

# **EEL 5764 Computer Architecture**

## **Lecture 37: Warehouse Scale Computing**

**Sandip Ray**

Department of Electrical and Computer Engineering  
University of Florida

# Introduction

- Warehouse-scale computer (WSC)
  - Provides Internet services
    - Search, social networking, online maps, video sharing, online shopping, email, cloud computing, etc.
  - Differences with HPC “clusters”:
    - Clusters have higher performance processors and network
    - Clusters emphasize thread-level parallelism, WSCs emphasize request-level parallelism
  - Differences with datacenters:
    - Datacenters consolidate different machines and software into one location
    - Datacenters emphasize virtual machines and hardware heterogeneity in order to serve varied customers

# Introduction

- Important design factors for WSC:
  - Cost-performance
    - Small savings add up
  - Energy efficiency
    - Affects power distribution and cooling
    - Work per joule
  - Dependability via redundancy
  - Network I/O
  - Interactive and batch processing workloads
  - Ample computational parallelism is not important
    - Most jobs are totally independent
    - “Request-level parallelism”
  - Operational costs count
    - Power consumption is a primary, not secondary, constraint when designing system
  - Scale and its opportunities and problems
    - Can afford to build customized systems since WSC require volume purchase

# Programming Models and Workloads

- Batch processing framework: MapReduce
  - **Map:** applies a programmer-supplied function to each logical input record
    - Runs on thousands of computers
    - Provides new set of key-value pairs as intermediate values
  - **Reduce:** collapses values using another programmer-supplied function

# Programming Models and Workloads

- Example:
  - **map (String key, String value):**
    - // key: document name
    - // value: document contents
    - for each word w in value
      - `EmitIntermediate(w,"1");` // Produce list of all words
  - **reduce (String key, Iterator values):**
    - // key: a word
    - // value: a list of counts
    - `int result = 0;`
    - for each v in values:
      - `result += ParseInt(v);` // get integer from key-value pair
    - `Emit(AsString(result));`

# Programming Models and Workloads

- MapReduce runtime environment schedules map and reduce task to WSC nodes
- Availability:
  - Use replicas of data across different servers
  - Use relaxed consistency:
    - No need for all replicas to always agree
- Workload demands
  - Often vary considerably

# Computer Architecture of WSC

- WSC often use a hierarchy of networks for interconnection
- Each 19" rack holds 48 1U servers connected to a rack switch
- Rack switches are uplinked to switch higher in hierarchy
  - Uplink has  $48 / n$  times lower bandwidth, where  $n = \#$  of uplink ports
    - "Oversubscription"
  - Goal is to maximize locality of communication relative to the rack

# Storage

- **Storage options:**
  - **Use disks inside the servers, or**
  - **Network attached storage through Infiniband**
- **WSCs generally rely on local disks**
- **Google File System (GFS) uses local disks and maintains at least three replicas**



# Array Switch

- **Switch that connects an array of racks**
  - **Array switch should have 10 X the bisection bandwidth of rack switch**
  - **Cost of  $n$ -port switch grows as  $n^2$**
  - **Often utilize content addressable memory chips and FPGAs**

# WSC Memory Hierarchy

- Servers can access DRAM and disks on other servers using a NUMA-style interface

