CS528
High Performance Computing

A Sahu

Dept of CSE, IIT Guwahati

High Performance Computing

- What is HPC?
- Who needs high performance systems?
- How do you achieve high performance?
- How to analyze or evaluate performance?

- **Power Performance Tradeoff : Green Computing**
- Best architecture/design for a problem
- Parallel Architecture: Design and Programming
- **Cloud Computing, FOG/EDGE Computing/IoT**

Course Website

- http://jatinga.iitg.ernet.in/~asahu/cs528/
  - MS team Group : Grp_CS528_2022
- Course Contents
  - **Mostly Algorithmic Nature: Require CS204/CS512**
- Text and Reference Books
- All lecture slides
- Summery of each class with references
- Other information
  - Benchmarks, Source Code,
  - Referred Papers, EBooks

Grading Policy & General

- Class timing & Venue
  - Tue, Wed, Thursday : 3PM -4 PM
  - Venue : MS Team, Online, Video recording, PPT will be shared in MS Team
- Grading (HPC-3-0-0-6) : **No programming assignment** ☺☺☺
  - 2 Pre-announce quizs, Mid Sem, End Sem
  - Quizs (30%), MidSem(30%), End Sem (30%), 10% Class participation
  - **Rapid Mode MCQ/SA Exam**
  - **Missing Exam/Quiz due to medical/Internet cases: Average of others exams**
- No single text book is available
  - Two books : Hager HPC Book and Paterson CA Book
  - Many other resources: Manuals and Papers

What are Supercomputers Used For?

- Scientific simulations
- Animated graphics
- Analysis of geological data
- Nuclear energy research and meteorology
- Computational fluid dynamics
- **Analysis of business data**
  - **Online Sales**
- **Analysis of social data**
  - **Social media, Facebook, Utube, LinedIn,…**

How do you achieve high performance?

- Performance: FLOPS or MIPS
- High Performance = => Increase FLOPS
- How?

How do you achieve high performance?

- How?
  - Increase number of FPU of the system
  - Increase number processor in the system
  - Increase amount of Register/Cache/RAM of system
  - Use different cache/RAM mapping/management policy
  - Restructure Program, Use different Language
  - Use different compiler
  - Use different algorithm/approaches for same problem
- Cost, AMC, Power Consumption

How ?

- Increase number of FPU of the system
  - Vector Processor (SSE, SIMD, MMX), GPU Accelerator
- Increase number processor in the system
  - **Core i3/i5/i7, Ryzen R3/R5/R7: Dual/Quad/Hexa/Octa cores**
  - Intel Xeon 4,6,8,10,12,16,18,20, 24, 38 cores
  - Intel Xeon Phi (KNL), 72 cores/288Thread
  - AMD Thread Ripper : 8, 16, 32, 64 cores
- Increase amount of Register/Cache/RAM of system
  - Big register file/Cache :Power Hungry
  - RAM/NVRAM /SSD: No disk moment fast but costly

Technology Trends

- Desktop 8086/80386
  - Processor, Mother Board, Co-Processor (Floating Point Unit), Graphics Card, RAM, Audio, Ethernet
- Desktop Pentium
  - Processor (Coprocessor inside) + Mother Board (Audio, Ethernet) + Graphics Card
- Desktop PIV
  - Processor + Mother Board (Graphics + Audio + Ethernet)
- Desktop Core
  - Processor (Graphics Inside) + Board (Audio, Ethernet)
- Mobile SOC
  - Processor + Graphics+ Board (Almost in Chip)

