

Declaration and statement of  
authorship

I, bearing Roll No. 106119100, agree and acknowledge  
that:

- 1) The assessment was answered by me as per instruction applicable to each assessment, and that I have not resorted to any unfair means to deliberately improve my performance.
- 2) I have neither impersonated anyone, nor have been impersonated by any person for the purpose of assessment.

Signature : Rajneesh Pandey

Full Name : Rajneesh Pandey

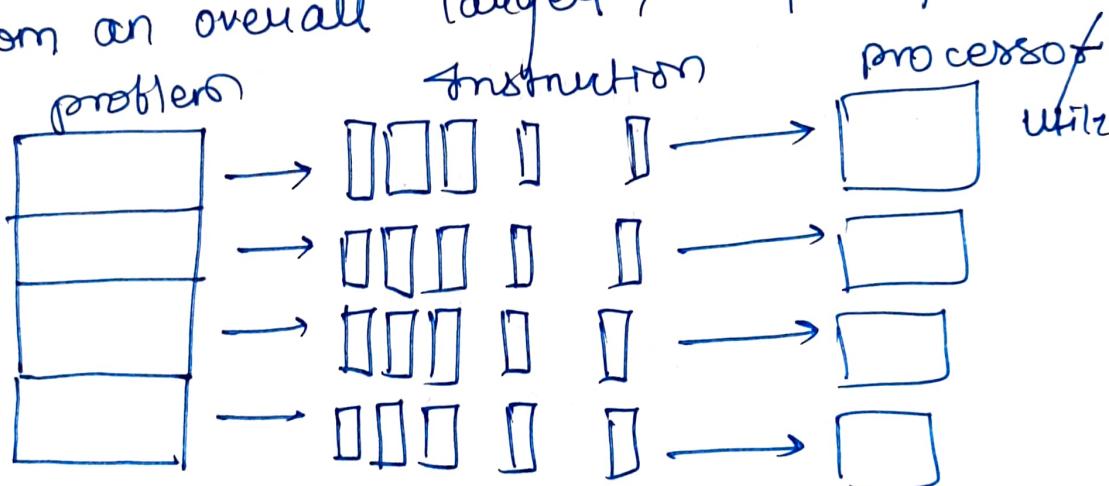
Roll No : 106119100

Sub code : CSPE 56

Mobile No : 8290968008

Cloud ComputingEnd-Semester ExamSET-2Rajneesh  
PandeyQuestion② :

Parallel Data computing in cloud is a computing architecture in which several processors simultaneously execute multiple, smaller calculation broken down from an overall larger, complex problem.



Major cloud computing companies have started to integrate frameworks of parallel data processing in their product portfolio, making it easy for customers to access the services and to deploy their programs. It uses processing frameworks which are currently used stem

from. from the field of cluster computing and disregard the particular nature of cloud.

Hence, the allocated computer resources may be inadequate for big parts of the submitted job and unnecessarily increase processing time and cost meants and demeants.

