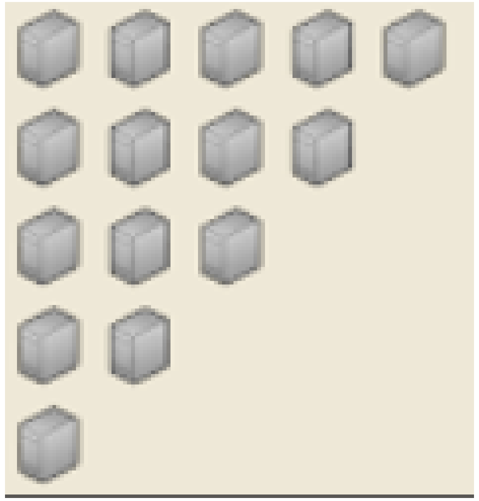
Databases

Hadoop & the Cloud

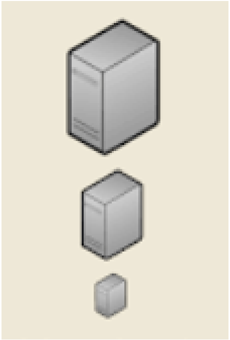
Dr. Roger Young

Distributed Systems

Horizontal Scaling



Vertical Scaling



Horizontal scaling means that you scale by adding more machines into your pool of resources where Vertical scaling means that you scale by adding more power (CPU, RAM) to your existing machine.

Classification of Distributed Systems

- These can be classified into 4 groups:
- Peer-to-peer networks
- Clusters
- Grids (We wont Cover)
- Clouds.

Peer-to-Peer Networks

- In a P2P network, every node (peer) acts as both a client and server. Peers act autonomously to join or leave the network. No central coordination or central database is needed. No peer machine has a global view of the entire P2P system. The system is self-organizing with distributed control.

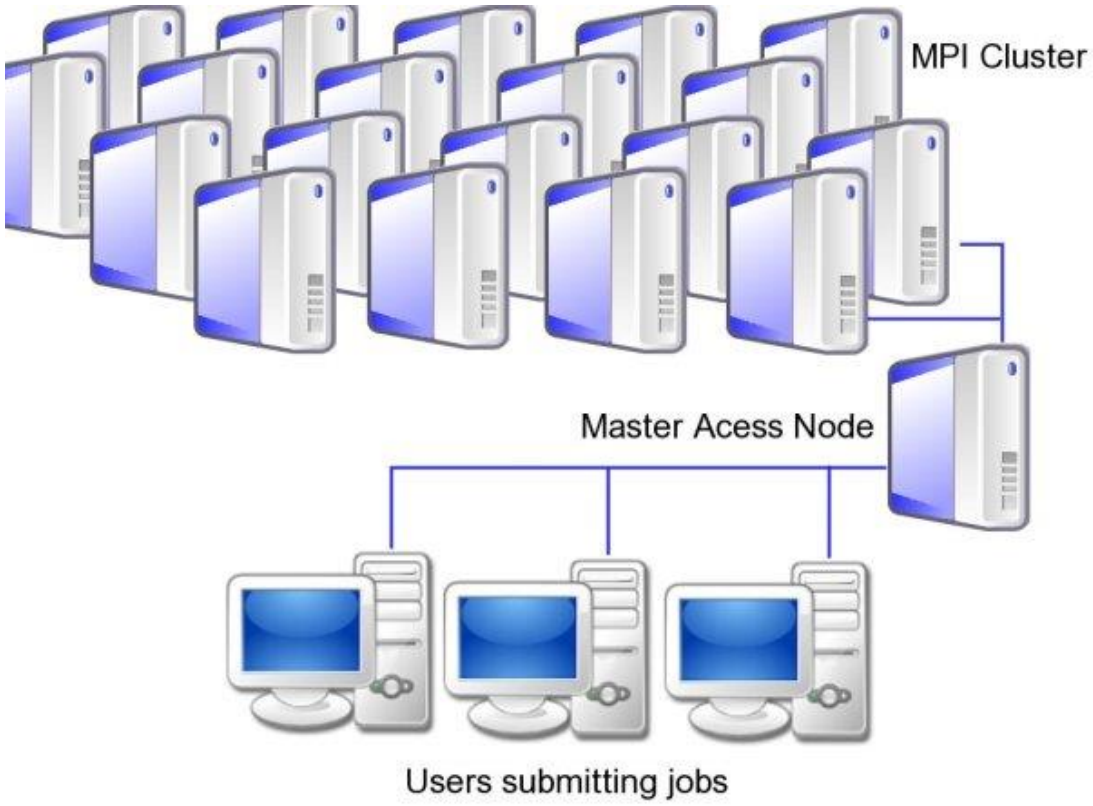
- Unlike the cluster or grid, a P2P network does not use dedicated interconnection network.

- P2P Networks are classified into different groups:

  Distributed File Sharing: content distribution of MP3 music, video, etc. E.g. Gnutella, Napster, BitTorrent.

  Collaboration P2P networks: Skype chatting, instant messaging, gaming etc.

Clusters



- A <u>computing cluster</u> consists of interconnected stand-alone computers which work cooperatively as a single integrated computing resource.
- The network of compute nodes are connected by LAN/SAN and are typically homogeneous with distributed control running Unix/Linux.

Hadoop Overview

- Hadoop is a large-scale distributed batch processing infrastructure.
- While it can be used on a single machine, its true power lies in its ability to scale to hundreds or thousands of computers, each with several processor cores.
- Hadoop is also designed to efficiently distribute large amounts of work across a set of machines.

How large an amount of work?

- Hundreds of gigabytes of data constitute the *low end* of Hadoop-scale.

