CLOUD COMPUTING

Dr. Manjunath Hegde
Dept. of Computer Applications
Manipal Institute of Technology, Manipal

SESSION OUTLINE

- Introduction to Cloud Computing
- Scalable Computing over the Internet,
- Technologies for Network based systems,
- System models for Distributed and Cloud Computing,
- Software environments for distributed systems and clouds

CLOUD COMPUTING

- A type of internet-based computation model that provides shared computer processing resources and data to computers and other devices on demand.
- Model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources (e.g., computer networks, servers, storage, applications and services), which can be rapidly provisioned and released with minimal management effort.
- Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in either privately owned, or third-party data centers that may be located far from the user–ranging in distance from across a city to across the world.

DIVERSIFIED COMPUTATION

- Personal Computing
- Mobile Computing
- Distributed Computing
  - Grid Computing
  - Cluster Computing
- Parallel Computing
- Super Computing
- Utility Computing
- Ubiquitous Computing (Ubi-comp)

PERSONAL COMPUTING

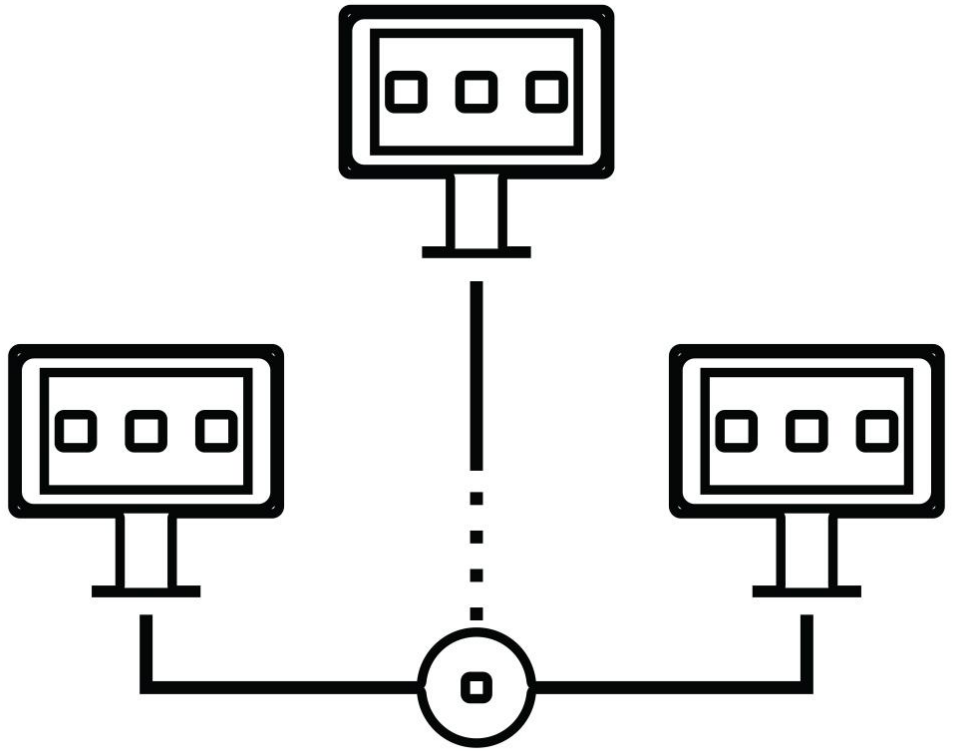
- Personal Computing System
- Local S/W installation, maintenance
- Local System maintenance
- Customizable to user needs
- High up-front cost
- Very low utilization

MOBILE COMPUTING

- Computing device used even when being mobile
- Local installations, maintenance
- Customizable to user expectations
- Low utilization compared to High up-front cost
- Range-bandwidth, health hazard, etc.