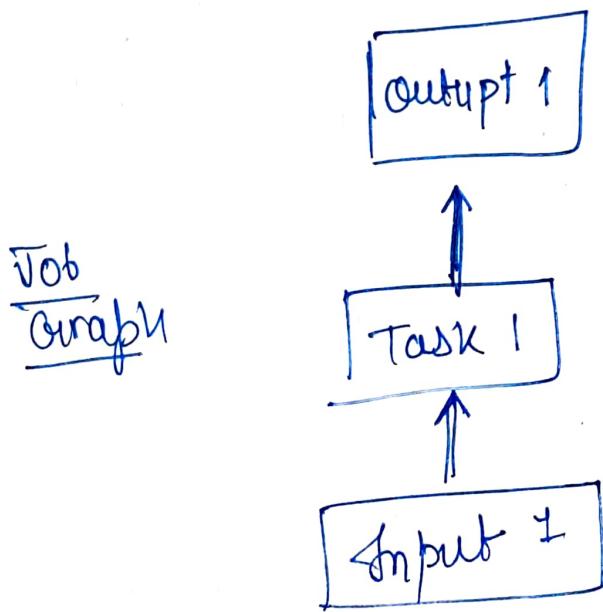
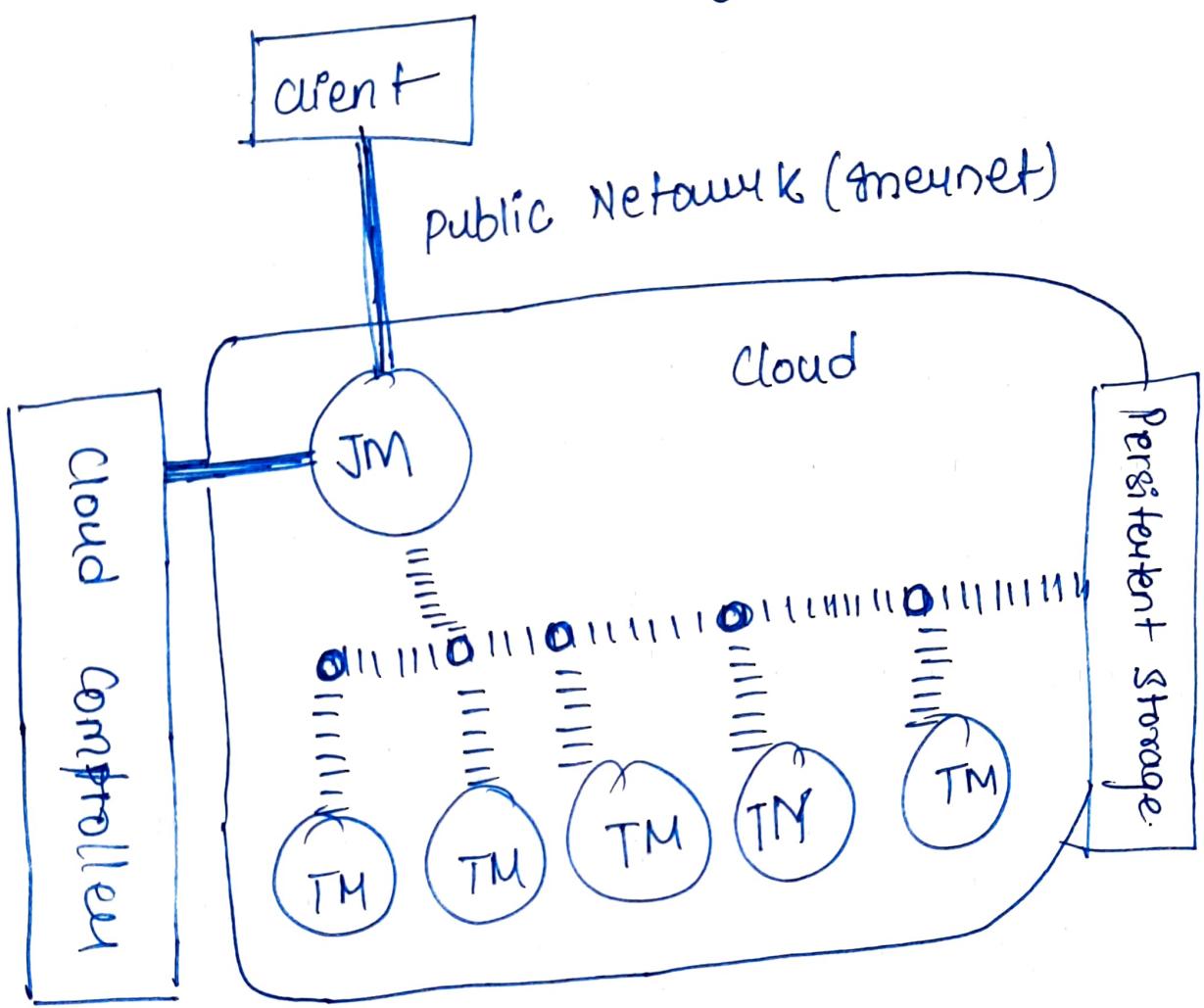
data processing framework to explicitly exploit the dynamic resource allocation offered by today's compute cloud for both, task scheduling and execution.

It allows assigning the particular task of a processing job to different types of virtual machines and takes care of their instantiation and termination during the job execution.

Based on this new framework, we perform evaluation on a compute cloud system and compare the result to the existing data processing

framework Hadoop

overall structure



### Question ③

#### Google file System (GFS)

It was developed in late 1990s, uses thousand of storage system built from inexpensive commodity components. to provide petabytes of storage to a large user community with diverse needs.

#### Motivation for GFS development :

- component failures is the norm
- files are huge
- The most common operation is to append to an existing file.
- the consistency model should be relaxed to simplify the system implementation. but without placing an additional burden on the application developers.

GFS node cluster is a single master with multiple chunk servers that are continuously accessed by different client system. chunk servers store data as linux files on local disks. stored data is divided into large chunks (64MB), which are replicated in the network a minimum of three times. The large chunk size reduces network overhead.