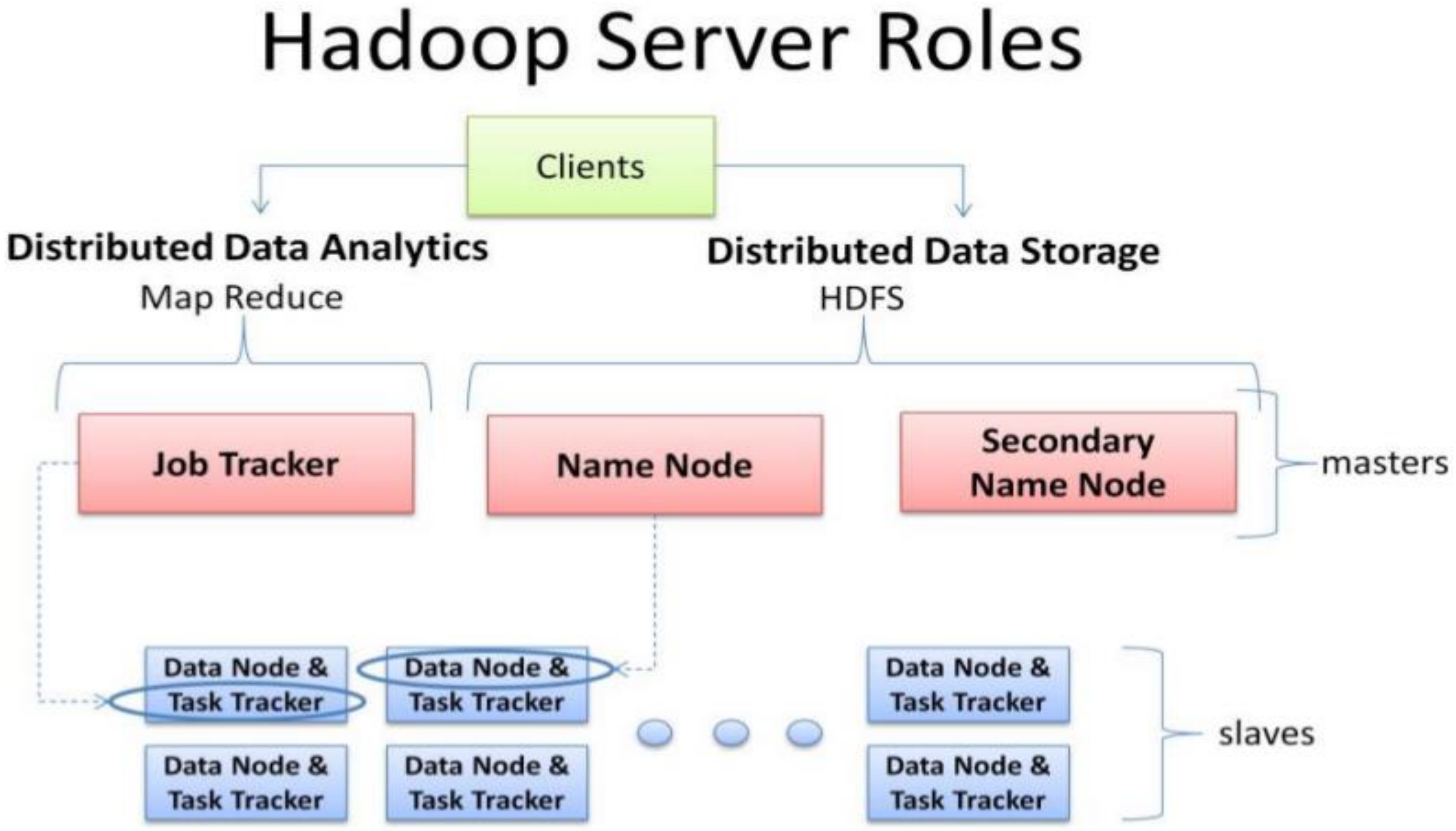
- Actually Hadoop is built to process "web-scale" data on the order of hundreds of gigabytes to terabytes or petabytes.

- At this scale, it is likely that the input data set will not even fit on a single computer's hard drive, much less in memory.

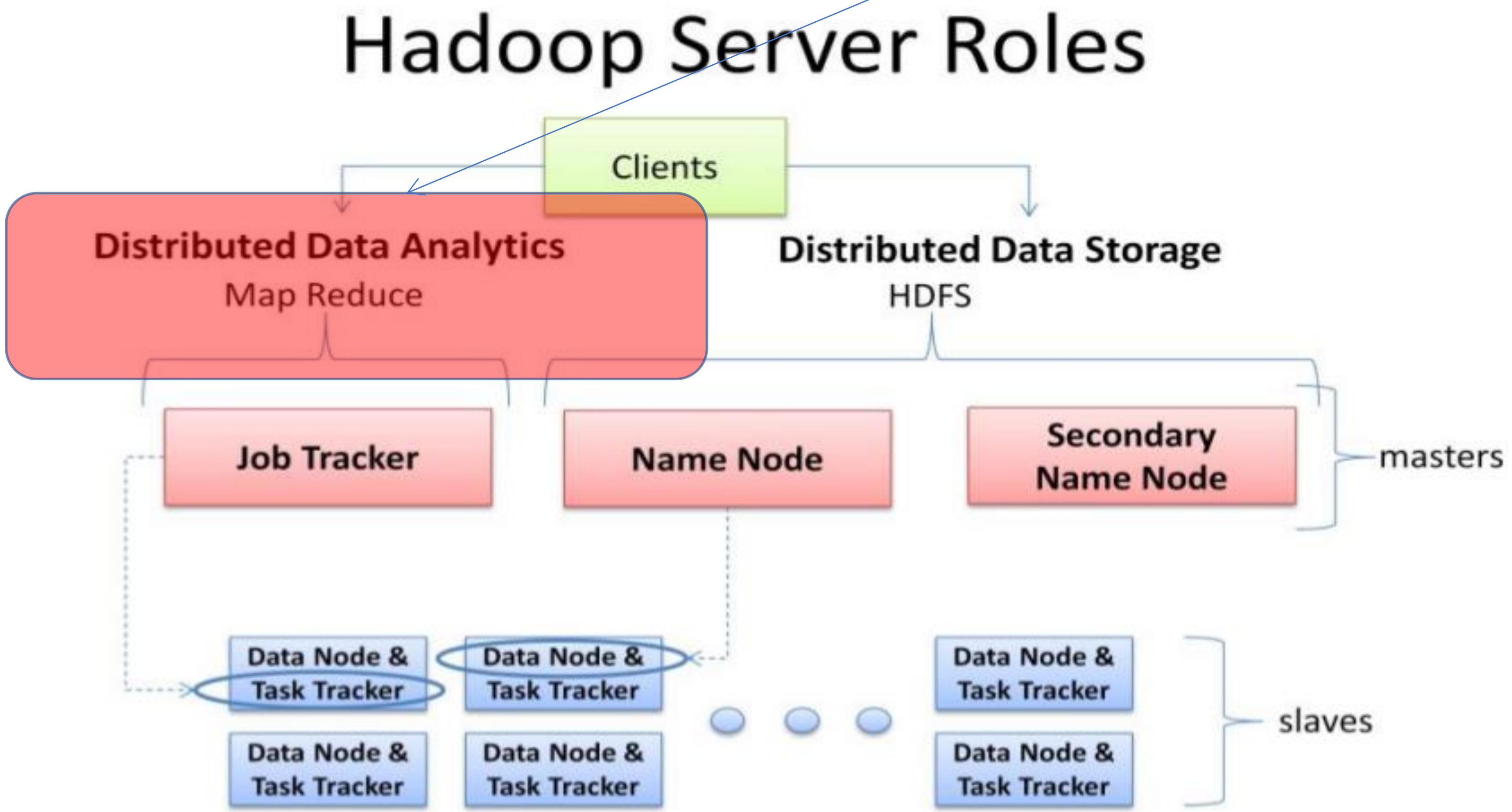
- So Hadoop includes a distributed file system which breaks up input data and sends fractions of the original data to several machines in your cluster to hold.

- This results in the problem being processed in parallel using all of the machines in the cluster and computes output results as efficiently as possible.

