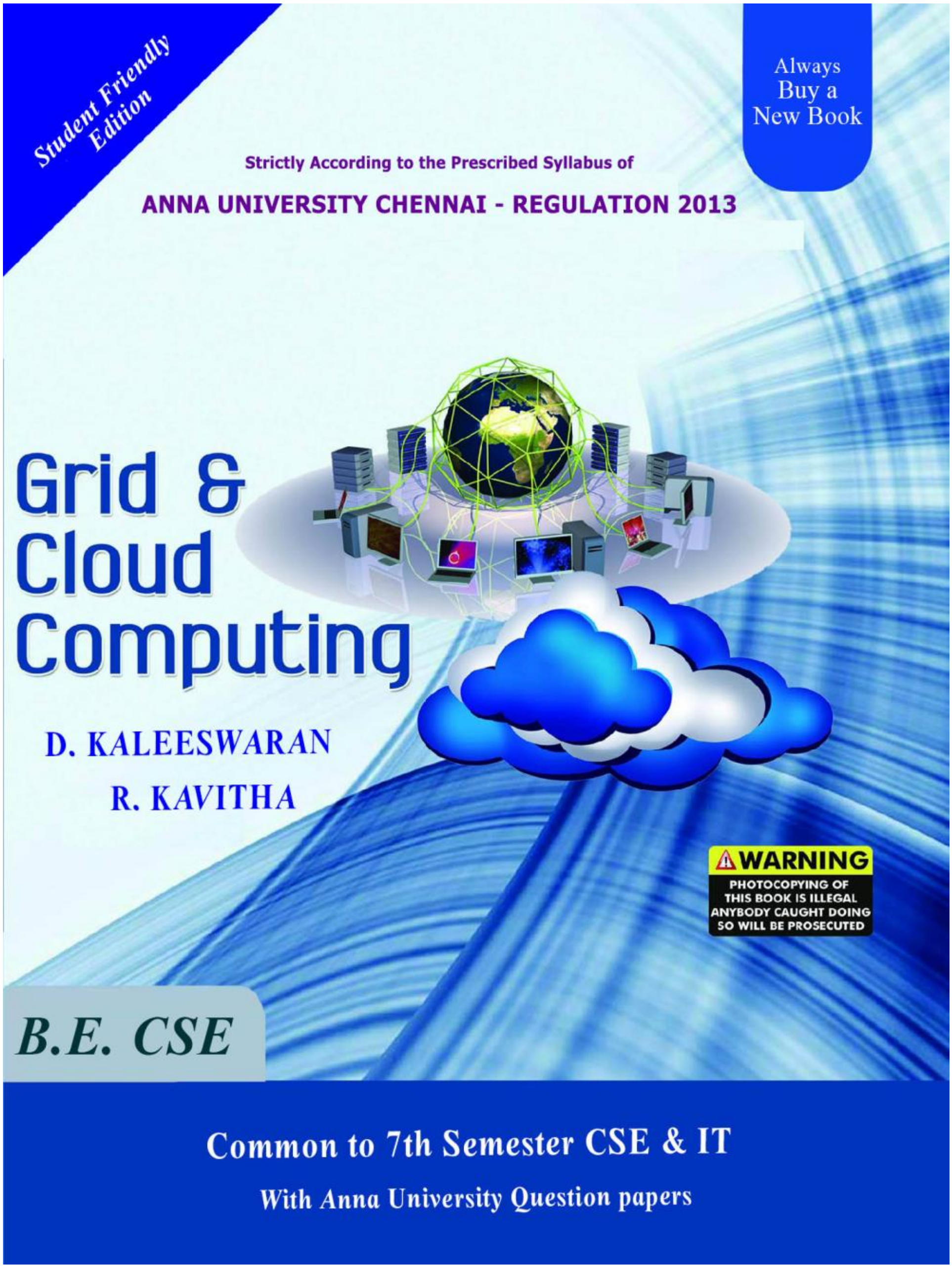
**We are providing 25% of Discount for Students to purchase the book.**

**Contact: eeswarpublicationmadu@gmail.com**



# GRID AND CLOUD COMPUTING

## (For Common to B.E CSE & IT Branches)

---

**AS PER THE LATEST SYLLABUS OF ANNA UNIVERSITY, CHENNAI**

---

**(Regulation 2013)**


**Mr. D.KALEESWARAN**
**Assistant Professor**
**Dept. of Computer Science and Engineering**
**St. Michael College of Engineering & Technology**
**Sivagangai, Tamil Nadu**