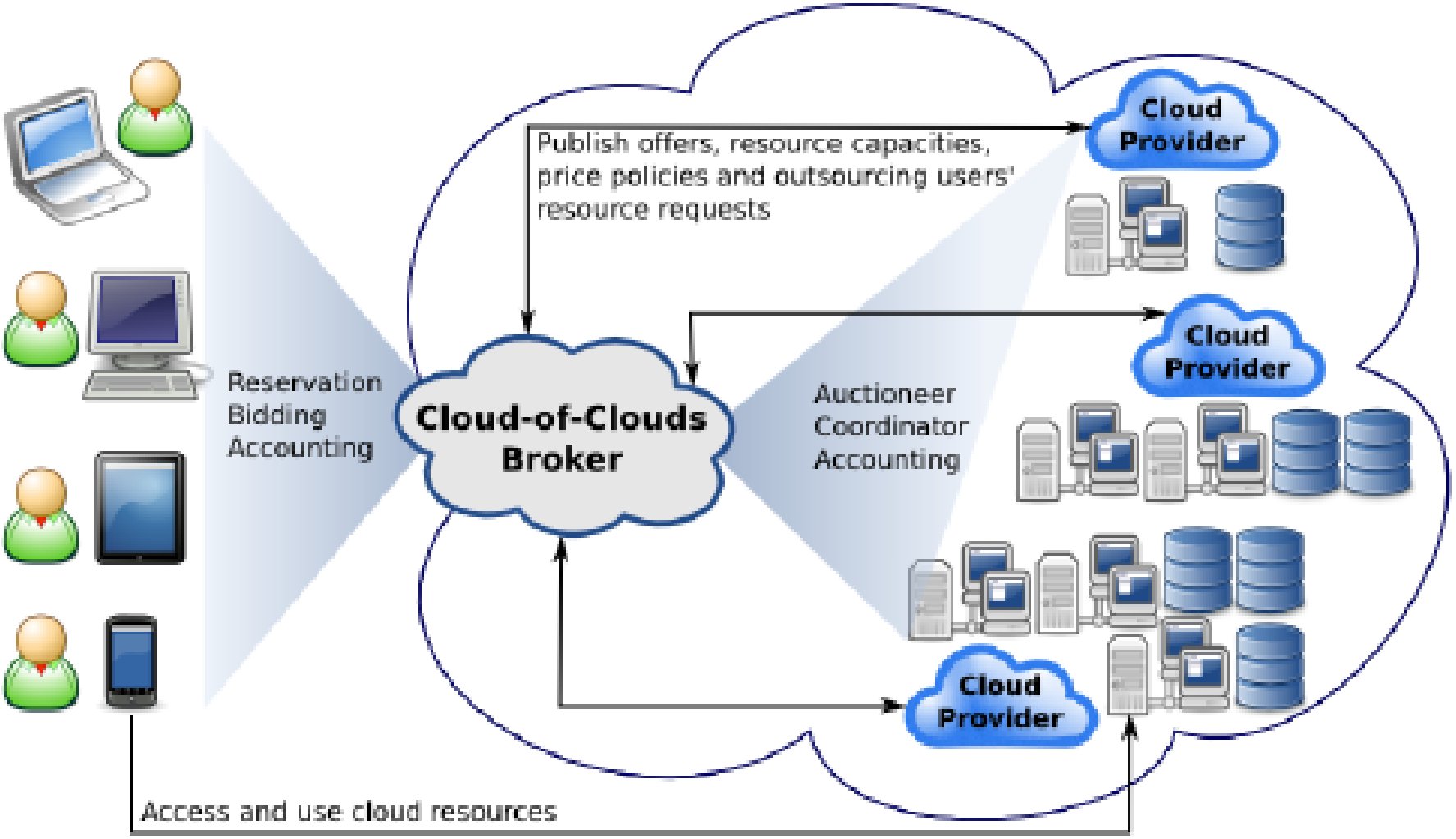
Technology Trend

- Performance is no longer is main issue
  - Power, Energy, Cost
  - DVFS : run at lower frequency to reduce power/energy consumption
- Most of modern day servers are
  - Under utilized (core, RAM)
  - Same for Laptop/Desktop/Mobile
- Under utilized
  - Wastage of resources, **can be shared with others**
  - Sharing methodology (virtualization)
  - Leads to Cloud Computing