The Apache Hadoop Software Library



The Apache Hadoop Software Library – Map Reduce

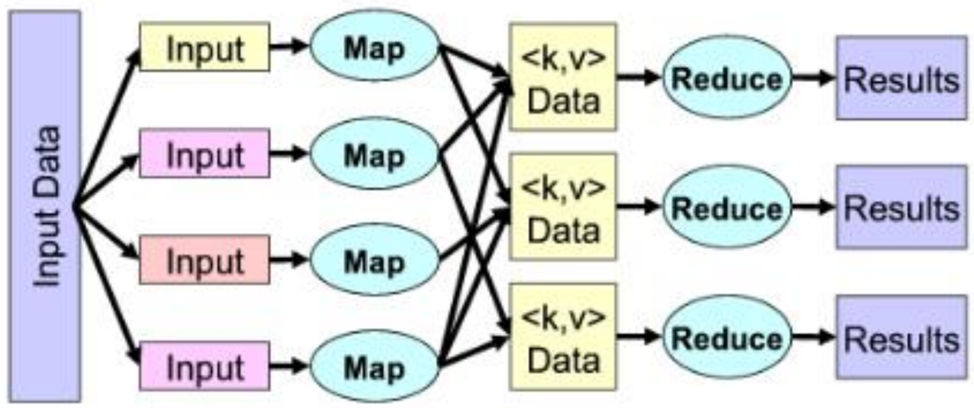


MapReduce

- **MapReduce** is a programming model for processing and generating large data sets with a parallel, distributed algorithm on a cluster
- A **Map()** procedure performs filtering and sorting e.g. sorting students by first name into queues, one queue for each name
- A **Reduce()** procedure performs a summary operation e.g. counting the number of students in each queue, yielding name frequencies

How MapReduce Works

- **Mappers** traverse the data and extracts what you are interested in
- This interesting data is then dealt like cards to the reducers
- Each reducer get its own set of cards
- **Reducers** aggregate or sum the data



How MapReduce Works - Example

- **Challenge**:
- Count the occurrence of particular words on the internet
  e.g. “**the**” “**and**” “**Athlone**”
- Write 2 functions (map and reduce)
  - Mappers traverses all words and tags them
  - Reducers then counts the number of occurrences of each word

- Map gets as input a web page and puts “1” beside every word
- Map just splits the page into words consisting of key/value pairs
- Example:

Key=”**the**” Values = “1” “1” “1”

Key =“**Athlone**” values=“1”

Key=“**and**” values = “1” “1” “1” “1” “1”

Final results

- Reducer 1: (“the", [1, 1,1]) -> (“the", 3)
- Reducer 2: (“Athlone", [1]) -> (“Athlone", 1)
- Reducer 3: (“and", [1,1,1,1,1]) -> (“and", 5)

Final Result: [(“the", 3), (“Athlone", 1), (“and", 5)]

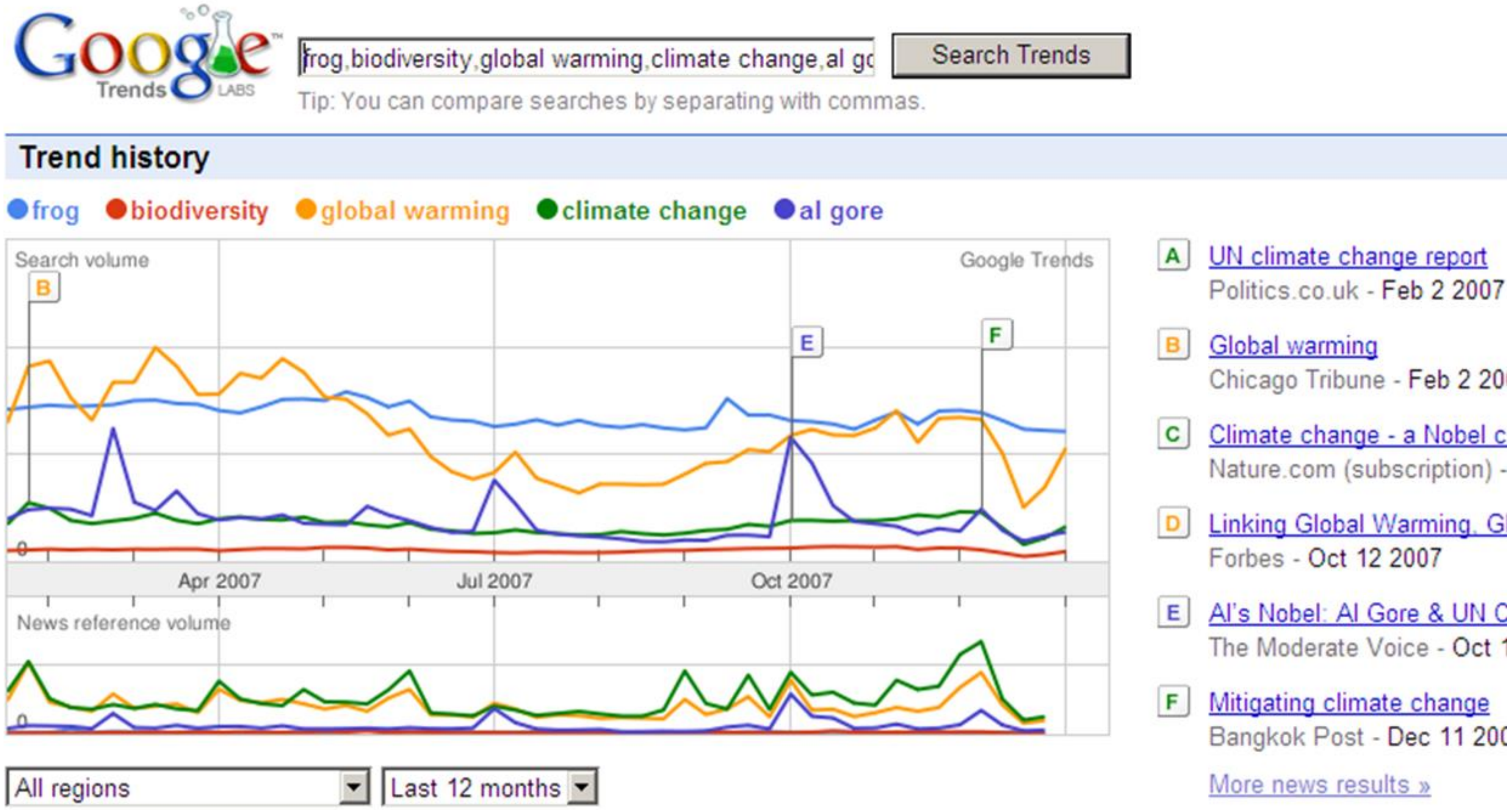
MapReducer Function: Google Trends

- Google uses MapReduce for applications such as Google Trends which shows the number of times particular words were searched for.

- The input to the MapReduce are the Google query logs.

- Mappers split up the logs into word counts and reducers then count the occurrence of each word.