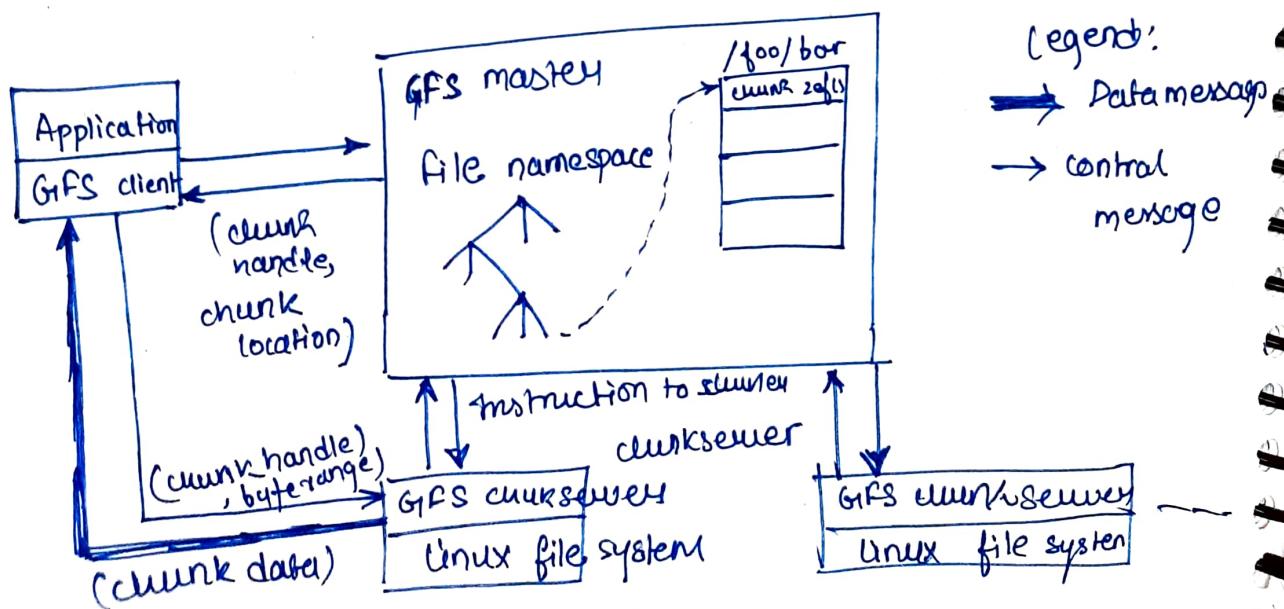
GFS is designed to accommodate Google's large cluster requirements without burdening applications. files are stored in hierarchical directories identified by path names.

## GFS APIs

- It provides a familiar interface though not POSIX.
- Supports : create , delete , open , close , read and write .
- Pins : snapshots and record append.

Snapshots → creates a file copy or a directory tree at a low cost.

record append → allows multiple client to append data to the same file concurrently while guaranteeing atomicity .



## Advantages of GFS :

- 1) very high availability and fault tolerance.  
replication:
  - a) chunk and master replication.
  - b) chunk and master recovery.
- 2) simple and efficient centralized design with a single master.  
i.e. large sequential reads.
- 3) concurrent writes to the same file region are not serializable. Thus, replicas might have duplicates but there is no interleaving of records. To ensure data integrity of its own copy using checksum.
- 4) Read operation span at least a few 64 kB blocks therefore the check summing cost
- 5) Bulk operation ensures no synchronization is needed at client end.
- 6) No caching eliminates cache coherence issues
- 7) GFS master constantly monitors each chunks through heartbeat message.

## Question 4

energy efficiency in cloud-computing:

Cloud computing provides various services based on internet computing.

It also provides computing resources through virtualization over internet.

Data center is the most prominent in cloud computing which contains collections of servers on which business information is stored and application runs.

Now, these centre, like consist of  $\text{CO}_2$ , releasing sources, like cable, air conditioner, network etc. - - consumes more power and releases huge amount of carbon dioxide ( $\text{CO}_2$ ).

Hence concept of energy utilization comes into picture.

Techniques used to apply it:

- 1) Dynamic Voltage & Frequency Scaling (DVFS)
- 2) Virtual Machine (VM)
- 3) Migration and VM consolidation.

