Set A

Countineous Assessment 2

CAP 470

Course Title: Cloud Computing

Name : Jayshri Lal Pandit

Section : D2112

Roll No. : RD2112A03

Reg. No. 12111670

(1.) Which platforms are used for large scale cloud computing ? Explain it briefly with example ?

Ans:- The Platforms for large-scale cloud computing are :-

(1.) Apache Hadoop :-

It is a open source framework coritten in java that allows distributed processing of large datasets across clusters of computers using simple programming models. The Hadoop framework application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

Hadoop Distributed File System

The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) and provides a distributed file system that is designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from

others distributed file systems are significant. It is highly fault-tolerant and is designed to be deployed on low-cost hardware. It provides hight throughput access to application data and is suitable for applications having large datasets.

Apart from the above-mentioned two core components, Hadoop framework also includes the following two modules :-

- **Hadoop common :-** These are java libraries and utilities required by other Hadoop modules.
- **Hadoop YARN :-** This is a framework for job scheduling and cluster resource management.

How Does Hadoop Work

It is quite expensive to build bigger servers with heavy configurations that handle large scale processing, but as an alternative, you can tie together many commodity computers with single-CPU, as a single functional distributed system and

Practically, the clustered machines can read the dataset in parallel and provide a much higher throughput. Moreover, it is cheaper than one high-end server. SO this is the first motivational factor behind using Hadoop that it runs across clustered and low-cost machines.

Hadoop runs code across a cluster of computers. This process includes the following core tasks that Hadoop performs :-

- Data is initially divided into directories and files. Files are divided into uniform sized blocks of 128 M and 64 M (preferably 128 M).
- These files are then distributed across various cluster nodes for further processing.
- HDFS, being on top of the local file system, supervises the processing.
- Blocks are replicated for handling hardware failure.
- checking that the code was executed successfully.

- performing the sort that takes place between the map and reduce stages.
- Sending the sorted data to a certain computer.
- Writing the debugging logs for each job.

(2.) Map Reduce :-

MapReduce is a programming model or pattern ~~cost~~ within the Hadoop framework that is used to access big data stored in the Hadoop File System (HDFS). It is a core component, integral to the functioning of the Hadoop framework.

For example, a Hadoop cluster with 20,000 inexpensive commodity servers and 256 MB block of data in each, can process around 5 TB of data at the same time. This reduces the processing time as compared to sequential processing of such a large data set.

With MapReduce, rather than sending data to where the

application or logic resides, the logic is executed on the servers where the data already resides, to expedite processing.

How MapReduce Works

At the crux of MapReduce are two function: Map and Reduce. They are sequenced one after the other.

- The **Map** function takes input from the disk as <Key, Value> pairs, processing them, and produces another set of intermediate <Key, Value> pairs as output.

- The **Reduce** function also takes inputs as <Key, Value> pairs, and produces <Key, Values> pairs as output

The types of Keys and Values differ based on the use case. All inputs and outputs are stored in the HDFS. While the map is a mandatory step to filter and sort the initial data, the reduce function is optional.

(2.) What is meant by Edge computing, and how it is related to cloud ?

Ans:- In the beginning, there was one Big computer. Then, in the Unix era, we learned how to connect to that computer using a dumb terminal. Next we had personal computers, which was the first time regular people really owned the hardware that did the work.

Right Now, in 2018, we're firmly in the cloud computing era. Many of us still own personal computers, but we mostly use them to access centralized services like Dropbox, Gmail, Office 365, and Slack. Additionally, devices like Amazon Echo, Google Chromecast, and the apple TV are powered by content and itelligence that's in the cloud – as opposed to the DVD box set of Little House on the Prairie or CD-ROM copy of Encarta you might've enjoyed in the personal computing era.

As centralized as this all sounds, the truly amazing thing about cloud computing is that a seriously large percentage of all companies in the world now rely on the infrastructure, hosting, machine learning, and compute power of a very select few cloud providers: Amazon, Microsoft, Google, and IBM.

Edge devices can contribute to a cloud, if the storage and computing capabilities provided by those devices at the endpoints of a ~~neut~~ network are abstracted, pooled, and shared across a network - essentially becoming part of a larger cloud infrastructure. Edge computing is not part of a cloud. What makes edge computing so useful is that it is purposefully separate from clouds and cloud computing.

Here's how we see it:

- Clouds are places where data can be stored or applications can run.

They are software - defined environments created by datacenters or server farms.

- Edges are also places where data is collected. They are physical environments made up of hardware outside a datacenter.

- Cloud computing is an act; the act of running workloads in a cloud.

- Edge computing is also an act; the act of running workloads on edge devices.

An edge (location) is not the same thing as edge computing (action). Collecting data at the edge of a network and transferring it to a cloud with minimal (if any) modification is not edge computing - it's just networking. But, if that data is collected and processed at the edge, then it's edge computing.

Edge computing is separate from

clouds 2 main reasons :-

(1) Time sensitivity :- The rate at which a decision needs to be made doesn't allow for the lag that would normally take place as data is collected by an edge device, transferred to a central cloud without modification and then processed before a decision is sent back to the edge device for execution.

(2) Data Volume :- The sheer volume of data collected is too much to send - unaltered - to a cloud.

(3) What is the difference between cloud and traditional datacenters?

Ans:-

Cloud

A cloud can be described as a term used to describe a group of services either a global or individual network of servers, which processes a unique function. Cloud is not a physical entity, but they are a group or network of remote servers which are arched together to operate as

a single entity for an assigned task.

Types of Cloud :-

- **public cloud** : It is a cloud methodology that is open to all with the Internet on pay-per-usage method.
- **private cloud** :- used by organisations to build their data centers that are accessible only with the permission of the organisation.
- **Hybrid cloud** :- that is a combination of public and private clouds. It serves different need of an organisation.
- **Community cloud** :- that provides services to a group of people in an organisation or a single community.

Data Center

A data center can be described as a facility / space of networked computers and associated components (like telecommunications and storage) which helps business and.

Organisations to function a large amount of data These Data centers allow the data to organise, process, store and dissemi upon the application used by businesses.

Types of Data center

- **Telecom Data center** :- It is a type of data center which are operated by telecommunications or service providers. It requires high-speed connectivity to function
- **Enterprise Data center** :- It is a type of data center which is built and owned by a company that may or may not be onsite.
- **Colocation Data Center** :- that consists of one data centere owner place which provides cooling to multiple enterprises and hyper-scale their customers.
- **Hyperscale Data center** :- data Center which are owned by and operated by the company itself.

Differences between cloud and Data centers—

	Cloud	Data Center
(i)	Cloud is virtual resorce that helps businesses to store, organize, and operate data efficiently.	Data center is a physical resource that helps businesses to store, organise, and operate and efficiently.
(ii)	The scalability of the cloud required less amount of investment.	The scalability of Data center is huge in investment as compared to the cloud.
(iii)	The maintenance cost is less than service providers maintain it.	The maintenance cost is high because developers of the organisation do maintenance.
(iv)	Third-party needs to be trusted for the organisation's data to be stored.	The organisation's developers are trusted for the data stored in data centers.
(v)	performance is huge as compared with investment.	performance is less than compared to investment.
(vi)	It requires a plan to customize the Cloud.	It is easily customizable without any hard plan.

(vii)	It requires a stable internet connection to provide the function	it may and may not require an internet connection.
(viii)	Cloud is easy to operate and is considered a viable option.	Data centers require experienced developers to operate and are considered not a viable option.

The End