**Dr. R. KAVITHA**
**Professor**
**Dept. of Computer Science and Engineering**
**Velammal College of Engineering & Technology**
**Madurai, Tamil Nadu.**

# GRID AND CLOUD COMPUTING

**By,**

**Mr. D. Kaleeswaran M.E., (Ph.D)**

**&**

**Dr. R. Kavitha, M.E., Ph.D.**

**First Edition : April 2016**

**Copy Right © Exclusive by the publisher with Author**

**All Right Reserved**

**No part of the publication to reproduced stored in a retrieval system or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without the prior written permission of the author.**

**Price Rs:- 280/-**

**Published by Copies can be had form:**

**Publication Name:**

**Eeswar Publishers Distributors Pvt Ltd**
**Mail id: eeswarpublicationmdu@gmail.com**

# SYLLABUS

## ANNA UNIVERSITY CHENNAI – REGULATION 2013

# GRID AND CLOUD COMPUTING

**UNIT I INTRODUCTION**

Evolution of Distributed computing: Scalable computing over the Internet – Technologies for network based systems – clusters of cooperative computers - Grid computing Infrastructures – cloud computing - service oriented architecture – Introduction to Grid Architecture and standards – Elements of Grid – Overview of Grid Architecture.

**UNIT II GRID SERVICES**

Introduction to Open Grid Services Architecture (OGSA) – Motivation – Functionality Requirements – Practical & Detailed view of OGSA/OGSI – Data intensive grid service models – OGSA services.

**UNIT III VIRTUALIZATION**

Cloud deployment models: public, private, hybrid, community – Categories of cloud computing: Everything as a service: Infrastructure, platform, software - Pros and Cons of cloud computing – Implementation levels of virtualization – virtualization structure – virtualization of CPU, Memory and I/O devices – virtual clusters and Resource Management – Virtualization for data center automation. **UNIT IV PROGRAMMING MODEL**

Open source grid middleware packages – Globus Toolkit (GT4) Architecture , Configuration – Usage of Globus – Main components and Programming model - Introduction to Hadoop Framework - Mapreduce, Input splitting, map and reduce functions, specifying input and output parameters, configuring and running a job – Design of Hadoop file system, HDFS concepts, command line and java interface, dataflow of File read & File write.

**UNIT V SECURITY**

Trust models for Grid security environment – Authentication and Authorization methods – Grid security infrastructure – Cloud Infrastructure security: network, host and application level – aspects of data security, provider data and its security, Identity and access management architecture, IAM practices in the cloud, SaaS, PaaS, IaaS availability in the cloud, Key privacy issues in the cloud.

# CONTENTS

# UNIT – I INTRODUCTION

## 1.1 EVOLUTION OF DISTRIBUTED COMPUTING

- A distributed computing system consists of multiple autonomous computers that communicate through a computer network.
- It accomplishes a task in a distributed way to achieve the results faster than computing with a single computer.

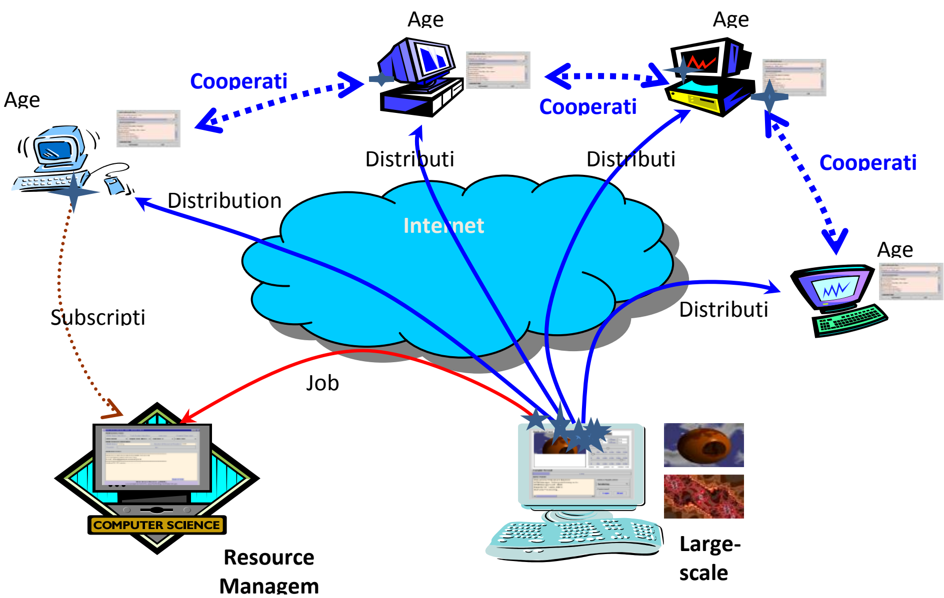



*Figure 1.1 Example of Distributed System*

- **Need for Distributed Computing**

- To perform Computing intensive tasks in a shorter time span as the tasks could consume a lot of time on computing.
  - Eg: Astronomical Calculations, Large scale molecular simulations etc..,
  - To perform Data intensive tasks that deals with a large amount or large size of Data files.
  - Eg: Facebook Data, LHC(Large Hadron Collider) data at CERN