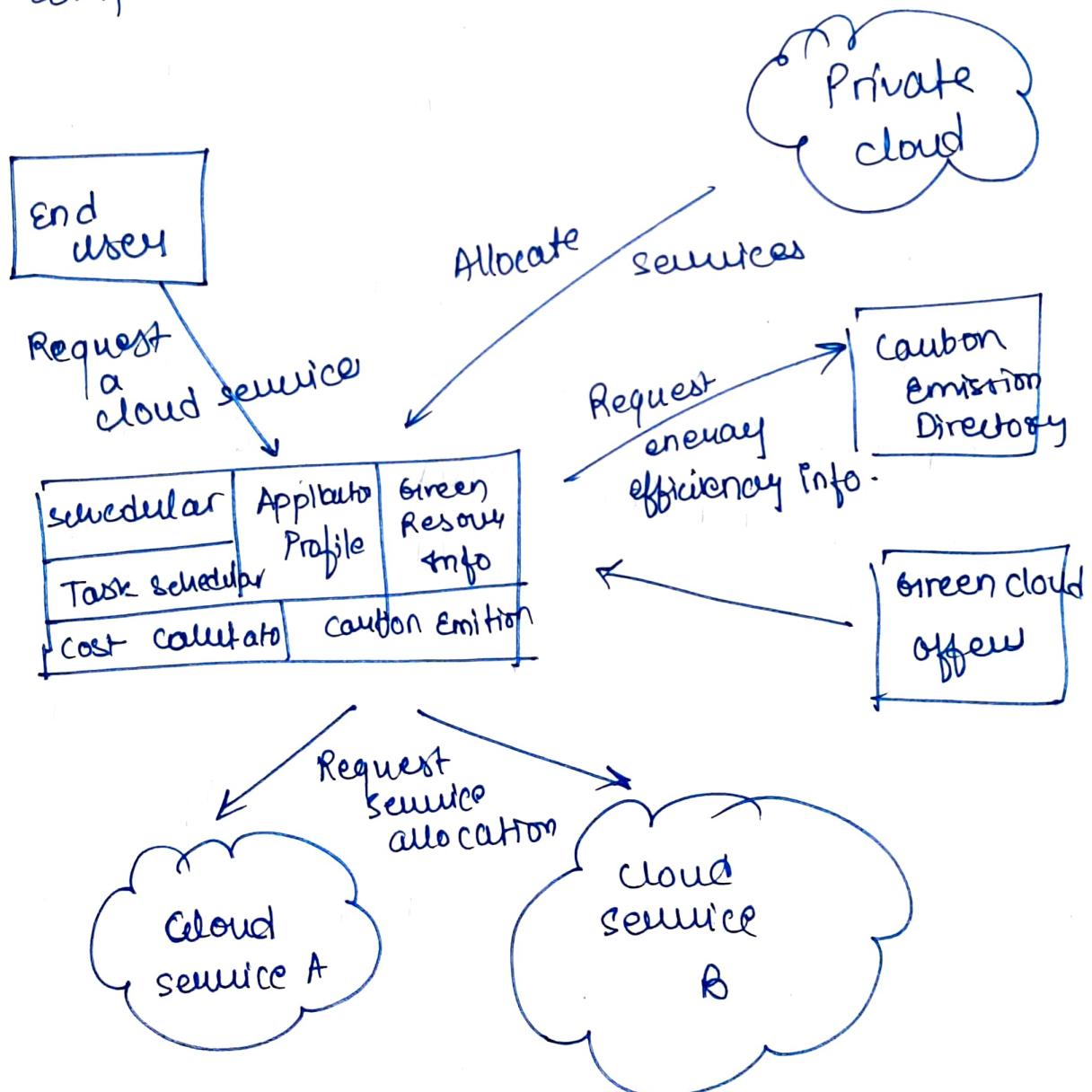
Also, the

Algorithms are:

1. Maximum Bin Packing
2. Power Emand Min-Max and Minimization Migrations.
3. Highest Potential growth.

Green computing:

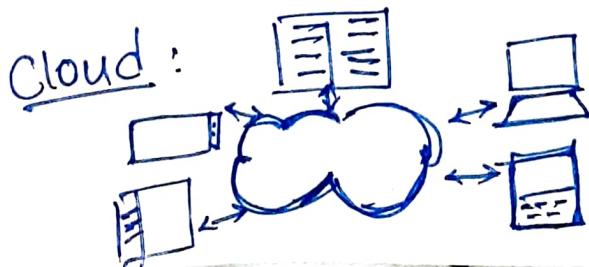
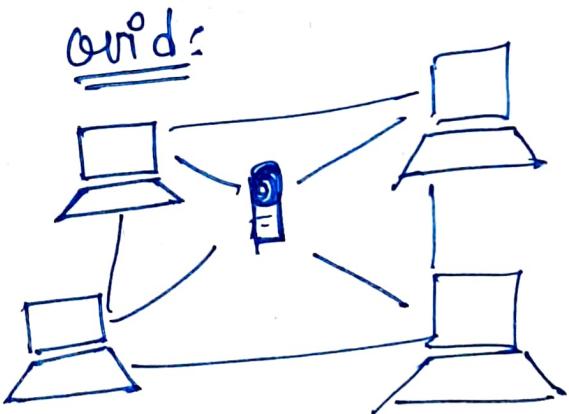
Green computing is the eco-friendly use of computers and their resources.



Architecture-

## Similarities between Grid & Cloud

- Grid computing are scalable as well as cloud computing is also scalable.
- SaaS (Software as a service), PaaS , All these technology are there in both of them i.e Grid and cloud.
- Grid computing is used for high intensive computation which involves number crunching and cloud computing is used for requirements where you require,
  - High CPU usage
  - High storage
  - High Bandwidth.



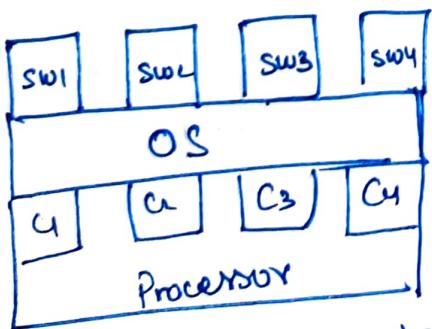
## Question 5

Multicore processing, and implementation of virtualization.

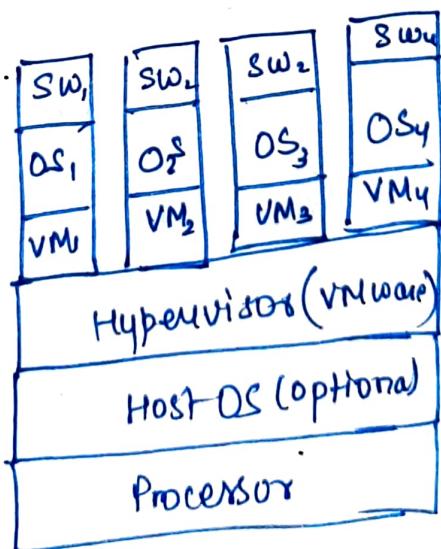
The implementation of virtualization is a combination of software and hardware features that creates virtual CPUs (VCPUs) or virtual systems on-chip (VSOC).

Virtualization helps to run multiple operating systems on a single computer.

Incorporation of multicore processing in virtualization is with the help of hardware software combination.