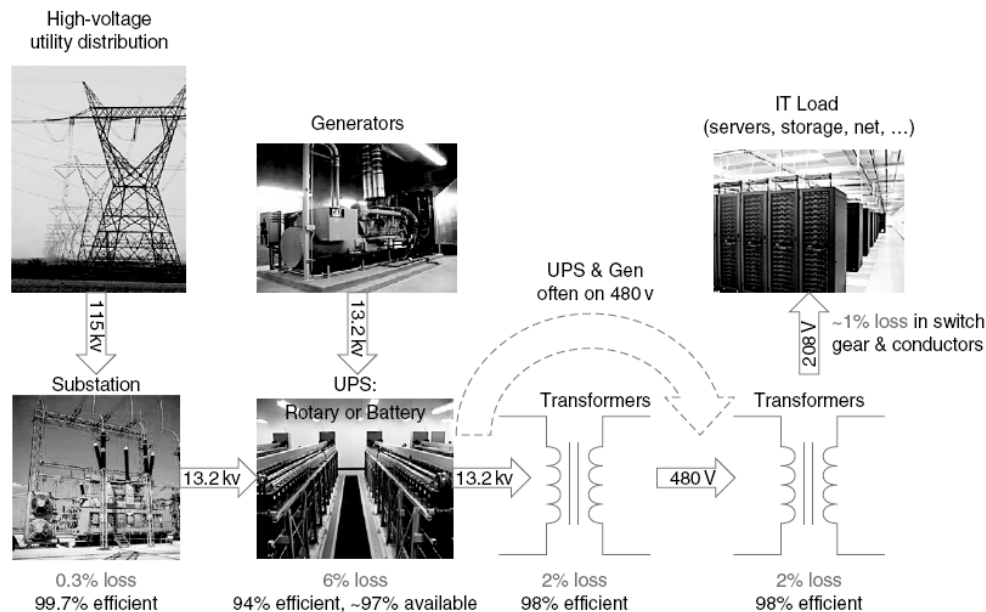
	Local	Rack	Array
DRAM latency (microseconds)	0.1	100	300
Disk latency (microseconds)	10,000	11,000	12,000
DRAM bandwidth (MB/sec)	20,000	100	10
Disk bandwidth (MB/sec)	200	100	10
DRAM capacity (GB)	16	1,040	31,200
Disk capacity (GB)	2000	160,000	4,800,000

# Infrastructure and Costs of WSC

## ■ Location of WSC

- Proximity to Internet backbones, electricity cost, property tax rates, low risk from earthquakes, floods, and hurricanes

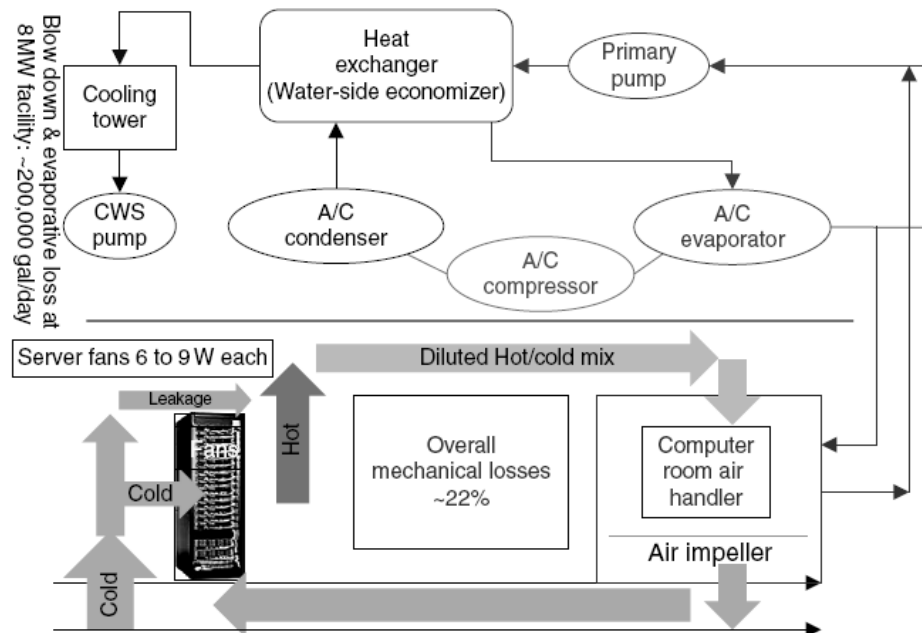
## ■ Power distribution



# Infrastructure and Costs of WSC

## ■ Cooling

- Air conditioning used to cool server room
- 64 F – 71 F
  - Keep temperature higher (closer to 71 F)
- Cooling towers can also be used
  - Minimum temperature is “wet bulb temperature”