Robust performance of process execution

- No SPOF (Single Point Of Failure

- **Properties**
  - Fault tolerance
  - Resource sharing
  - Load Sharing
  - Dispatching several tasks to each nodes can help share loading to the whole system.
- **Network based systems**
  - The distributed computing needs a networked system for process management. There are three types of network systems;
  - LAN – Local Area Network – Connecting local servers
  - NAS/SAN – Network Assisted Storage/Storage Area Network
  - VMWare – Virtual Machines
  - The evolution internet into a low-cost network medium has enabled distribution of process over remote locations by means of 'Cloud Computing'.
- **NAS Vs SAN**

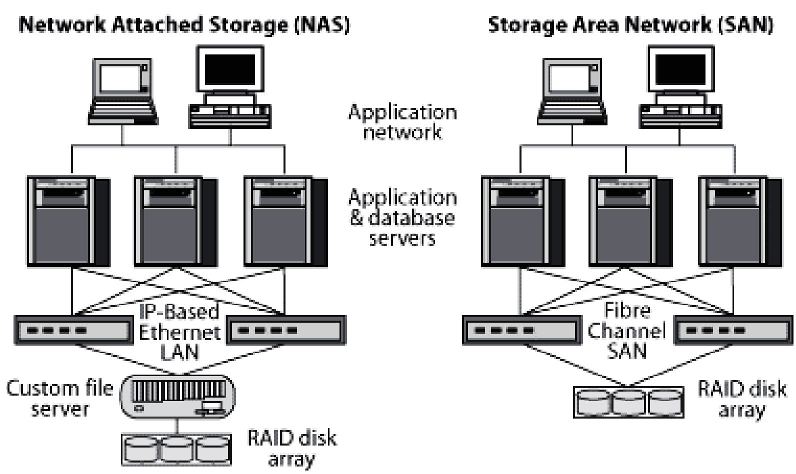


***Figure 1.2 NAS Vs SAN***

- A NAS is fundamentally a bunch of disks, usually arranged in a disk array.
- Users and servers attach to the NAS primarily using TCP/IP over Ethernet.
- SANs are used for mission-critical data such as big databases, where reliability and performance are key.

## 1.2 SCALABLE COMPUTING OVER THE INTERNET

- It is a evolutionary changes in machine architecture, operating system platform, network connectivity, and application workload.
- Distributed computing becomes data-intensive and network-centric.
- It identifies the applications of modern computer systems that practice parallel and distributed computing.

1. The Age of Internet Computing
2. Scalable Computing Trends and New Paradigms
3. The Internet of Things and Cyber-Physical Systems

### 1.2.1 The Age of Internet Computing

- The emergence of computing clouds instead demands high-throughput computing (HTC) systems built with parallel and distributed computing technologies
- The Platform Evolution:
  - The general computing trend is to leverage shared web resources and massive amounts of data over the Internet.
  - Figure 1.1 illustrates the evolution of HPC and HTC systems. On the HPC side, supercomputers massively parallel processors or MPPs) are gradually replaced

by clusters of cooperative computers out of a desire to share computing resources.

- High-performance computing (HPC) is used to applications is no longer optimal for measuring system performance.
- High-throughput computing (HTC) is used improve fast servers, storage systems, and high-bandwidth networks.