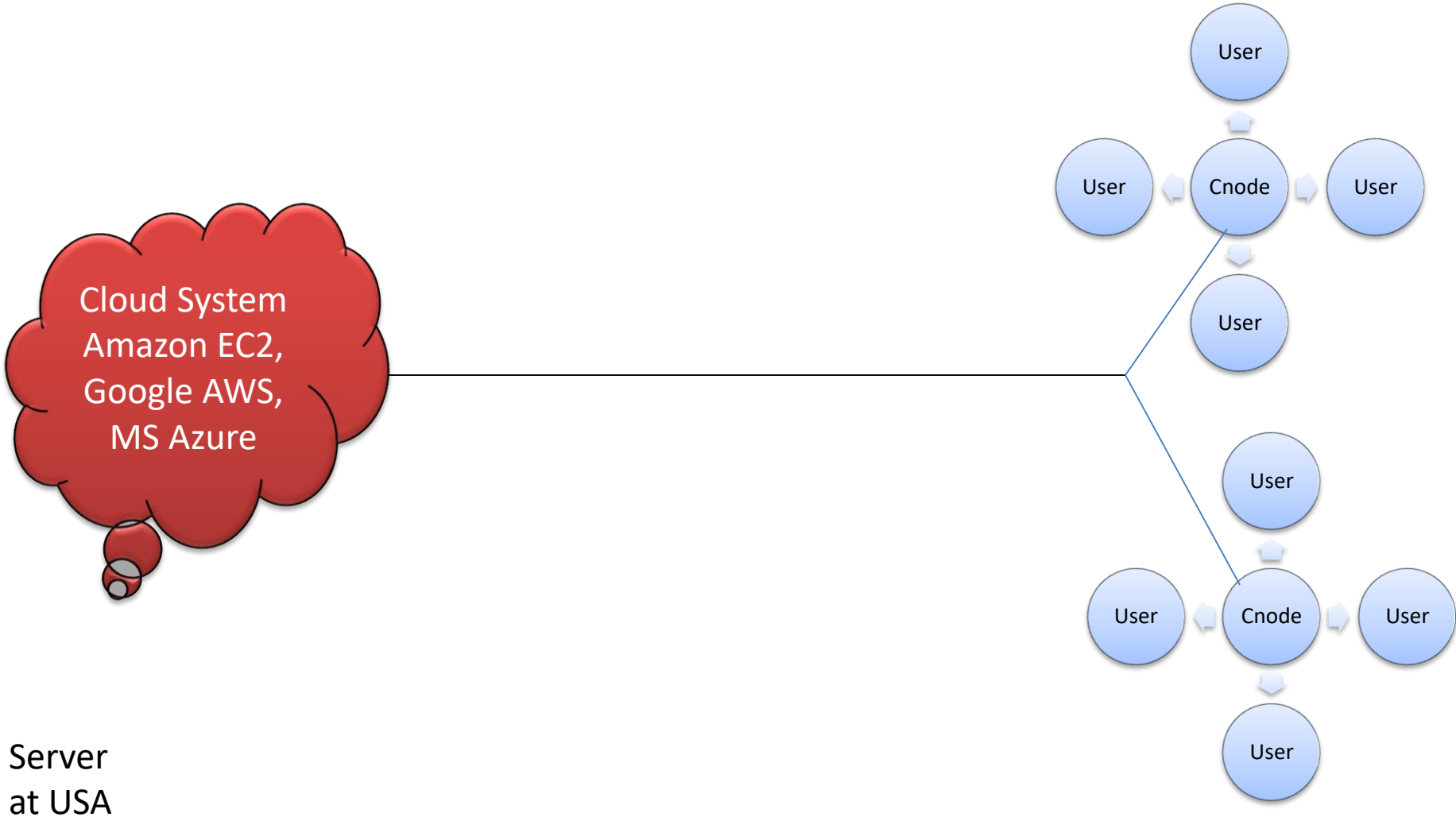
Technology Trend

- Cloud Computing
  - Economy: Similar to OLA/UBER
  - Renting Model
- IoT : Many things on Internet
  - Control and Management of Big Work
  - Sensors and actuators
- FOG
  - Peers Computing, Multiple Level
- Edge
  - Computing at Edge not far, Latency sensitive