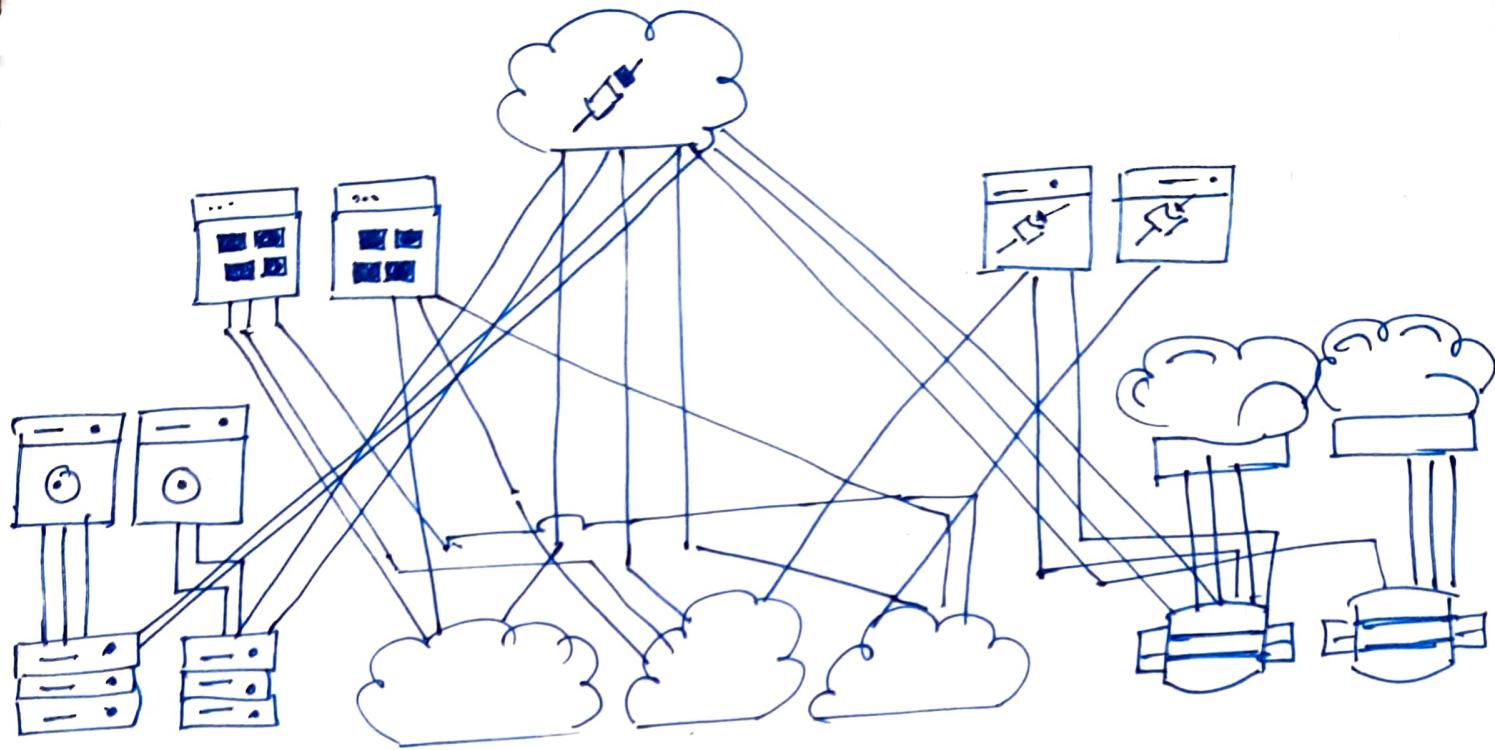
Multicore processing (MCP).



Virtualization via VMs.



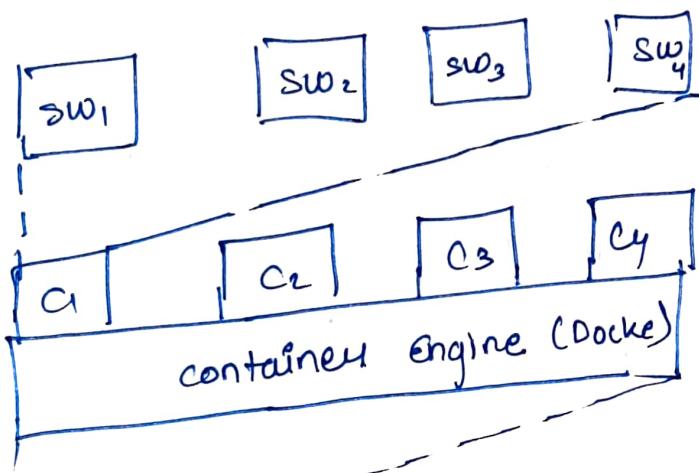
Virtualization via containers.



It's the property that facilitates unrestricted sharing and use of data or resources between disparate systems via local area networks (LANs) or wide area networks (WANs).

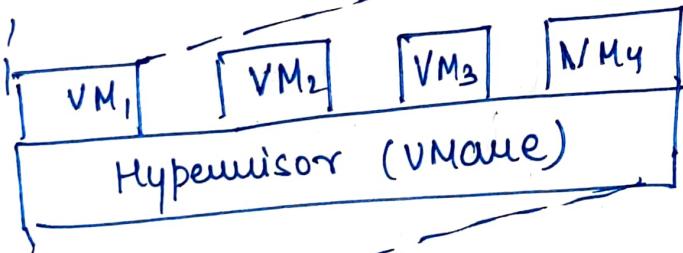
There are two types of data interoperability - syntactic interoperability, which is a prerequisite to semantic interoperability and enables different software components to cooperate, facilitating two or more systems to communicate and exchange data, and semantic interoperability, which refers to the ability of computer systems to exchange meaningful data with unambiguous, shared meaning.

- multiple software application can be allocated to a single container.
- multiple software application can be allocated to a single virtual machine.
- multiple virtual machine can be allocated to a single core.
- multiple core are contained by a single multicore processor.