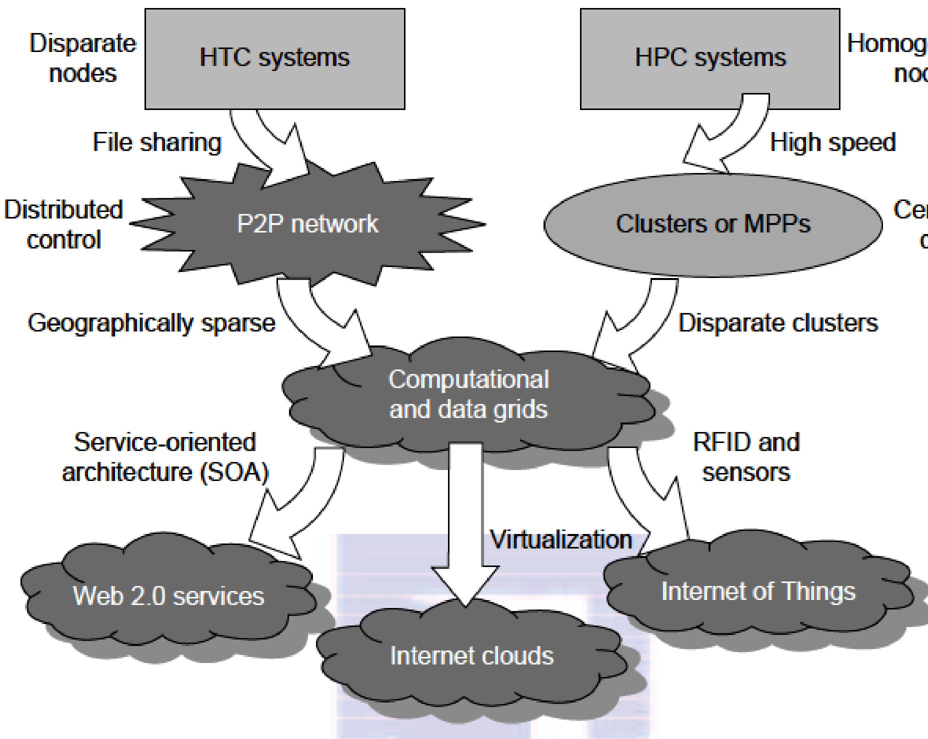



*Figure 1.2.1 Evolution of HPC and HTC systems*

**Three New Computing Paradigms**

- Advances in virtualization make it possible to see the growth of Internet clouds as a new computing paradigm
    1. Radio-frequency identification (RFID)
    2. Global Positioning System (GPS)
    3. Sensor technologies
- Computing Paradigm Distinctions
    - The high-technology community has argued for many years about the precise definitions of centralized computing, parallel computing, distributed computing, and cloud computing.

1. **Centralized computing:**
    - All computer resources are centralized in one physical system.
    - All resources (processors, memory, and storage) are fully shared and tightly coupled within one integrated OS.
    - Many data centers and supercomputers are centralized systems, but they are used in parallel, distributed, and cloud computing applications
2. **Parallel computing:**
    - All processors are either tightly coupled with central shard memory or loosely coupled with distributed memory.
    - Parallel computing is a type of computation in which many calculations are carried out simultaneously, operating on the principle that large problems can often be divided into smaller ones, which are then solved at the same time.

### 3. Distributed computing:

- A distributed system consists of multiple autonomous computers, each having its own private memory, communicating through a computer network.
- Information exchange in a distributed system is accomplished through message passing.

### Cloud computing:

- An Internet cloud of resources that may be either centralized or decentralized. The cloud apples to parallel or distributed computing or both. Clouds may be built from physical or virtualized resources.
- Cloud computing, often referred to as simply "the cloud," is the delivery of on-demand computing resources - everything from applications to data centers - over the Internet on a pay-for-use basis.

### Distributed System Families

- To establish wide area computing infrastructures, known as computational grids or data grids.
- Grids and clouds are disparity systems that place great emphasis on resource sharing in hardware, software, and data sets.
- Demand in computing power in terms of
    1. Throughput
    2. Efficiency
    3. Scalability
    4. Reliability.
- These factors are defined bellow.

- **Efficiency:**
  - It measures the utilization rate of resources in an execution model by exploiting massive parallelism in HPC
- **Dependability:**
  - It measures the reliability and self-management from the chip to the system and application levels.
- **Adaptation in the programming model:**
  - It measures the ability to support billions of job requests over massive data sets and virtualized cloud resources under various workload and service models.
- **Flexibility in application deployment:**
  - It measures the ability of distributed systems to run well in both HPC (science and engineering) and HTC (business applications)

## 1.3 TECHNOLOGIES FOR NETWORK BASED SYSTEMS

- It explores hardware, software, and network technologies for distributed computing system design and applications.
- Approaches to building distributed operating systems for handling massive parallelism in a distributed environment.
  1. Multicore CPUs and Multithreading Technologies
  2. GPU Computing to Exascale and Beyond
  3. Memory, Storage, and Wide-Area Networking
  4. Virtual Machines and Virtualization Middleware
  5. Data Center Virtualization for Cloud Computing

### 1.3.1 Multicore CPUs and Multithreading Technologies:

- The growth of component and network technologies is HPC and HTC systems.