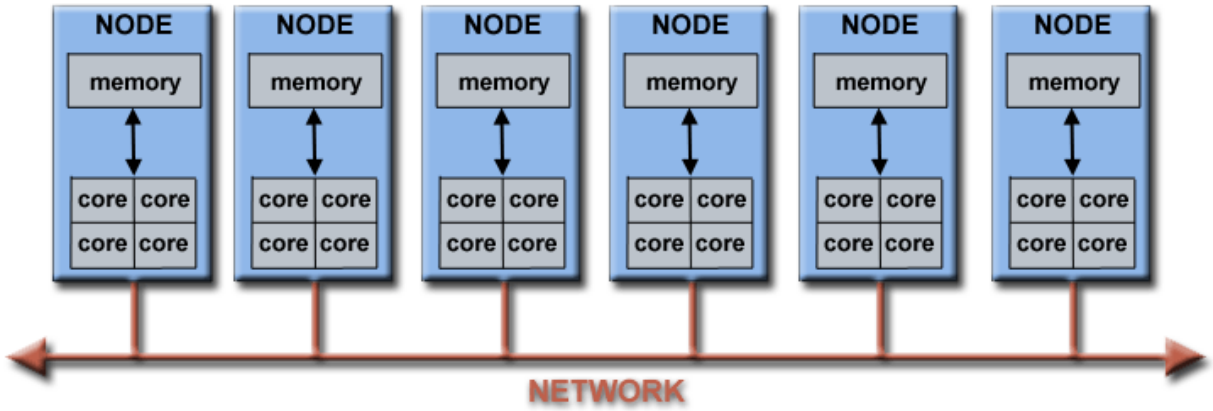
DISTRIBUTED COMPUTING

- A problem is divided into many tasks, each of which is solved by one or more computers
- Computing components located on networked computers
- Communicate and coordinate their actions by passing messages, to achieve a common goal



PARALLEL COMPUTING

- Large problems can often be divided into smaller ones, which are solved concurrently
- High performance Computing
- Multicore architecture



- Communication and Synchronization between the different subtasks !!!

SUPER COMPUTING

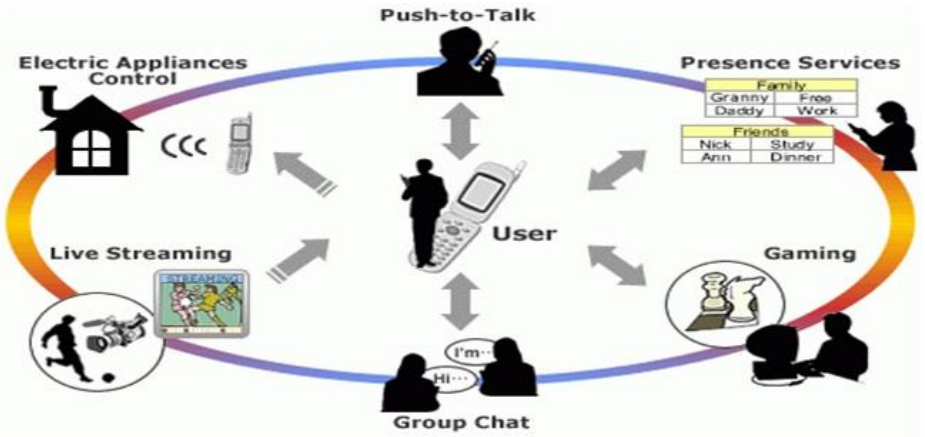
- Distributed approach: individual clients receive, complete small tasks and a central server integrates the task-results into an overall solution.
- Cluster approach: dedicated processors in close proximity work together
- Tremendous power consumption, huge up-front cost.
- Very complicated maintenance issues. . . . .

UTILITY COMPUTING

- Packaging of computing resources, such as computation, storage and services, as a metered service.
- Low or no initial cost to acquire computer resources; instead, computational resources are essentially rented.
- ON-DEMAND computing
- Pay-per-Use

UBIQUITOUS COMPUTING

- Computing is made to appear everywhere in everything
- Internet Services, advanced network and middleware
- Stock, weather, news reports
- Home security.
- Domestic, corporate, scientific, research, military, etc.



CLOUD COMPUTING

- Avail Ready-made FACILITY
- Applications, Services and Infrastructure:
- Remotely hosted: Services or data are hosted on remote infrastructure.
- Pay for use, as per need:
- scale up and down in capacity and functionalities
- H/w and S/w services are available to
- general public, enterprises, corporations and business markets
- Ubiquitous: Services or data are available everywhere.
- Commoditised: Result is a utility computing model, like gas and electricity - you pay for what you would want!

