## CLUSTERS AND WAREHOUSE SCALE COMPUTERS

Warehouse-scale computers form the foundation of internet services. The present days WSCs act as one giant machine. The main parts of a WSC are the building with the electrical and cooling infrastructure, the networking equipment and the servers.

> ***A cluster is a collection of desktop computers or servers connected together by a local area network to act as a single larger computer. A warehouse-scale computer (WSC) is a cluster comprised of tens of thousands of servers***

### WSCs as Servers

The following features of WSCs that makes it work as servers:

- **Cost-performance**: Because of the scalability, the cost-performance becomes very critical. Even small savings can amount to a large amount of money.
- **Energy efficiency:** Since large numbers of systems are clustered, lot of money is invested in power distribution and for heat dissipation. Work done per joule is critical for both WSCs and servers because of the high cost of building the power and mechanical infrastructure for a warehouse of computers and for the monthly utility bills to power servers. If servers are not energy-efficient they will increase
  - cost of electricity
  - cost of infrastructure to provide electricity
  - cost of infrastructure to cool the servers.
- **Dependability via redundancy:** The hardware and software in a WSC must collectively provide at least 99.99% availability, while individual servers are much less reliable. Redundancy is the key to dependability for both WSCs and servers. WSC architects rely on multiple cost-effective servers connected by a low cost network and redundancy managed by software. Multiple WSCs may be needed to handle faults in whole WSCs. Multiple WSCs also reduce latency for services that are widely deployed.
- **Network I/O:** Networking is needed to interface to the public as well as to keep data consistent between multiple WSCs.

- **Interactive and batch-processing workloads**: Search and social networks are interactive and require fast response times. At the same time, indexing, big data analytics etc. create a lot of batch processing workloads also. The WSC workloads must be designed to tolerate large numbers of component faults without affecting the overall performance and availability.

**Differences between WSCs and data centers**

| Data Centers | WSCs |
|---|---|
| Datacentreshostsservicesformultiple providers. | WSCsarerunbyonlyoneclient. |
| Therewillbelittlecommonalitybetween software. | Homogenous hardwareandsoftware hardwareand management. |
| Thirdpartysoftwaresolutions. | In-house middleware. |

**WSC are not servers:**

The following features of WSCs make them different from servers:

- **Ample parallelism:**
- Servers need not to worry about the parallelism available in applications to justify the amount of parallel hardware.
- But in WSCs most jobs are totally independent and exploit request-level parallelism.
- **Request-Level parallelism (RLP)** is a way of representing tasks which are set of requests which are to be to run in parallel.
- Interactive internet service applications, the workload consists of independent requests of millions of users.
- Also, the data of many batch applications can be processed in independent chunks, exploiting data-level parallelism.
- **Operational costs count:**
- Server architects normally design systems for peak performance within a cost budget.
- Power concerns are not too much as long as the cooling requirements are

maintained. The operational costs are ignored.

- WSCs, however, have a longer life times and the building, electrical and cooling costs are very high.
- So, the operational costs cannot be ignored. A
- ll these add up to more than 30% of the costs of a WSC in 10 years.
- Power consumption is a primary, not secondary constraint when designing the WSC system.
- **Scale and its opportunities and problems**:
- The WSCs are massive internally, so it gets volume discounts and economy of scale, even if there are not too many WSCs.
- On the other hand, customized hardware for WSCs can be very expensive, particularly if only small numbers are manufactured.
- The economies of scale lead to cloud computing, since the lower per-unit costs of WSCs lead to lower rental rates.
- Even if a server had a Mean Time To Failure (MTTF) of twenty five years, the WSC architect should design for five server failures per day.

**Architecture of WSC**

The height of the servers is measured by **rack units.** A typical rack is 42 rack units.

But the standard dimension to hold the servers is 48.26 cm.