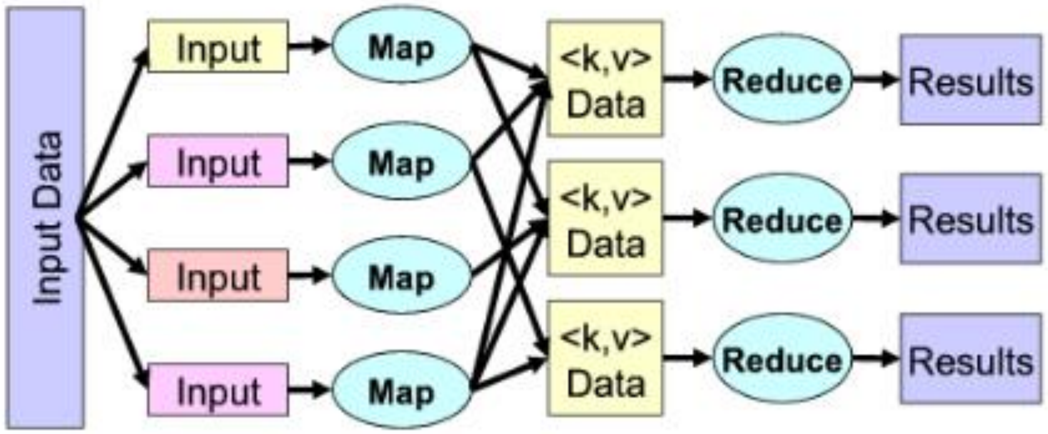
MapReducer Function: Google Trends



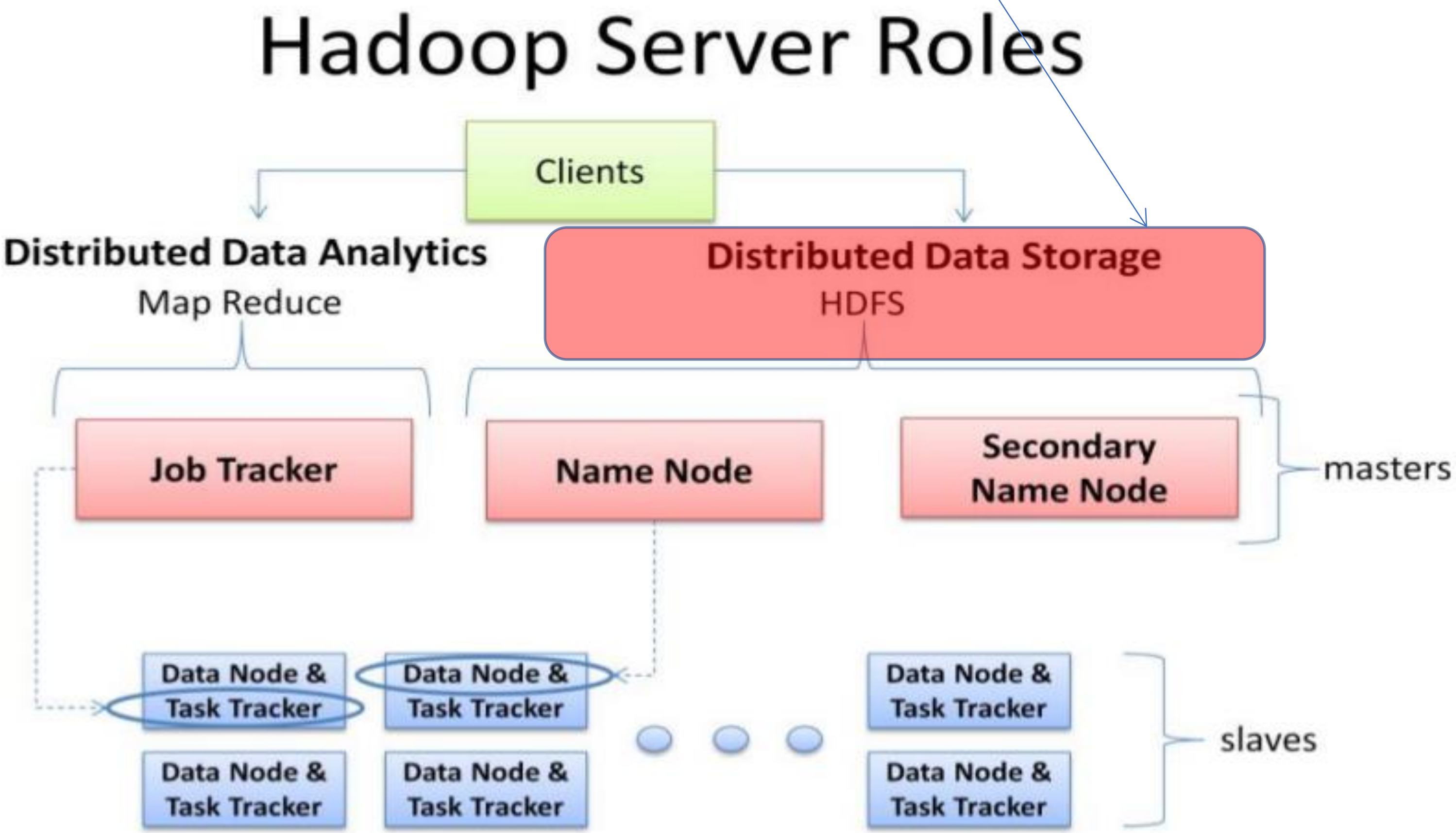
- Application: Show the number of times words were searched for

MapReduce ctd

- The programmer writes the map and reduce functions.
- Hadoop will then automatically handle the code distribution, data distribution, parallel processing, worker scheduling, worker failures etc.



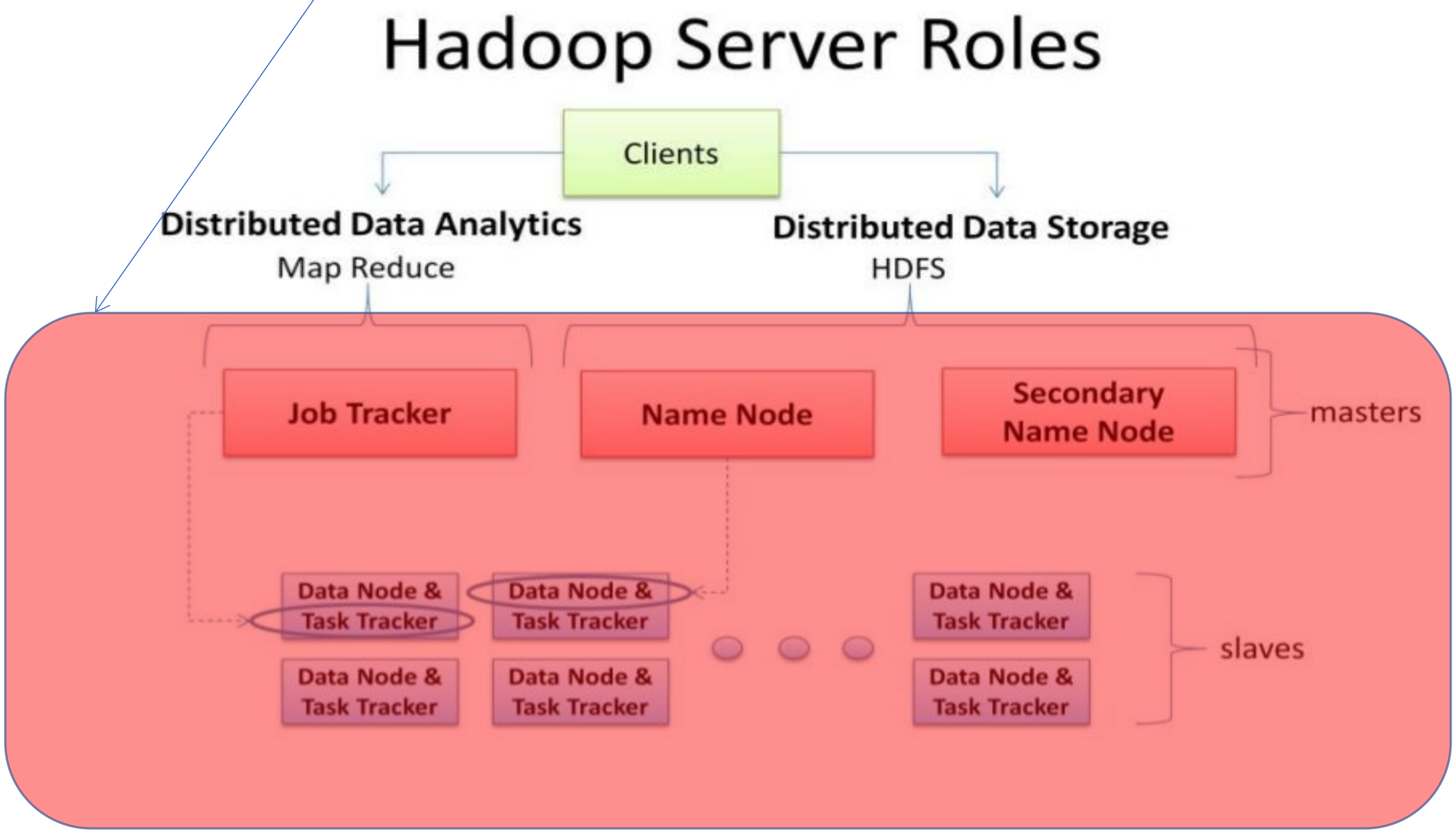
The Apache Hadoop Software Library – Distributed Data Storage