TERMS IN CLOUD COMPUTING

- Resource sharing
- Virtualization
- Multi-tenancy
- Service model: IaaS, PaaS, SaaS
- Deployment model: Public, Private, Community, Hybrid
- Cloud Computing Architecture (Big data, Hadoop, Map-reduce)
- Privacy and Security

SOME CLOUD SERVICES

Personal Services

- Email
- Social Networking
- Data Storage – Store mp3, video, photo and documents online instead of at personal system.
- Data Sharing – Google Doc's, allows quicker updates and faster project completion times
- Processing Power - Amazon's EC2 virtual computing environment

Enterprise Services

- Enterprise resource planning (ERP)

SCALABLE COMPUTING OVER THE INTERNET

SCALABLE COMPUTING OVER THE INTERNET

- Computing technology has undergone many of platform and environment changes from past several years.
- Evolutionary changes in machine architecture, operating system platform, network connectivity, and application workload.
- Instead of using a centralized computer to solve computational problems, a parallel and distributed computing system uses multiple computers to solve large-scale problems over the Internet.
- Thus, distributed computing becomes data-intensive and network-centric.
- These large-scale Internet applications have significantly enhanced the quality of life and information services in society today.

THE AGE OF INTERNET COMPUTING

- Billions of people use the Internet every day. As a result, supercomputer and large data centers should provide high-performance computing services.
- This high demand made to think out of the box. The results are:
- High-performance computing (HPC):
  - High Performance Computing most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than desktop computer or workstation in order to solve large/complex problems.
  - In HPC systems, problems will be split up into many smaller "problems" called threads, corresponding to each core. The cores can communicate with each other efficiently, and the system will be organized well.
- High-throughput computing (HTC) :
  - High-throughput computing (HTC) is a computer science term to describe the use of many computing resources over long periods of time to accomplish a computational task.
  - HTC gives more attention to high-flux computing. The main application for high-flux computing is in Internet searches and web services by millions or more users simultaneously.

