

Guest Lecture 1: January 23

*Lecturer: TO BE FILLED**Scribe: Daniel Thiyagu*

1.1 Application Agnostic RPC

SOAP: SOAP is an application agnostic RPC Protocol and is a web services definition language. RPC's hook into cloud computing, recently amazon deprecated SOAP and replaced it with REST.

1.2 Differences between HPC and Data Center

High Performance Computing:

- Generally managed and used by scientific community.
- They focus on high computational workloads.
- The types of machines are Highly Parallel Mainframes.
- Example: Used in weather forecasting.

Data Center:

- Generally managed and associated with enterprises.
- They focus on collecting data.
- They don't focus on high computational workloads.

1.3 Types of Data Center

Traditional Data Center:

- Sys admins monitor and manage servers.
- They schedule process to be run.
- Applications run on physical servers.
- They use Storage Array Networks(SAN) & Network Attached Storage(NAS). SAN and NAS : Provides an illusion of shared disk among servers. NAS: File system abstraction, attached file system to all servers in a network.

Modern Data Center:

- Dynamic Larger Scale
- They transfer process to get better efficiency.
- Applications run on virtual machines.
- Increased automation allows larger scales.

Modular Data Center: They have an inbuilt cooling infrastructure and inside shipping containers are placed data centers and they are like plug and play.

1.3.1 How does FB distributes wall feed?

Data is not present in one data center, they touch a lot of servers and hence more complex process like load balancing and pipelining would be different. FB wall is generally formed on demand and prioritize preference and popularity.

1.3.2 Does Archived Old Data have specialized storage centers ?

Maybe. It is dependent on the company and depends on ratio of storage to compute.

1.4 MGHPCC - Massachusetts Green High Performance Computing Cluster

:

- MGHPCC - multi tenant Datacenter - Only infrastructure level optimization possible, whereas in a single-tenant DC we can move applications around.
- They are managed remotely, servers have a black pane ie. remote bios and can install new OS. They also have a remote power switch. They also have internet connectivity.
- Electricity is cheaper in holyoke, land cost is cheaper, fiber networking(Springfield to boston available, so reduces network latency) and Water is used as a form of coolant, which makes Holyoke ideal. Power used is Green - renewable energy.
- It requires 15 MWatts to be in operation

1.5 VM Migration in a LAN

: It migrates memory state, migrating disk state is more complex. In Data Centers, disk memory is shared and therefore migration is simple. When we go outside of LAN, it becomes WAN, and IP is based on geographical location. Data Center Networking is a research area,

1.6 Server Virtualization

:

- VM aware machine: Trap into lower level hypervisor
- HVM : Hardware Virtual Machine Technology
- KVM : Linux
- Xen : Used by research community for migration, load balancing research, etc. It is cheap virtualization for the masses. They worked with Microsoft team for virtualizing Microsoft Windows.

1.6.1 Why was x86 not virtualized previously ?

: It was not possible because privileged instructions were an issue, they interrupt and call handler to handle it. VMware was the first to virtualize it by inserting dynamic calls into the lower level handlers. Then x86 were fixed to be free from this issue in recent years.

1.6.2 Virtualization in Data Centers

: **Virtual Servers:**

- Balance/consolidate load
- Faster Deployment
- Easy Maintenance

Virtual Desktops:

- host employee desktops in VM
- Thin Client Model

1.7 Data Centers

Resource Management:

- We need to keep it as highly utilized as possible
- Apps have variable/unpredictable workload.
- Want High Performance and Low Cost
- Automated Resource management

Energy Efficiency:

- Servers consume a lot of energy
- Be Green
- Save Money

In general, it takes 50% power to run the computers and the remaining power to cool them.

Reliability challenges:

- 0.5% overheat
- Power Distribution unit failures
- Network Rewiring - as you add/remove systems, etc.

Data Center Cost:

$$PowerUsageEffectiveness = \frac{ITPower}{TotalPower}$$

Economy of Scale: Location impacts:

- presence of customers nearby
- cooling
- Generation of power.
- presence of already existing grid power infrastructure.

1.8 Cloud

What is Cloud? It is remote, you pay as you go, we get high scalability, shared infrastructure.

IaaS Infrastructure as a Service: Eg:Google, Aws

PaaS Platform as a Service: Eg:Azure, Google App Engine

SaaS Software as a Service: Eg:Applications like Salesforce, Gmail.

Hybrid Cloud: It is a mix of private and public cloud usage.

Programming Models:

- Client Server(Interactive)
- Batch Processing(Not Interactive)
- Map reduce.(Not Interactive)

Future Challenges:

- Privacy/Security
- Extreme Scalability
- Programming Models