Hadoop Data Distribution

- In a Hadoop cluster, data is distributed to all the nodes of the cluster as it is being loaded in.
- The Hadoop Distributed File System (HDFS) will split large data files into chunks which are managed by different nodes in the cluster.
- In addition to this each chunk is replicated across several machines, so that a single machine failure does not result in any data being unavailable.
- An active monitoring system then re-replicates the data in response to system failures

The Apache Hadoop Software Library – Node Types



The Apache Hadoop Software Library

- “Running Hadoop” means running a set of daemons, or resident programs, on the different servers in your network.

- A basic distributed Hadoop cluster consists of:
  - The NameNode & JobTracker daemons;
  - Secondary NameNode daemon

  - DataNode & TaskTracker daemons.

- These daemons have specific roles; some exist only on one server, some exist across multiple servers.

The Apache Hadoop Software Library – Master Nodes

- **NameNode** : The NameNode is the master of HDFS that directs the slave DataNode daemons to perform the low-level I/O tasks.

- The NameNode is the bookkeeper of HDFS; it keeps track of how your files are broken down into file blocks, which nodes store those blocks, and the overall health of the distributed filesystem. This is the single point of failure in a Hadoop cluster: its failure will impact the overall.

- **Secondary NameNode** : The Secondary NameNode (SNN) is an assistant daemon for monitoring the state of the cluster HDFS. It communicates with the NameNode to take snapshots of the HDFS metadata at intervals defined by the cluster configuration.

The Apache Hadoop Software Library – Master Nodes

- **JobTracker** : The JobTracker daemon is the liaison between your application and Hadoop.

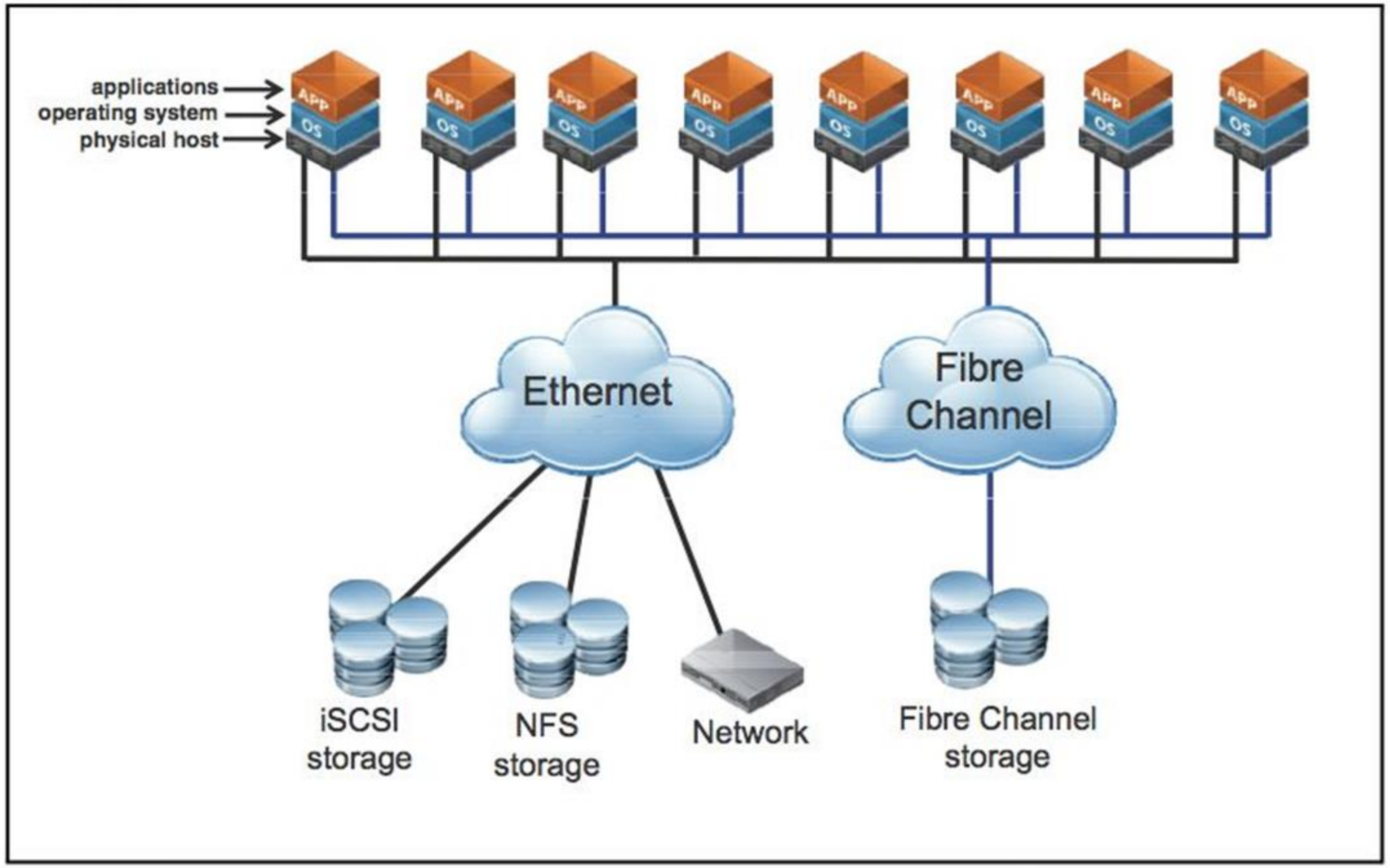
- Once you submit your code to your cluster, the JobTracker **determines the execution plan** by determining which files to process, assigns nodes to different tasks, and monitors all tasks as they’re running.