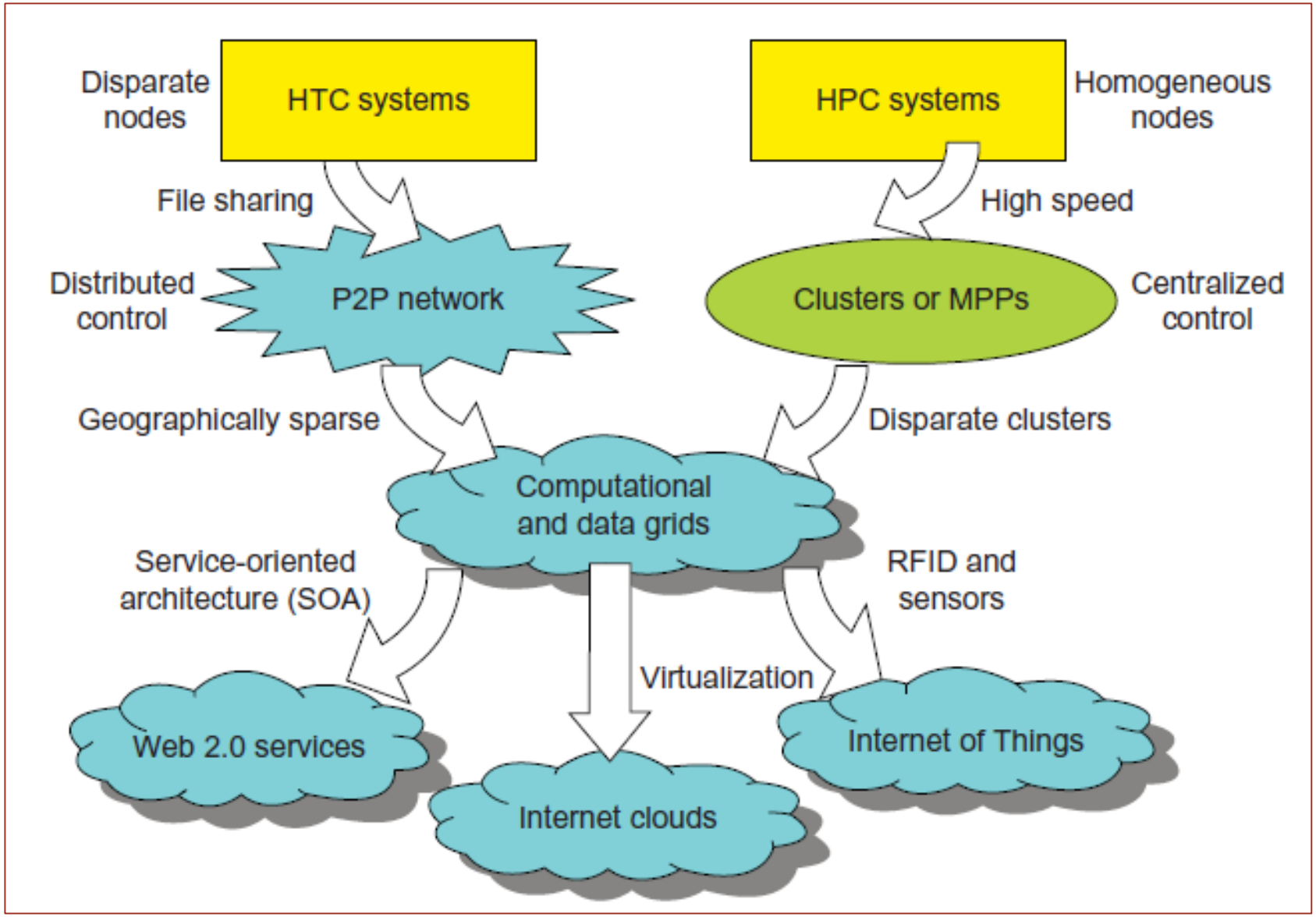


Fig 1. Differences in HPC and HTC

NEW COMPUTING PARADIGMS

- Advances in virtualization make it possible to see the growth of Internet clouds as a new computing paradigm.
- Radio-frequency identification (RFID)
- Global Positioning System (GPS)
- Internet of Things (IoT)

COMPUTING PARADIGM DISTINCTIONS

- Centralized computing - This is a computing paradigm by which all computer resources are centralized in one physical system.
- Parallel computing - In parallel computing, all processors are either tightly coupled with centralized shared memory or loosely coupled with distributed memory.
- Distributed computing - has multiple autonomous computers, each having its own private memory, communicating through a computer network.
- Cloud computing - An Internet cloud of resources can be either a centralized or a distributed computing system.

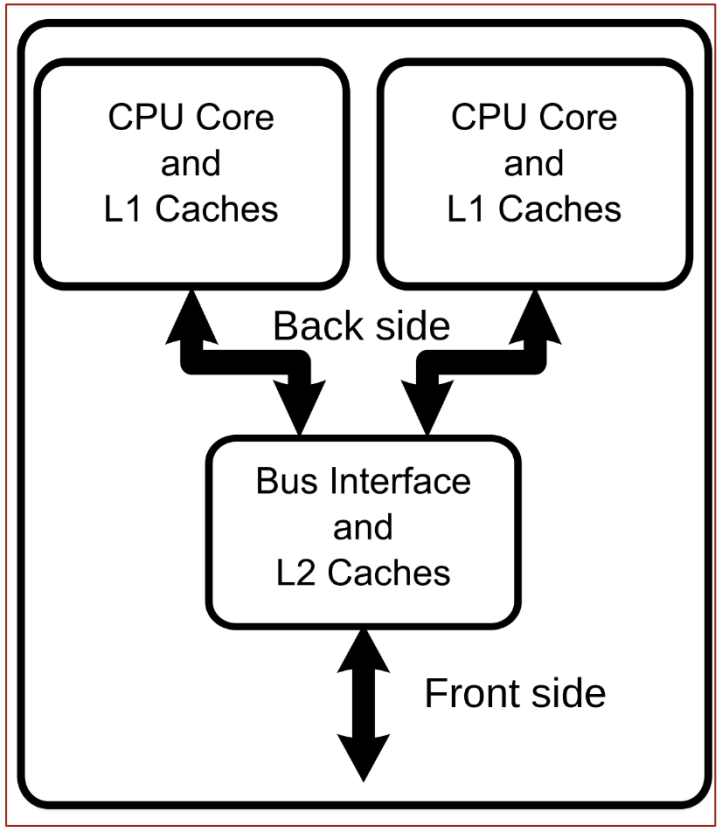
TECHNOLOGIES FOR NETWORK-BASED SYSTEMS

INTRODUCTION

- Scalable computing explores hardware, software, and network technologies for distributed computing system design and applications.
- In particular, we will focus on viable approaches to building distributed operating systems for handling massive parallelism in a distributed environment.
- Here we need to understand about
  - Multicore CPUs and Multithreading Technologies
  - GPU Computing to Exascale and Beyond
  - Memory, Storage, and Wide-Area Networking
  - Virtual Machines and Virtualization Middleware
  - Data Center Virtualization for Cloud Computing