# Infrastructure and Costs of WSC

- **Cooling system also uses water (evaporation and spills)**
  - E.g. 70,000 to 200,000 gallons per day for an 8 MW facility
- **Power cost breakdown:**
  - Chillers: 30-50% of the power used by the IT equipment
  - Air conditioning: 10-20% of the IT power, mostly due to fans
- **How many servers can a WSC support?**
  - Each server:
    - “Nameplate power rating” gives maximum power consumption
    - Must consider actual, measured power under actual workloads
  - Oversubscribe cumulative server power by 40%, but monitor power closely

# Measuring Efficiency of a WSC

- **Power Utilization Effectiveness (PEU)**
  - = Total facility power / IT equipment power
  - Median PUE on 2006 study was 1.69
- **Performance**
  - Latency is important metric because it is seen by users
  - Bing study: users will use search less as response time increases
  - Service Level Objectives (SLOs)/Service Level Agreements (SLAs)
    - E.g. 99% of requests be below 100 ms

# Cost of a WSC

- **Capital expenditures (CAPEX)**
  - **Cost to build a WSC**
- **Operational expenditures (OPEX)**
  - **Cost to operate a WSC**

# Cloud Computing

- **WSCs offer economies of scale that cannot be achieved with a datacenter:**
  - **5.7 times reduction in storage costs**
  - **7.1 times reduction in administrative costs**
  - **7.3 times reduction in networking costs**
  - **This has given rise to cloud services such as Amazon Web Services**
    - **“Utility Computing”**
    - **Based on using open source virtual machine and operating system software**

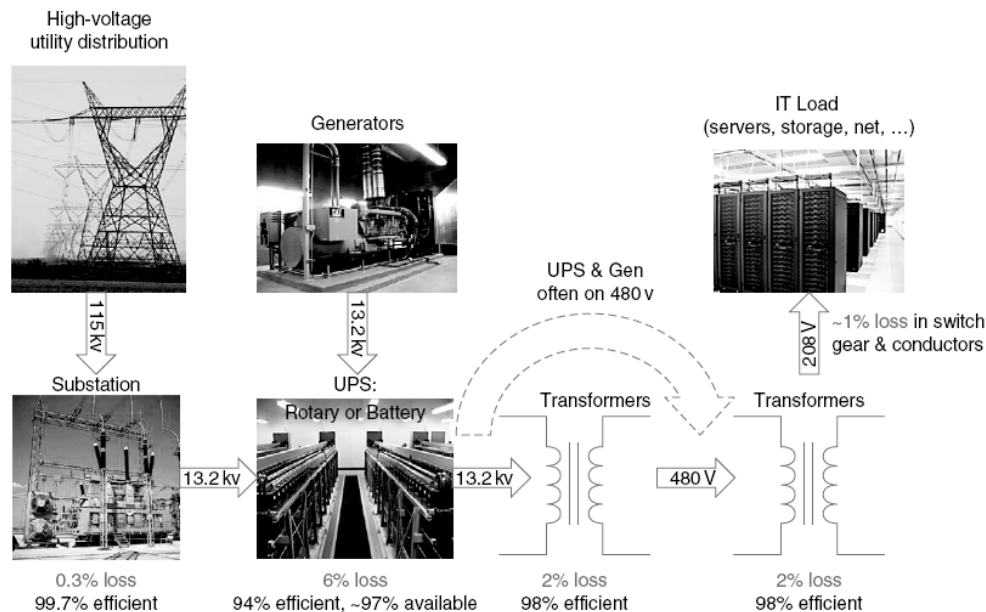


# Infrastructure and Costs of WSC

## ■ Location of WSC

- Proximity to Internet backbones, electricity cost, property tax rates, low risk from earthquakes, floods, and hurricanes