- Processor speed is measured in millions of instructions per second (MIPS) and network bandwidth is measured in megabits per second (Mbps) or gigabits per second (Gbps).

**Advances in CPU Processors:**

- Advanced CPUs or microprocessor chips assume a multicore architecture with dual, quad, six, or more processing cores.
- These processors exploit parallelism at ILP and TLP levels.
- ILP mechanisms include,
    1. Multiple-issue superscalar architecture
    2. Dynamic branch prediction
    3. Speculative execution, among others.
- DLP and TLP are highly explored in graphics processing units (GPUs) that adopt many-core architecture with hundreds to thousands of simple cores.

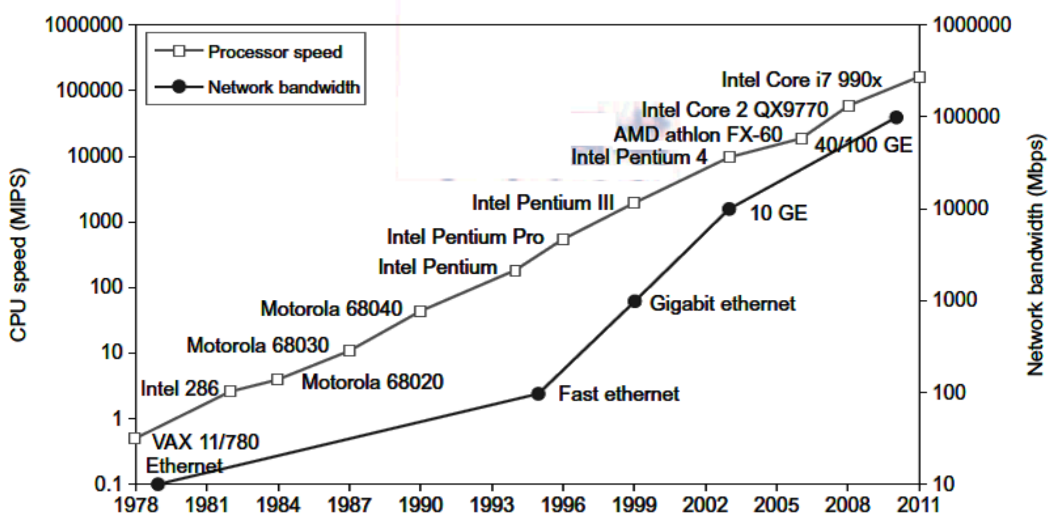



*Figure 1.3.1 generations of microprocessors or CMPs*

- Each core is essentially a processor with its own private cache (L1 cache). Multiple cores are housed in the same chip with an L2 cache that is shared by all cores.

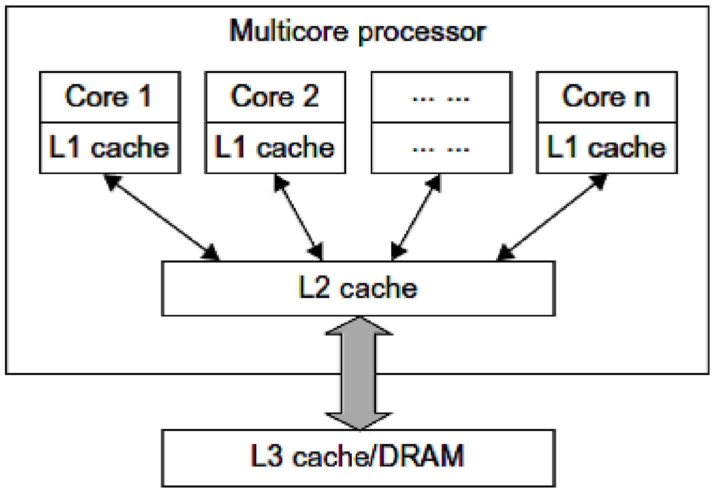



*Figure 1.3.2 Architecture of a typical multicore processor*

- Multicore and multi-threaded CPUs are equipped with many high-end processors, including the Intel i7, Xeon, AMD Opteron, Sun Niagara, IBM Power 6, and X cell processors.

**Multicore CPU and Many-Core GPU Architectures:**

- Multicore CPUs may increase from the tens of cores to hundreds or more in the future. But the CPU has reached its limit in terms of exploiting massive DLP due to the aforementioned memory wall problem.
- In the future, the processor industry is also keen to develop asymmetric or heterogeneous chip mul-tiprocessors that can house both fat CPU cores and thin GPU cores on the same chip.