**1 rack unit (U)=1.75 inches or 44.45 mm.**

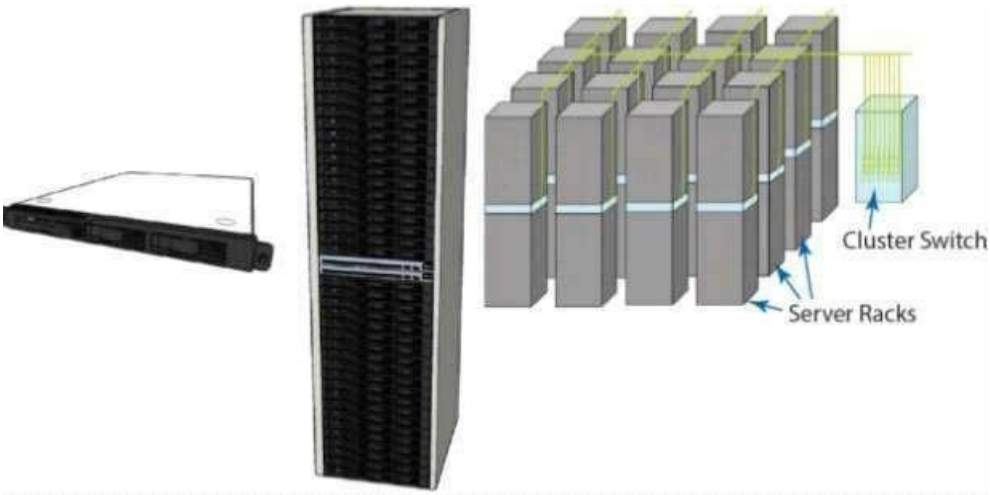


**Fig 1: Architecture of WSCs**

***Source: Miles J. Murdocca and Vincent P. Heuring, — "Computer Architecture and Organization: An Integrated approach"***

The fig 1 shows a WSC system with 1Unit server, 7 inch rack with an Ethernet switch. This figure shows a high end server. But low end servers are of 1U size mounted within a rack and connected with Ethernet switch. These rack level switches use 1 or 10 Gbps links with a number of uplink connections to cluster level switches. The second level switching can span more than 10,000 individual servers.

**Programming model for WSC**

There is a high variability in performance between the different WSC servers because of:

- varying load on servers
- file may or may not be in a file cache
- distance over network can vary
- hardware anomalies

A WSC will start backup executions on other nodes when tasks have not yet completed and take the result that finishes first. Rely on data replication to help with read performance and availability. A WSC also has to cope with variability in load. Often WSC services are performed with in-house software to reduce costs and optimize for performance.

**Storage of WSC**

- A WSC uses local disks inside the servers as opposed to network attached storage (NAS). The Google file system (GFS) uses local disks and maintains at least three replicas to improve dependability by covering not only disk failures, but also power failures to a rack or a cluster of racks by placing the replicas on different clusters.
- A read is serviced by one of the three replicas, but a write has to go to all three replicas.
- Google uses a relaxed consistency model in that all three replicas have to eventually match, but not all at the same time.

**WSC networking**

- A WSC uses a hierarchy of networks for interconnection.
- The standard rack holds 48 servers connected by a 48-port Ethernet switch. A rack switch has 2 to 8 uplinks to a higher switch.
- So the bandwidth leaving the rack is 6 (48/8) to 24 (48/2) times less than the bandwidth within a rack.
- There are array switches that are more expensive to allow higher connectivity.
- There may also be Layer 3 routers to connect the arrays together and to the Internet.
- The goal of the software is to maximize locality of communication relative to the rack.

**Performance**

Power Utilization Effectiveness (PUE) is widely used metric to estimate the performance of WSCs.

$$\text{PUE} = \frac{Total\_utility\_power}{IT\_equipment\_power}$$

Bandwidth is an important metric as there may be many simultaneous user requests or metadata generation batch jobs. Latency is also equally important metric as it is seen by users when they make requests. Users will use a search engine less as the response time increases. Also users are more productive in responding to interactive information when the system response time is faster as they are less distracted.