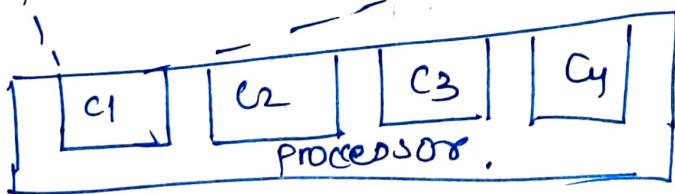


multiple software Application

virtualization via  
( container  
multiple virtual OS )



virtualization via  
( VMs )  
( Multiple VMMs )



multicore Processing  
( MCP )  
( Multiple Actual Cores )

challenges.

Multicore and virtualization add additional complexity to the architecture, and also better performance safety and security.

Google File Systems. →

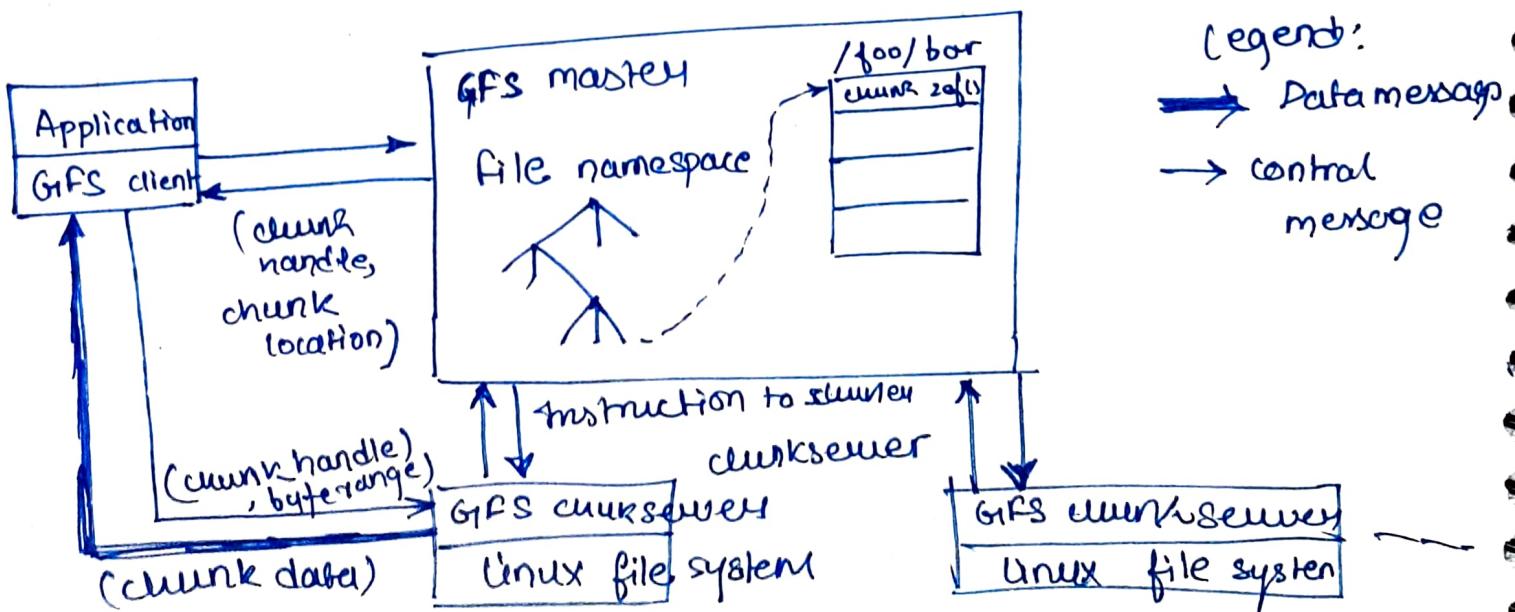
GFS is designed to accommodate Google's large cluster requirements without burdening applications. files are stored in hierarchical directories identified by path names.

### GFS APIs

- ff provides a familiar interface though not POSIX.
- supports: create , delete , open , close , read and write .
- Pins : snapshots and record append.

Snapshots → creates a file copy or a directory tree at a low cost!

record append → allows multiple client to append data to the same file concurrently while guaranteeing atomicity.