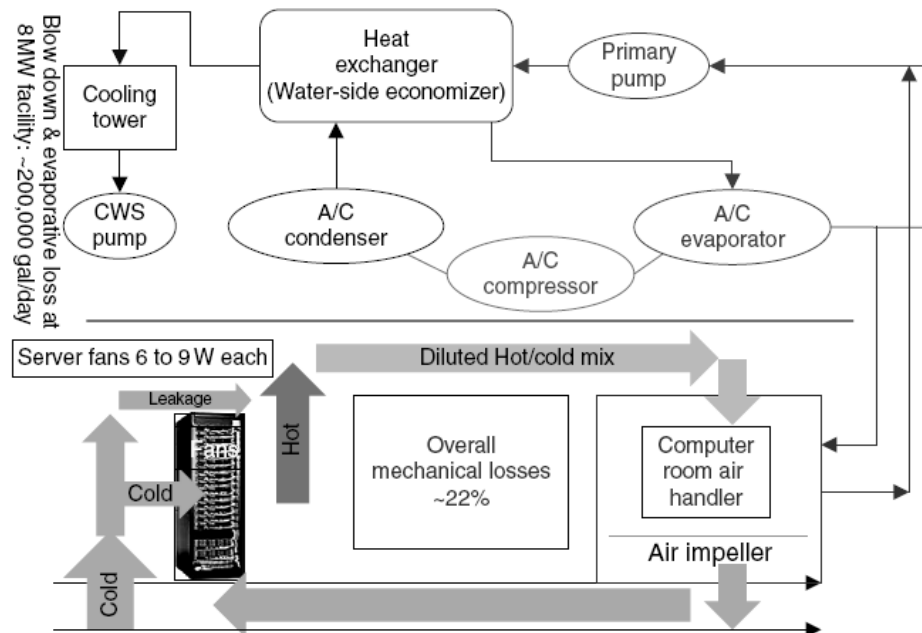
## ■ Power distribution



# Infrastructure and Costs of WSC

## ■ Cooling

- Air conditioning used to cool server room
- 64 F – 71 F
  - Keep temperature higher (closer to 71 F)
- Cooling towers can also be used
  - Minimum temperature is “wet bulb temperature”



# Infrastructure and Costs of WSC

- **Cooling system also uses water (evaporation and spills)**
  - E.g. 70,000 to 200,000 gallons per day for an 8 MW facility
- **Power cost breakdown:**
  - Chillers: 30-50% of the power used by the IT equipment
  - Air conditioning: 10-20% of the IT power, mostly due to fans
- **How many servers can a WSC support?**
  - Each server:
    - “Nameplate power rating” gives maximum power consumption
    - Must consider actual, measured power under actual workloads
  - Oversubscribe cumulative server power by 40%, but monitor power closely

# Measuring Efficiency of a WSC

- **Power Utilization Effectiveness (PUE)**
  - = Total facility power / IT equipment power
  - Median PUE on 2006 study was 1.69
- **Performance**
  - Latency is important metric because it is seen by users
  - Bing study: users will use search less as response time increases
  - Service Level Objectives (SLOs)/Service Level Agreements (SLAs)
    - E.g. 99% of requests be below 100 ms

# Cost of a WSC

- **Capital expenditures (CAPEX)**
  - **Cost to build a WSC**
- **Operational expenditures (OPEX)**
  - **Cost to operate a WSC**

# Cloud Computing

- **WSCs offer economies of scale that cannot be achieved with a datacenter:**
  - **5.7 times reduction in storage costs**
  - **7.1 times reduction in administrative costs**
  - **7.3 times reduction in networking costs**
  - **This has given rise to cloud services such as Amazon Web Services**
    - **“Utility Computing”**
    - **Based on using open source virtual machine and operating system software**