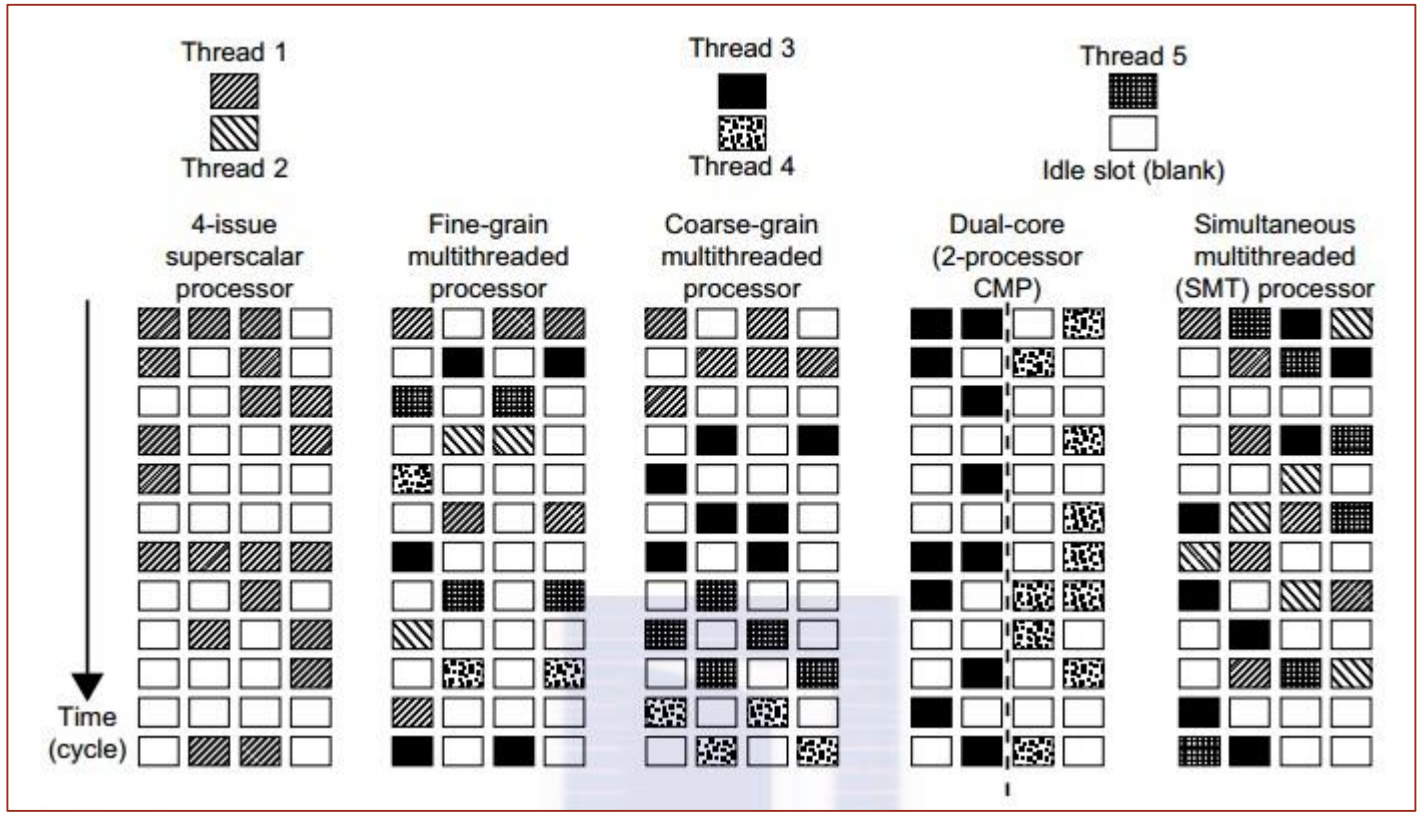
MULTICORE CPUS AND MULTITHREADING TECHNOLOGIES

- Both multi-core CPU and many-core GPU processors can handle multiple instruction threads at different magnitudes today.
- 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.
- In the future, multiple Core Multi-Processing (CMPs) could be built on the same CPU chip with even the L3 cache on the chip.