Question 6:

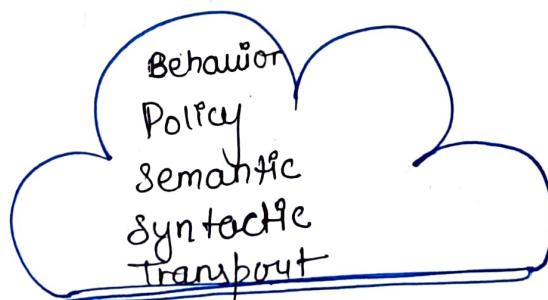
106119100

### : Inter - Operability:

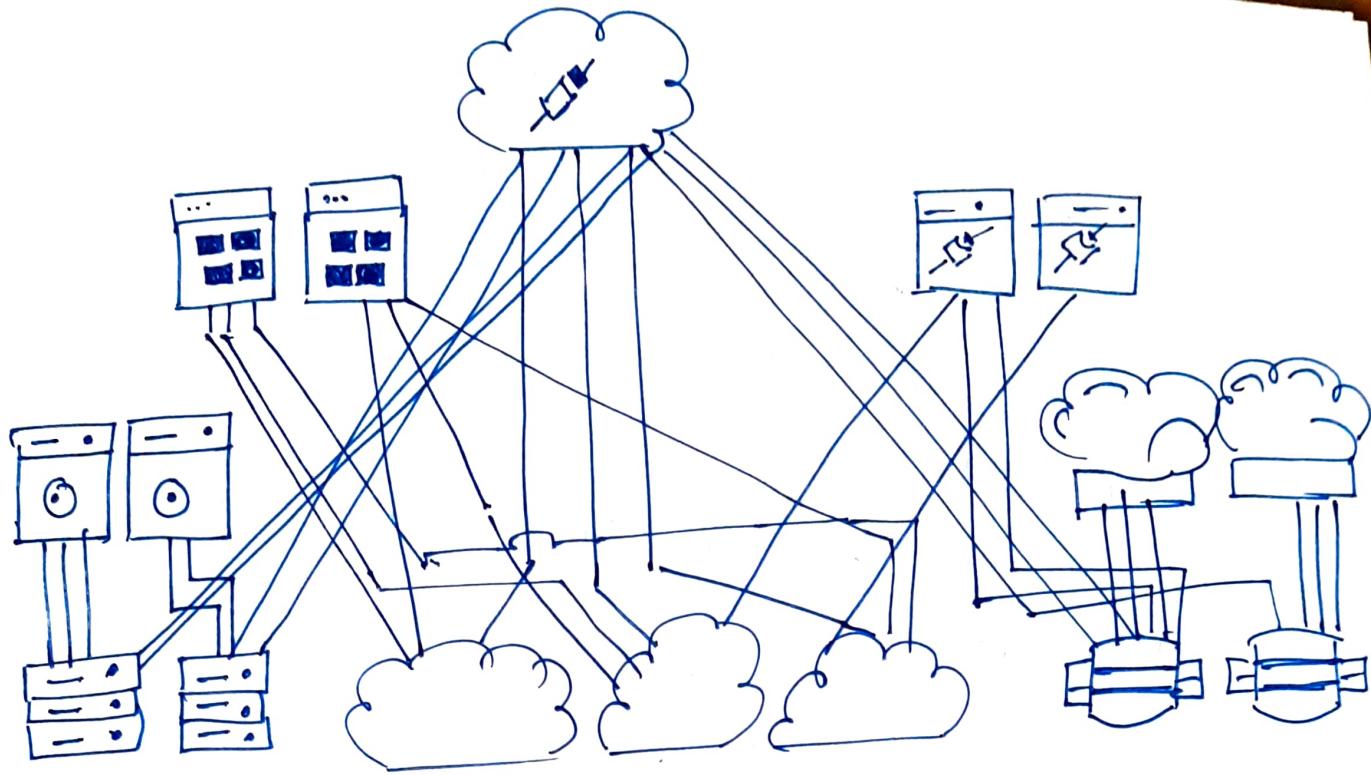
It is the ability of two or more systems or application to exchange information and to mutually use the information that has been exchanged.

Cloud interoperability is the ability of a customer system to interact with a cloud service or the ability for one cloud service to interact with other cloud services by exchanging information according to a prescribed method to obtain predictable result.

The two noteworthy dimensions of interoperability - connectivity and usability - have been divided into five layers, as is illustrated in the diagram below.



It's also the basic ability of different computerized products or systems to readily connect and exchange information, with one another, in either implementation or access, without restriction.



It's the property that facilitates unrestricted sharing and use of data or resources between disparate systems via local area networks (LANs) or wide area networks (WANs).

There are two types of data interoperability - syntactic interoperability, which is a prerequisite to semantic interoperability and enables different software components to cooperate, facilitating communication and exchange of data, and semantic interoperability, which refers to the ability of computer systems to exchange meaningful data with unambiguous, shared meaning.