**Multithreading Technology:**

- Each of the three multithreaded processors is four-way multithreaded over four functional data paths. In the dual-core processor, assume two processing cores, each a single-threaded two-way superscalar processor.

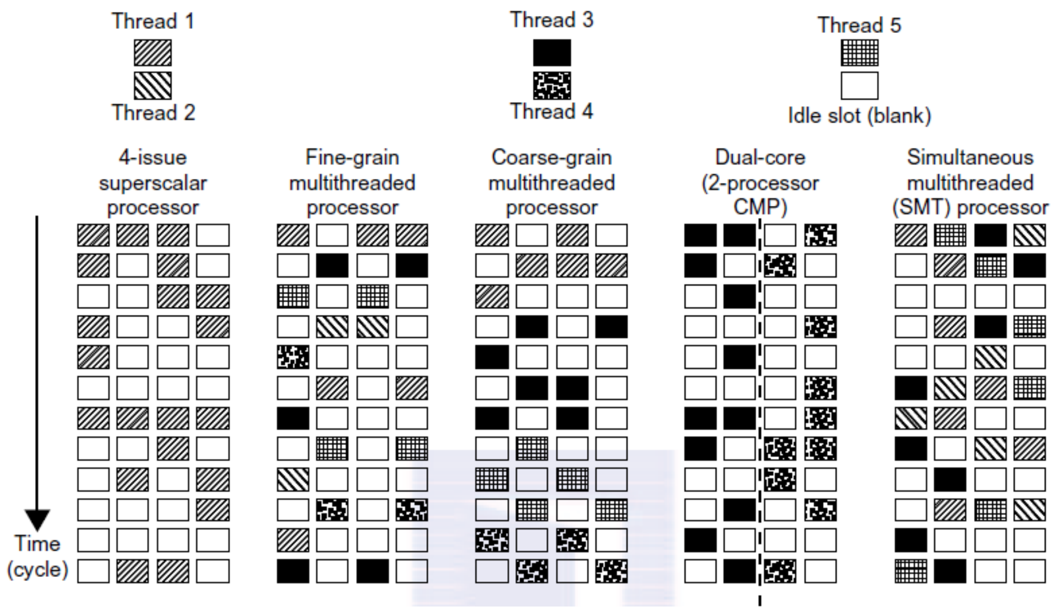



*Figure 1.3.. five micro-architectures in modern CPU processors*

- Instructions from different threads are distinguished by specific shading patterns for instructions from five independent threads.

  1. Only instructions from the same thread are executed in a superscalar processor.
  2. Fine-grain multithreading switches the execution of instructions from different threads per cycle.
  3. Course-grain multi-threading executes many instructions from the same thread for quite a few cycles before switching to another thread.

4. The multicore CMP executes instructions from different threads completely.
5. The SMT allows simultaneous scheduling of instructions from different threads in the same cycle.

### 1.3.2 GPU Computing to Exascale and Beyond

- A GPU is a graphics coprocessor or accelerator mounted on a computer's graphics card or video card. A GPU offloads the CPU from tedious graphics tasks in video editing applications.
- General-purpose computing on GPUs, known as GPGPUs, have appeared in the HPC field.
- Parallel GPUs or GPU clusters have been garnering a lot of attention against the use of CPUs with limited parallelism.

**How GPUs Work:**

- GPUs are designed to handle large numbers of floating-point operations in parallel.
- Modern GPUs are not restricted to accelerated graphics or video coding.
- Conventional GPUs are widely used in mobile phones, game consoles, embedded systems, PCs, and servers.
- The NVIDIA CUDA Tesla or Fermi is used in GPU clusters or in HPC systems for parallel processing of massive floating-pointing data.

**GPU Programming Model**

- The GPU has a many-core architecture that has hundreds of simple processing cores organized as multiprocessors.
- Each core can have one or more threads. Essentially, the CPU's floating-point kernel computation role is largely offloaded to the many-core GPU. The CPU instructs the GPU to perform massive data processing.