The Apache Hadoop Software Library – Slave Nodes

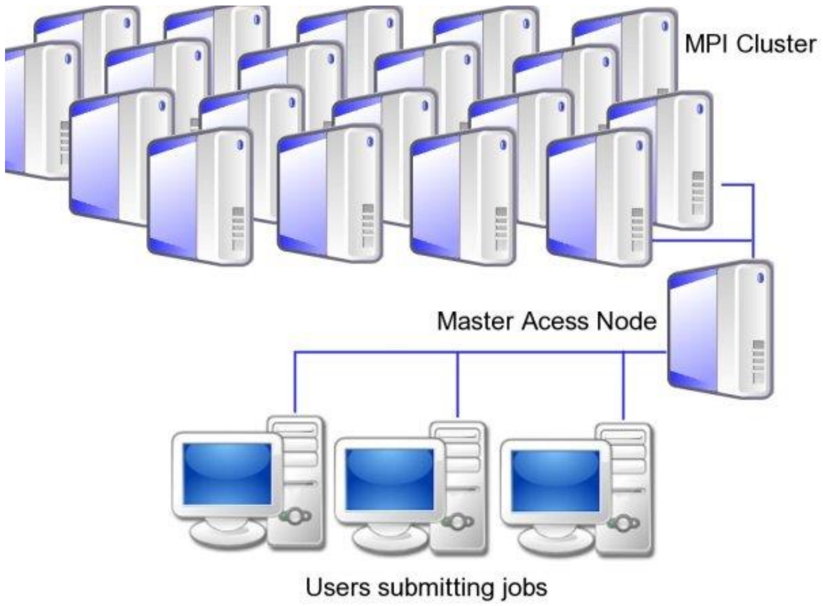
- **DataNode** : Each slave machine in the cluster will host a DataNode daemon to perform the work of the distributed filesystem—reading and writing HDFS blocks to actual files on the local filesystem.

- **TaskTracker** : Each TaskTracker is responsible for executing the individual tasks that the JobTracker assigns. There is a single TaskTracker per each slave node.

Traditional Datacentres



Clusters



- A computing cluster consists of interconnected stand-alone computers which work cooperatively as a single integrated computing resource.
- The network of compute nodes are connected by LAN/SAN and are typically homogeneous with distributed control running Unix/Linux.

The Cost of Searching - “You don’t Buy the Best Machine” for a cluster

- Google's goal is to keep data processing per dollar to a minimum
- For this reason “standard” computers are used

- A typical machine configuration would have
  - 2-4 CPU
  - 1-6 attached disks
  - 4-16GB ram

- An average computer has a usage expectancy of 3 years.
- In a cluster of 10,000 computers you can expect 3 breakages per day.

- You’re system is running on thousands of machines you need a approach which will accept that failures occur

Dealing with Computer Failure

- If a machine is dead it is dead
- The worst situation is where a machine responds to pings from the master but is very slow completing the final reducer in a job.
- 1999 machines could have processed 199,999 records but you are still waiting for the final reducer to finish the job.
- A refinement of the master is to kill off poorly operating machines

Running Traditional Datacentres

- Traditionally, operating systems and software run on a physical computer.

- Several challenges exist to running a large number of physical servers in a datacentre. The model is not flexible and can be inefficient.

- The planning and cost of proper infrastructure (square footage, rack space, power, cooling, cabling, and server provisioning) are but a few of the problems that IT staff must address.

Traditional Datacentres

- Provisioning physical servers is a time consuming process.

- In non-virtualized environments time must be allotted to procure new hardware, place it in the datacentre, install an operating system, patch the operating system and install and configure the required applications can take weeks.

- This process also includes a myriad of other tasks to integrate the system into the infrastructure. For example, configuring firewall rules, enabling switch ports and provisioning storage.

Data Centre Heat Production

- Data centres produce massive amounts of heat
- The air conditioning costs for such buildings is very costly
- Irelands temperate climate means that external air can be taken directly to cool such buildings

Data Centre Heat Production

- Where tech companies like Facebook and Google prefer to move their data centers to colder countries to reduce their air conditioning bill, Microsoft has come up with an even better home for data centers while cutting high energy costs for cooling them: **Under the Sea**.

- Here's what Microsoft says:

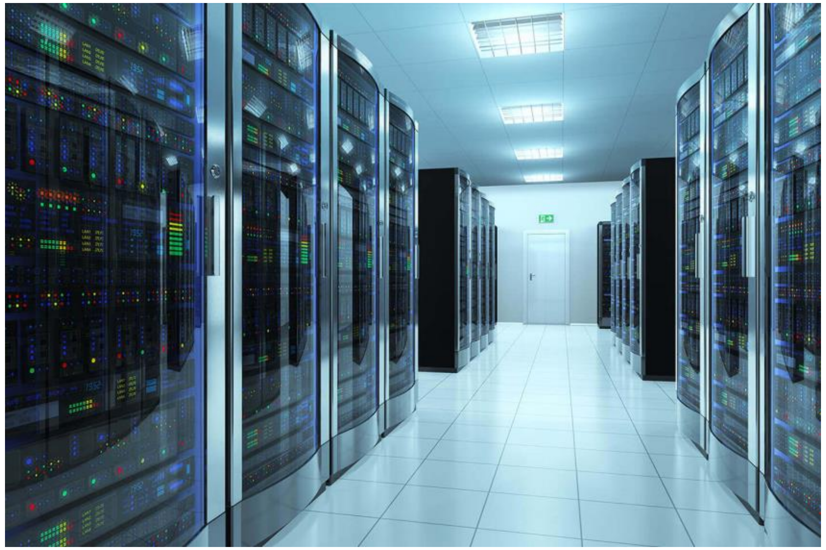
- ***"50% of us live near the coast. Why doesn't our data?"***

- Building massive data centers underwater might sound crazy, but it is exactly something Microsoft is testing with its first submarine data center.

