

## Internet Data Centers

Kai Shen

4/13/2011

CSC 258/458 - Spring 2011

1

## Data Centers

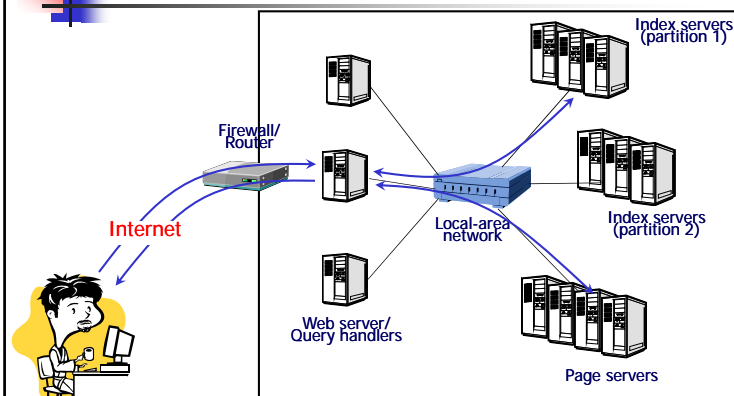
- Large distributed systems of computers and storage
- Often host interactive Internet services, but some are also used for batch processing and data mining
- Can be at very large scale:
  - Up to 10s/100s of thousands of machines (millions of processors)
  - megawatts or more in power usage
  - Serve millions of users (thousands at a time)
  - Host large amount of data: partitioning and replication

4/13/2011

CSC 258/458 - Spring 2011

2

## Data Center Hosting Web Search



4/13/2011

CSC 258/458 - Spring 2011

3


## Data Center Architecture

- Big multiprocessors vs. cluster of commodity machines
- Earlier days
  - Scientific computing is parallelizable, but complex synchronizations glue the parallel processes
  - Distributed databases are not very scalable
  - Many points of failure in commodity machine clusters
- Late 90s, early 2000
  - Embarrassingly parallel web services with explosive growth, e.g., e-commerce
  - Killer application - web search
  - Turn independent failures from a curse to a strength, migrate a server farm

4/13/2011

CSC 258/458 - Spring 2011


4



## Data Center Storage Architecture

- Ideally, all machines can access all the data, with high performance and reliability
- Centralized storage, connected to all machines through network, managed by databases or file systems
- Every machine has some local storage
  - More scalable
  - Asymmetric performance
  - Complex management
  - Poor reliability/durability


4/13/2011 CSC 258/458 - Spring 2011 5



## Data Center Data Management

- Distributed data management in a unified namespace and access interface
- Distributed file system that manages distributed storage on many machines, but present the view of a single FS
  - Google file system
- Data access interface:
  - File system
  - Relational SQL
  - Bigtable (Google)


4/13/2011 CSC 258/458 - Spring 2011 6



## Scalability

- Performance at scale
  - Overhead doesn't scale as much
- Parallel processing and result aggregation
  - Load partitioning and distribution
  - Bound by the slowest but slowest can be really slow (or it can fail altogether)
- Mapreduce (Google)
- Partial aggregation if it meant partial degradation of quality
  - Search the same thing twice and get different results


4/13/2011 CSC 258/458 - Spring 2011 7



## Reliability

- Why does a machine fail?
  - Hardware problems
  - Software problems
  - Neither
- Reliability rate for one machine is  $p$ , for  $n$  machines is  $p^n$
- Redundancy
  - Memory ECC
  - Machines
  - Data storage
    - Replication of mutable states introduces consistency problems
  - Networking
  - Power supply
- Lose sleep over correlated failures


4/13/2011 CSC 258/458 - Spring 2011 8



## More on Reliability

- Correlated failures
  - Deterministic parallel execution???
  - Load-triggered failures
  - Malicious activities
- Data collection and statistical analysis
  - Overhead
  - Careful consideration of data collection scenario (memory fails more often when the system is more heavily used)
- Recovery-oriented computing
  - Components are inherently unreliable, design a system in which components can be independently recovered (rebooted) without significant effects on others


4/13/2011 CSC 258/458 - Spring 2011 9



## Management

- Easy to add/remove components
  - Container-based data center modules


4/13/2011 CSC 258/458 - Spring 2011 10



## Power

- Power consumption of a large data center is significant
- The power distribution system is also expensive
  - In building cost and power lost during distribution
- People care about the total energy usage and peak power
- What do people do?
  - Build your data center next to a hydro power plant
  - Energy conservation through server consolidation
    - Weigh spare capacity against load spikes and speed of server powerup
  - Lower power spikes through load shifting or throttling

4/13/2011 CSC 258/458 - Spring 2011 11



## Resource Provisioning

- Power provisioning
  - Labeled peak power is never reached
  - A machine spends most of its time under-utilized, away from peak power
  - Machines don't reach peak power simultaneously
    - Really?
- Provisioning of computing/storage resources
  - Curse of hugely successful Internet services
  - Cloud computing

4/13/2011 CSC 258/458 - Spring 2011 12