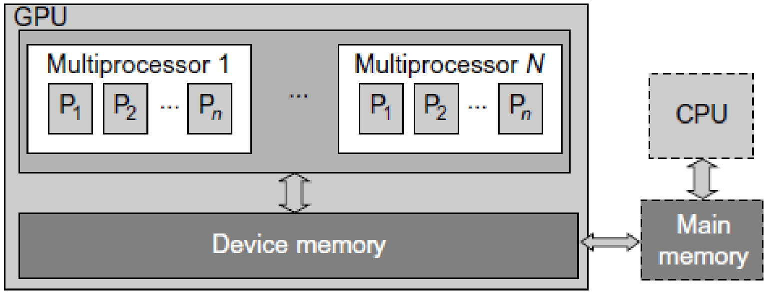
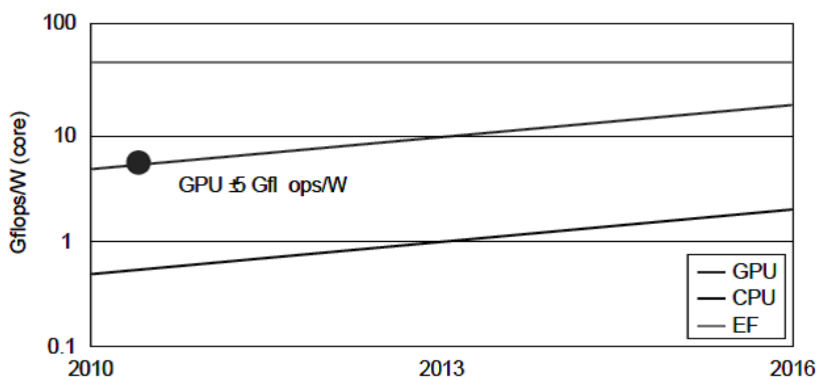



*Figure 1. 3.3Interaction between a CPU and GPU in performing parallel execution*

**Power Efficiency of the GPU**

- It estimated that the CPU chip consumes about nJ/instruction.
- The CPU is optimized for latency in caches and memory, while the GPU is optimized for throughput with explicit management of on-chip memory.




- The GPU performance (middle line, measured 5 Gflops/W/core in 2011), compared with the lower CPU performance (lower line measured 0.8 Gflops/W/core in 2011) and the estimated 60 Gflops/W/core performance in 2011 for the Exascale (EF in upper curve) in the future.

## 1.3.3 Memory, Storage, and Wide-Area Networking

**Memory Technology**

- The growth of DRAM chip capacity from 16 KB in to 64 GB.

- The capacity increase of disk arrays will be even greater in the years to come. Faster processor speed and larger memory capacity result in a wider gap between processors and memory.

**Disks and Storage Technology**

- The rapid growth of flash memory and solid-state drives (SSDs) also impacts the future of HPC and HTC systems.
- Eventually, power consumption, cooling, and packaging will limit large system development. Power increases linearly with respect to clock frequency and quadratic ally with respect to voltage applied on chips.

**System-Area Interconnects**

- The nodes in small clusters are mostly interconnected by an Ethernet switch or a local area network (LAN).
- A storage area network (SAN) connects servers to network storage such as disk arrays.
- Network attached storage (NAS) connects client hosts directly to the disk arrays.

**Wide-Area Networking**

- Network links with 1,000, 1,000, 100, 10, and 1 Gbps bandwidths were reported, respectively, for international, national, organization, optical desktop, and copper desktop connections.

### 1.3.4 Virtual Machines and Virtualization Middleware

- Virtual machines (VMs) offer novel solutions to underutilized resources, application inflexibility, software manageability, and security concerns in existing physical machines.
- Today, to build large clusters, grids, and clouds, we need to access large amounts of computing, storage, and networking resources in a virtualized manner.

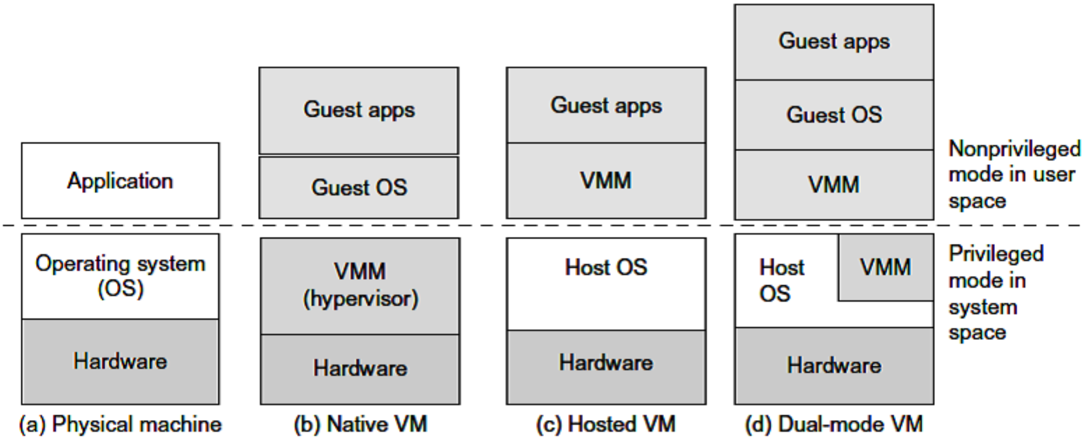



***Figure 1.3.4 The architectures of three VM configurations***

## Virtual Machines

- The VM is built with virtual resources managed by a guest OS to run a specific application.
- Between the VMs and the host platform, one needs to deploy a middleware layer called a virtual machine monitor (VMM).
- A native VM installed with the use of a VMM called a hypervisor in privileged mode.