- http://thehackernews.com/2016/02/microsoft-underwater-datacenter.html

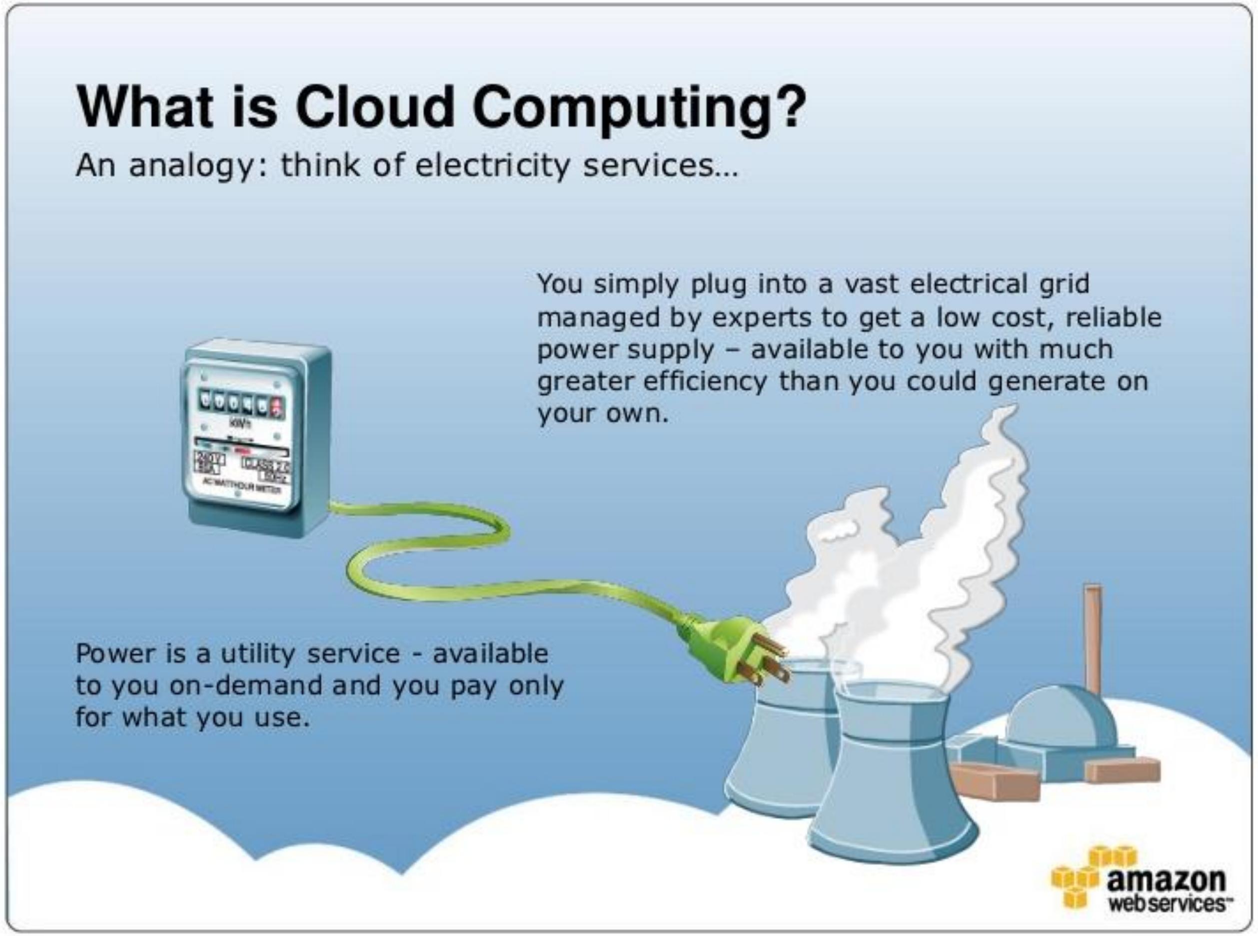
Cloud Computing

What is Cloud Computing?



Definition

- According to the official NIST definition, "cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction


What is Cloud Computing?
An analogy: think of electricity services…
You simply plug into a vast electrical grid managed by experts to get a low cost, reliable power supply – available to you with much greater efficiency than you could generate on your own.
Power is a utility service - available to you on-demand and you pay only for what you use.
kWh
AC WATTHOUR METER
amazon web services

What is Cloud Computing?

Cloud Computing is also a utility service - giving you access to technology resources managed by experts and available on-demand.

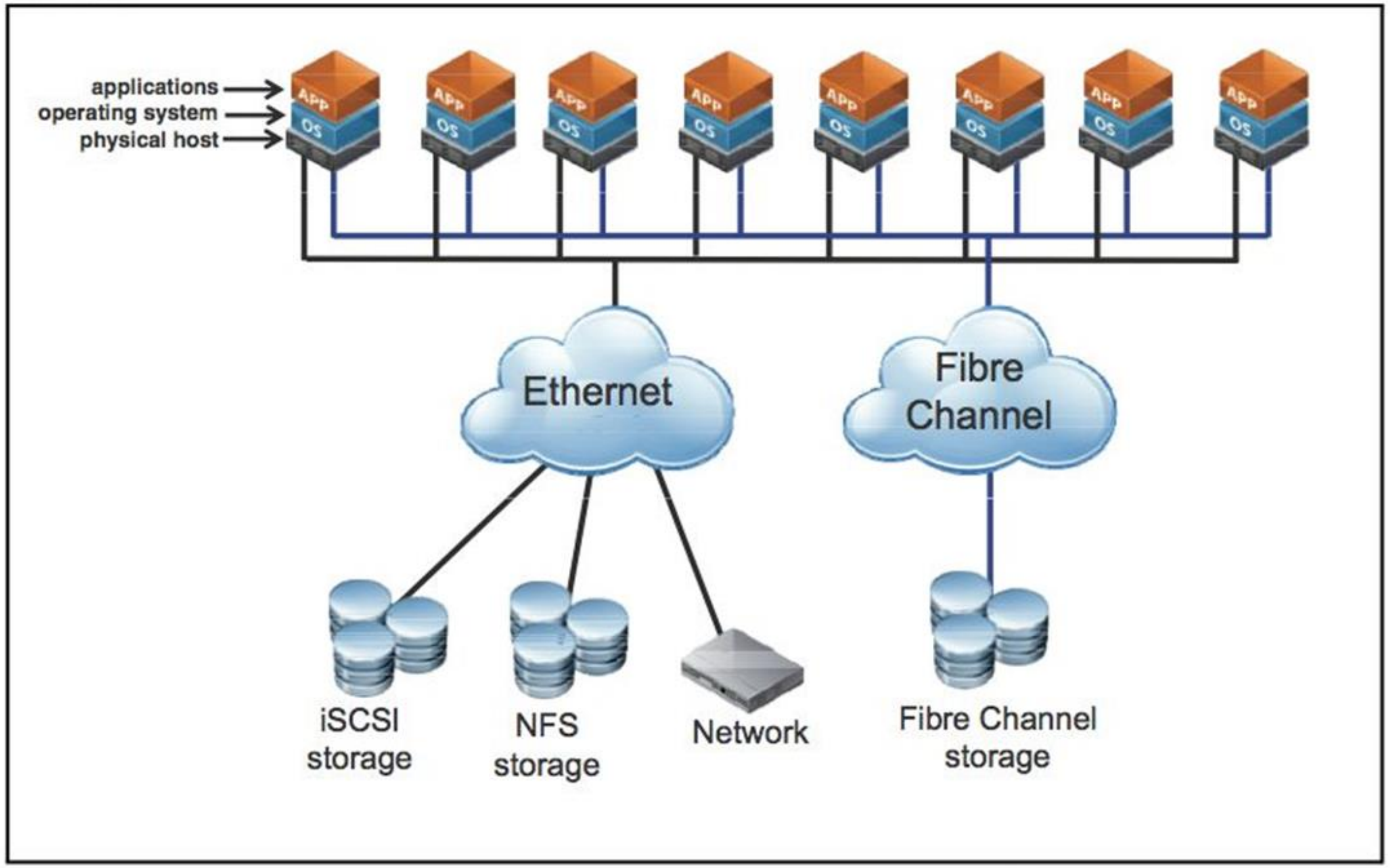
You simply access these services over the internet, with no up-front costs and you pay only for the resources you use.

amazon webservices

Cloud Attributes\Advantages

- No up front capital expenditure (capex)
- Pay as you go and pay only for what you use, billed hourly
- Auto Scaling – true elastic capacity, scale up or down
  - Well suited for applications that experience variability in usage
- Improves time to market
- You get to focus engineers on what differentiates you vs. managing the undifferentiated infrastructure

Physical infrastructure



Traditional Datacentres

- Typically, a 1:1 relationship exists between a physical computer and the software that it runs.
- This relationship leaves most computers vastly underused, leaving between only 5-10% of physical server capacity in use.
- The cost of the space and power required to house, run and keep these systems cool can be expensive.

Virtualization

Although virtualization exists for years, advances in this data centre technology have been significant in the last number of years.