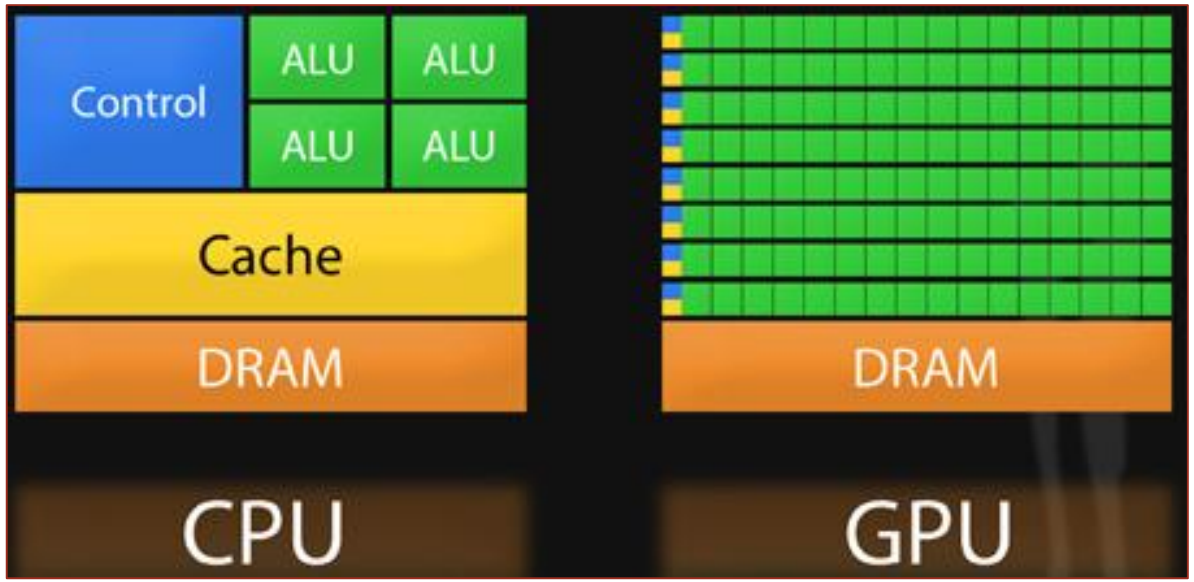
MULTITHREADING TECHNOLOGY

- In computer architecture, multithreading is the ability of a CPU to provide multiple threads of execution concurrently.
- This approach differs from multiprocessing.
- In a multithreaded application, the threads share the resources of a single or multiple cores, which include the computing units and the CPU caches.



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 complex graphics tasks in video editing applications.
- GPUs have a throughput architecture that shows massive parallelism by executing many concurrent threads.
- In the future, GPUs may appear in Exascale (EFLOPs* or PFLOPs) systems.