## VM Primitive Operations

- The VMM provides the VM abstraction to the guest OS.
- With full virtualization, the VMM exports a VM abstraction identical to the physical machine so that a standard OS.
  - a) VMs can be multiplexed between hardware machines.
  - b) Second, a VM can be suspended and stored in stable storage.
  - c) Third, a suspended VM can be resumed or provisioned to a new hardware platform.
  - d) Finally, a VM can be migrated from one hardware platform to another.

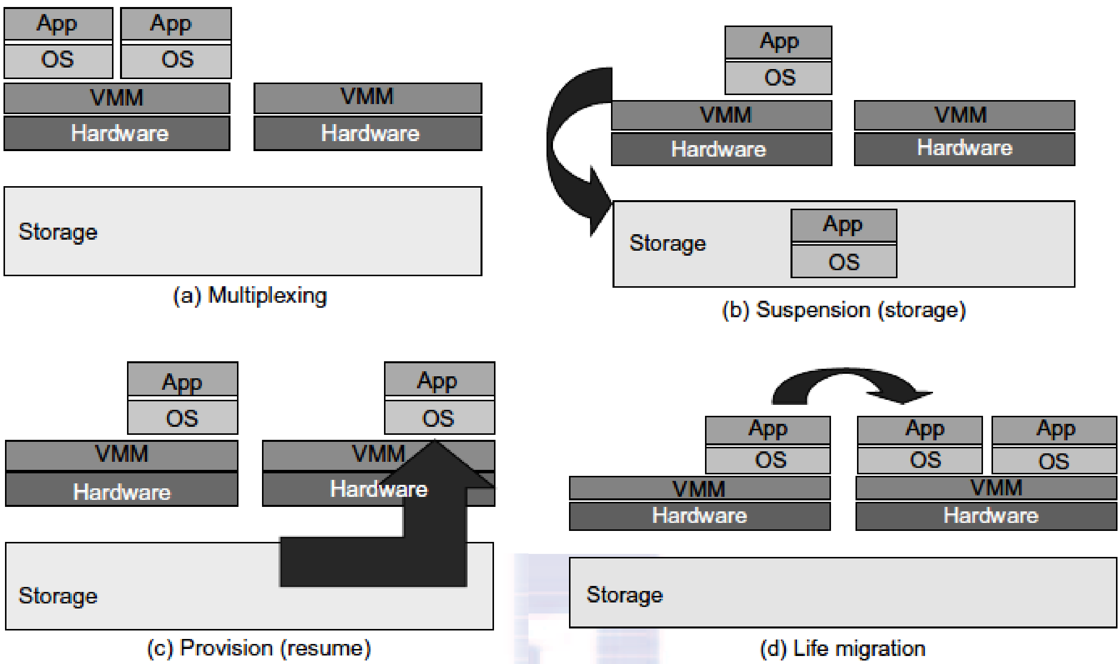



***Figure 1.3.5 VM multiplexing, suspension, provision, and migration in a distributed computing environment***

**Virtual Infrastructures**

- Physical resources for compute, storage, and networking are mapped to the needy applications embedded in various VMs at the top.
- Virtual infrastructure is what connects resources to distributed applications. The result is decreased costs and increased efficiency and responsiveness.

## 1.3.5 Data Center Virtualization for Cloud Computing

- Cloud architecture is built with commodity hardware and network devices. Data center design emphasizes the performance/price ratio over speed performance alone.

**Data Center Growth and Cost Breakdown**

- A large data center may be built with thousands of servers. Smaller data centers are typically built with hundreds of servers. 60 percent of the cost to run a data center is allocated to management and maintenance.

**Low-Cost Design Philosophy**

- High-end switches or routers may be too cost-prohibitive for building data centers. The software layer handles network traffic balancing, fault tolerance, and expandability.

**Convergence of Technologies**

- Cloud computing is enabled by the convergence of technologies in four areas:

  (1) hard-ware virtualization and multi-core chips.

  (2) utility and grid computing.

  (3) SOA, Web 2.0, and WS mashups.

  (4) Atonomic computing and data center automation.