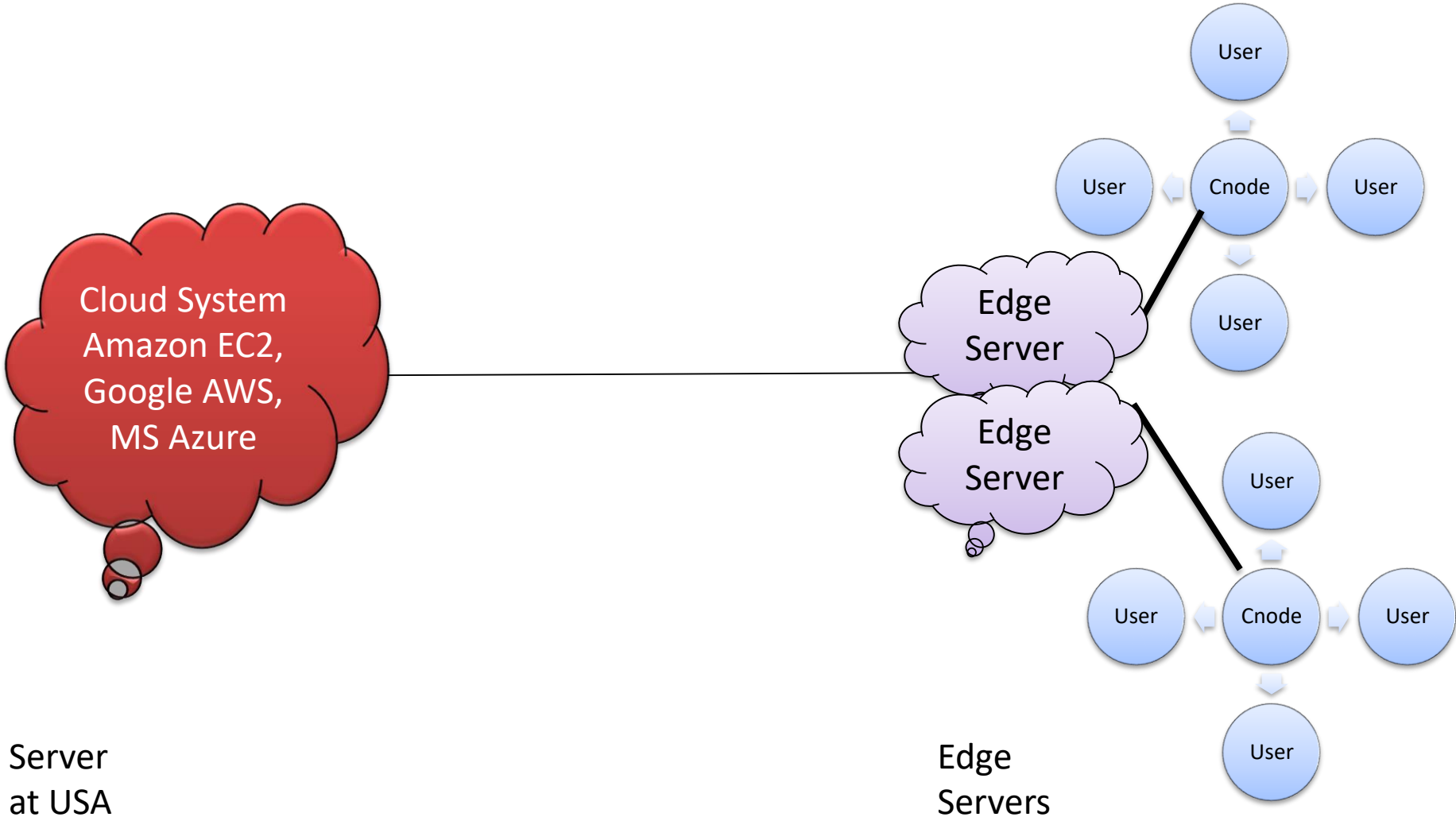
Cloud/IoT/Edges/FoG



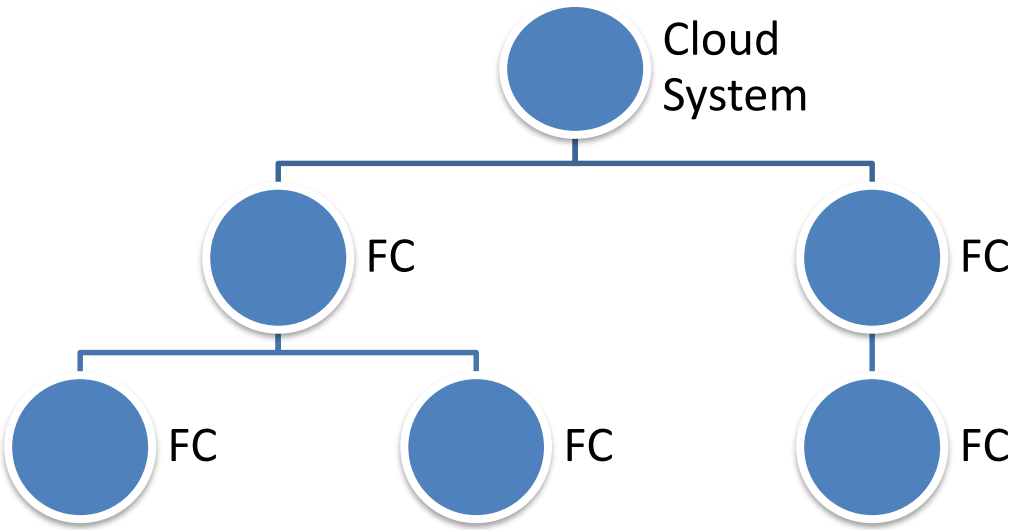
Cloud/IoT/Edges/FoG



Cloud/IoT/Edges/FoG



Cloud/IoT/Edges/FoG



Technology Trend

- Single processor/Single Computer
  - Single processor with SIMD instruction
- Multi Computer
  - Cluster, Data need to travel outside PC via LAN cable
- Multi processor
  - Tightly coupled, Data no need to travel out side PC, out side board
- Processor + Accelerator
  - PCI or Board level Communication
- Processor and Accelerator in the same chip
  - On chip, High BW , Intel Core (Graphics are in Chip)
- 3D chip

Quest for Performance

Quest for Performance

- Pipelining
- Superscalar Architecture
- Out of Order Execution
- Caches, SMT
- ISA Advancements

Single Processor
Past research

- Parallelism
  - Multi-core processors
  - Clusters
  - **Grid, Cloud System**

This is the current and future

Trend of HPC

- **HPC system**
  - Multi Nodes/Computer/Blades
  - **Programming Model MPI**
- **Nodes are Multicore**
  - Node have accelerators
  - **Programming Model : OpenMP, OpenCL/Cuda**
- **Core**
  - Multi Threaded
  - With vector instructions
  - 4 issue OOO Pipelines, Multilevel Caches,
  - **Programming Model: gcc optimized, vectorized code, OpenMP**

Need to study in HPC: User Prospects

- Single Processor
  - Architecture: Core Pipeline, Core Multithreading, Cache Hierarchy, SIMD