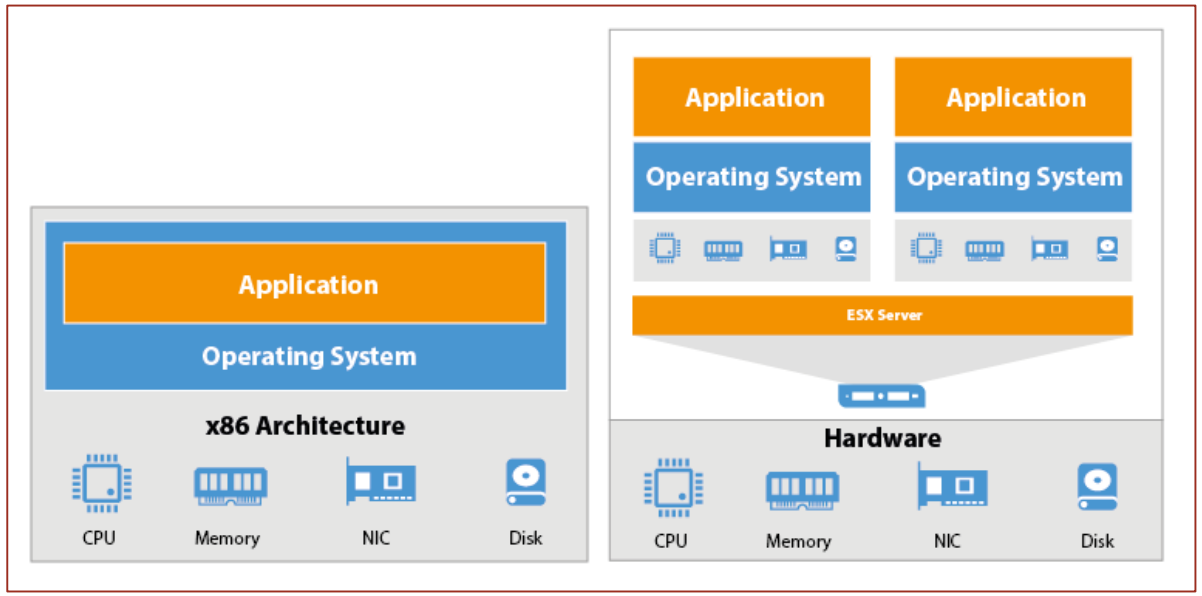
*FLOP - Floating-point arithmetic: It is a measure of computer performance, useful in fields of scientific computations that require floating-point calculations. More info: https://en.wikipedia.org/wiki/FLOPS

MEMORY, STORAGE, AND WIDE-AREA NETWORKING

- Memory Technology :
  - The memory chips have experienced a 4x increase in capacity every three years.
  - Memory access time did not improve much in the past.
  - The memory wall problem is getting worse as the processor gets faster.
- Disks and Storage Technology
  - Storage technology approximately increases 10x in capacity every eight years.
  - 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.
  - The memory wall may become even worse a problem limiting the CPU performance in the future.
- Wide-Area Networking
  - Rapid growth of Ethernet bandwidth usage
  - High-bandwidth networking increases the capability of building massively distributed systems.
  - Most data centers are using Gigabit Ethernet as the interconnect in their server clusters.

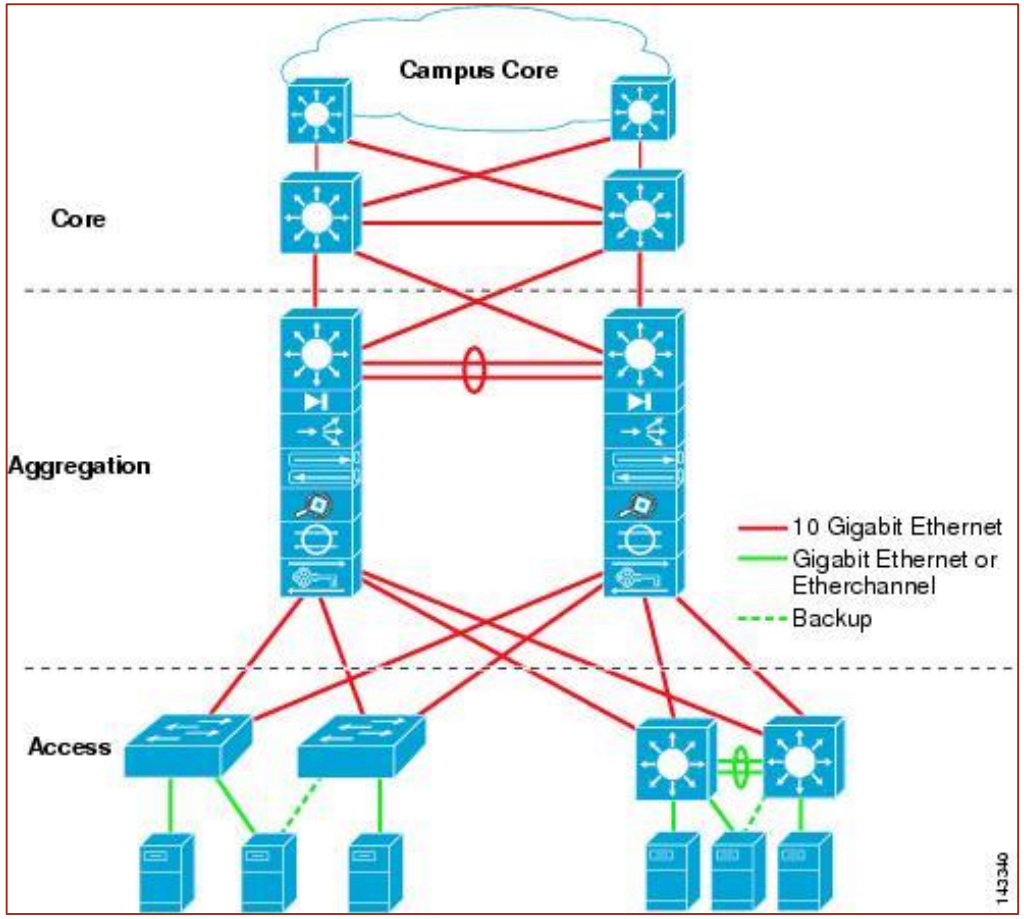
VIRTUAL MACHINES AND VIRTUALIZATION MIDDLEWARE

