Chapter 6: Warehouse Computers **to Exploit Request-Level and Data-Level Parallelism**

Introduction

- Warehouse-scale computer (WSC) Provides Internet services Search, social networking, online maps, video sharing, online shopping, email, cloud computing, etc.
- Houses 50,000 to 100,000 servers
- Differences with HPC “clusters”:
  - Clusters have higher performance processors and network
  - Clusters emphasize thread-level parallelism, WSCs emphasize request-level parallelism

Important design factors for WSC

- Cost-performance Small savings add up
- Energy efficiency Affects power distribution and cooling
- Dependability via redundancy (Availability, 99.99%) down < 1 hr/year
  - Redundancy management (multi-WSC)
- Network I/O
- Both interactive (like search) and batch processing workloads (parallel batch programs to compute metadata useful to search, for instance)

Important design factors for WSC

- Request Level Parallelism
  - Software as a service (SaaS): millions of independent users
  - Most jobs are totally independent
  - Little need to coordinate or synchronize.
- Operational costs count
  - Power consumption is a primary, not secondary, constraint when designing system
  - Energy, power distribution, and cooling: more than 30% of cost over 10 years.
- Scale and its opportunities and problems
  - Can afford to build customized systems since WSC require volume purchase
  - Expect one disk failure per hour for 50,000 severs WSC

Outages and Anamolies

Approx. number events in 1st year	Cause	Consequence
1 or 2	Power utility failures	Lose power to whole WSC; doesn't bring down WSC if UPS and generators work (generators work about 99% of time).
4	Cluster upgrades	Planned outage to upgrade infrastructure, many times for evolving networking needs such as recabling, to switch firmware upgrades, and so on. There are about 9 planned cluster outages for every unplanned outage.
1000s	Hard-drive failures	2% to 10% annual disk failure rate [Pinheiro 2007]
	Slow disks	Still operate, but run 10x to 20x more slowly
	Bad memories	One uncorrectable DRAM error per year [Schroeder et al. 2009]
	Misconfigured machines	Configuration led to ~30% of service disruptions [Barroso and Hölzle 2009]
	Flaky machines	1% of servers reboot more than once a week [Barroso and Hölzle 2009]
5000	Individual server crashes	Machine reboot, usually takes about 5 minutes

Figure 6.1 List of outages and anomalies with the approximate frequencies of occurrences in the first year of a new cluster of 2400 servers. We label what Google calls a cluster an *array*; see Figure 6.5. (Based on Barroso [2010].)

Programming Models and Workloads for WSC

- Batch Programming Framework: MapReduce
- Google uses MapReduce
- Facebook runs Hadoop which is open source
- Map runs the same function to each logical input record on lots of computers and gets an intermediate result of Key-Value pair.
- Reduce collects these results from each of the distributed tasks and collapse them using a programmer specified function

Big Data: Use for MapReduce

- Complex structured and unstructured data.
- Heterogeneous and continuously generated data.
- Data generated from a variety of sources: sensors, cameras, scientific instruments, records of transactions, etc.
- Applications: intelligence, health-care, business analytics, disease prevention, crime prevention, etc.

Main Points of MapReduce

- MapReduce runtime environment schedules
- map and reduce task to WSC nodes
- Availability:
  - Use replicas of data across different servers
- Workload demands
    - Often vary considerably
- More details on MapReduce in Bahadar Ali's presentation

Infrastructure and Cost of WSC

- Location:
  - Proximity to Internet backbones, electricity cost,
  - low risk from earthquakes, floods, and hurricanes
- Power and Cooling
  - Power efficiency is important
  - Cooling can be reduces by having equipment that work in higher temperatures.
  - Cooling be done by using either Air conditioning or water evaporation.

Measuring Efficiency of a WSC

- Power Utilization Effectiveness (PUE) = Total facility power / IT equipment power
  - Median PUE on 2006 study was 1.69

- WSC perform both batch processing and interactive processing.
  - Throughput is important but priority should be given to latency.
  - Bing study: users will use search less as response time increases