- Many different types of virtualization:
    - Network Virtualization
    - Desktop Virtualization
    - Operating system Virtualization
    - Storage Virtualization
    - Memory Virtualization
    - Server virtualization

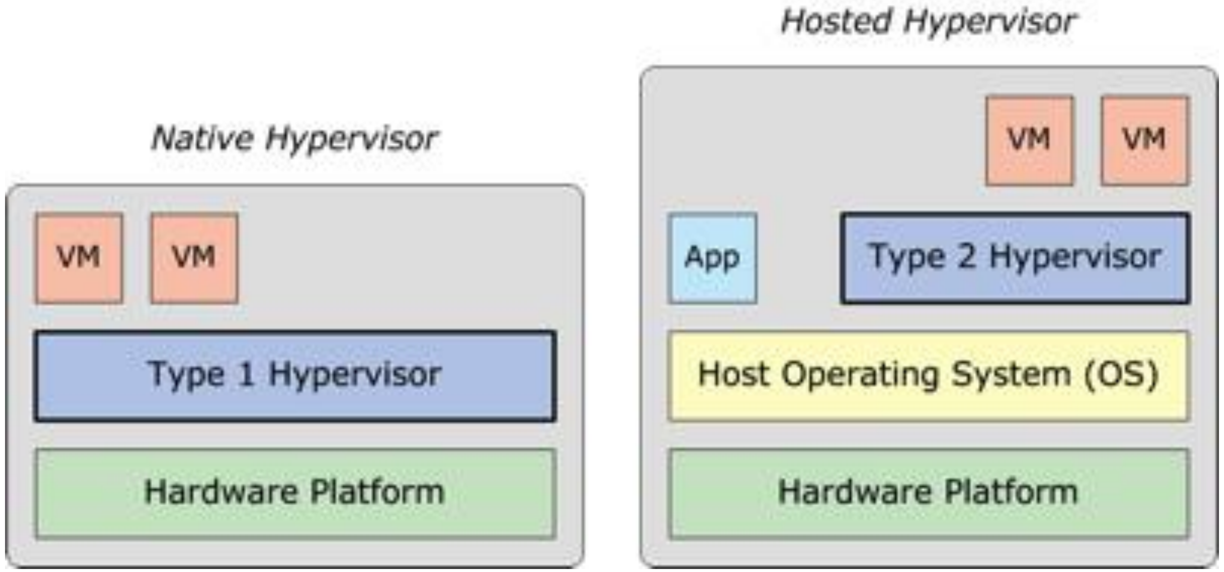
Virtualization

Server Virtualization –It abstracts the underlying physical resources and presents these as a set of virtual machines

There are many kinds of management layers (or hypervisors) that facilitate this abstraction.

- **Type 1** hypervisors such as Microsoft Hyper-V or VMWare ESX run on bare hardware.
- **Type 2** hypervisors such as Virtualbox or VMWare Workstation run on a host operating system.

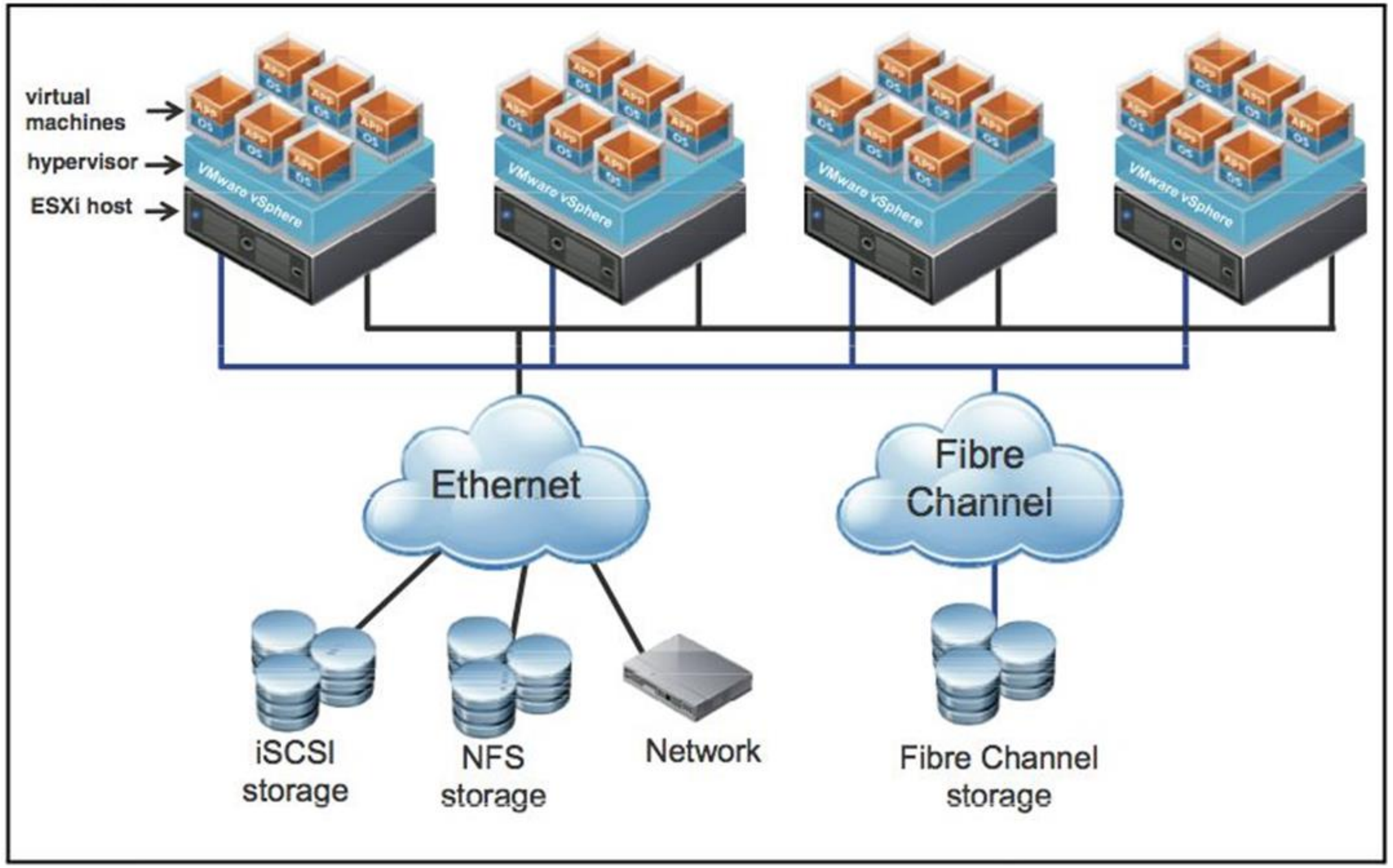
Virtualization



- Two types of hypervisors
  - Type 1 for enterprise solutions (better performance)
  - Type 2 small enterprise, individual

- **Type 1** hypervisors such as Microsoft Hyper-V or VMWare ESX run on bare hardware.
- **Type 2** hypervisors such as Microsoft Virtual Server or VMWare Workstation run on a host operating system.

Virtual Infrastructure



Benefits of Virtualization

- Virtualization enables you to run more workloads on a single server by consolidating the environment so that your applications run on virtual machines.
- Converting to a virtualized datacentre reduces the required datacentre square footage, rack space, power, cooling, cabling, storage and network components by reducing the sheer number of physical machines.
- The reduction of physical machines can be realized by converting physical machines to virtual machines and consolidating the converted machines onto a single host.

Physical and virtual infrastructure