  - C/C++ Optimization Methods: gcc, OpenMP, Simidization, cache optimized code
- Multicore node
  - Multicore, Accelerator, Interconnections
  - OpenMP Model, Cuda Model, Accelerated Model
- HPC Server
  - Multiple Nodes/Blades, Interconnection, Storages
  - Programming Model : MPI

HPC : overall

- Top 500 HPC : Multiprocessor, Accelerator based
- Applications : Programming Model, Management
- Cost of HPC: Initial cost (System: Racks, Rack server, SAS) , Place, AC, ..
- Running Cost of HPC : AMC, Energy, Management
- HPC on Rent :
  - VM, Management, Revenue Model, Cost Model
  - Cloud Model, IasS, PasS, SaaS (Infra/Platform/Software)

HPC Course Contents (Abstract)

- Parallel/Multicore Architecture
  - Multicore, GPU, Xeon Phi
- Programming Model : Thread, OpenMP, MPI, Cilk, Cuda, Intel MKL
- **Scheduling and Management**
  - **Resources: Core, RAID/NAS**
- **Benchmarking and Analysis**
- **Cloud : Virtualization, Cost, Revenue Model**
- **Energy Efficient and Power Efficiency**
- **Theme FOG/EGDE = Mobile Cloud Computing**

HPC course

- 1st Half : Before Mid Sem
  - HPC, Architecture, Programming, Code Optimization
  - Scheduling, Energy Efficiency, Power Efficiency
- 2nd Half : After Mid Sem
  - Cloud Computing
  - Mobile Cloud, FOG, EGDE, IoT

Books : Text

- Hager G and Wellein G . ***Introduction to High Performance Computing for Scientists and Engineers*** (1st ed.). CRC Press,, India, 2010.
- **Some user manuals**
- **Some recent papers**