- A conventional computer has a single OS image.
- This offers a rigid architecture that tightly couples application software to a specific hardware platform.
- Some software running well on one machine may not be executable on another platform with a different instruction set under a fixed OS.
- Virtual machines (VMs) offer novel solutions to underutilized resources, application inflexibility, software manageability, and security concerns in existing physical machines.



DATA CENTER VIRTUALIZATION FOR CLOUD COMPUTING

- A large data center may be built with thousands of servers. Smaller data centers are typically built with hundreds of servers.
- The cost to build and maintain data center servers has increased over the years.
- Necessory to implement efficient Low-Cost Design Philosophy



CONT..

- Essentially, cloud computing is enabled by the convergence of technologies in four areas:
  - hard-ware virtualization and multi-core chips,
  - Utility and grid computing,
  - SOA, Web (2.0, 3.0, etc), and WS mashups, and
  - Autonomic computing and data center automation.
- Hardware virtualization and multicore chips enable the existence of dynamic configurations in the cloud.
- Utility and grid computing technologies lay the necessary foundation for computing clouds.
- SOA, Web (2.0, 3.0, etc), and WS mashups of platforms are pushing the cloud another step forward.
- Autonomic computing and automated data center operations contribute to the rise of cloud computing.

SOFTWARE ENVIRONMENTS FOR DISTRIBUTED SYSTEMS AND CLOUDS

SERVICE-ORIENTED ARCHITECTURE (SOA)

- Service-oriented architecture (SOA) is a style of software design where services are provided to the other components by application components, through a communication protocol over a network.
- A SOA service is a discrete unit of functionality that can be accessed remotely and updated independently
- SOA is also intended to be independent of vendors, products and technologies.
- SOA has four properties:
  - It logically represents a business activity with a specified outcome.
  - It is self-contained.
  - It is a black box for its consumers, meaning the consumer does not have to be aware of the service's inner workings.
  - It may consist of other underlying services.

TRENDS TOWARD DISTRIBUTED OPERATING SYSTEMS

- Tanenbaum identifies three approaches for distributing resource management functions in a distributed computer system.
- The first approach is to build a network OS over a large number of heterogeneous OS platforms. Such an OS offers the lowest transparency to users, and is essentially a distributed file system, with independent computers relying on file sharing as a means of communication.
- The second approach is to develop middleware to offer a limited degree of resource sharing, similar to the MOSIX/OS developed for clustered systems.
- The third approach is to develop a truly distributed OS to achieve higher use or system transparency.

PARALLEL AND DISTRIBUTED PROGRAMMING MODELS

- There are three of programming models, for distributed computing have scalable performance and application flexibility
- Message-Passing Interface (MPI)
  - This is the primary programming standard used to develop parallel and concurrent programs to run on a distributed system.
  - MPI is essentially a library of subprograms that can be called from C or FORTRAN to write parallel programs running on a distributed system.
- MapReduce
  - This is a web programming model for scalable data processing on large clusters over large data sets.
  - The user specifies a Map function to generate a set of intermediate key/value pairs. Then the user applies a Reduce function to merge all intermediate values with the same intermediate key.
- Hadoop Library
  - Hadoop offers a software platform that was originally developed by a Yahoo! group.
  - The pack-age enables users to write and run applications over vast amounts of distributed data.
  - Users can easily scale Hadoop to store and process petabytes of data in the web space.