SNS COLLEGE OF TECHNOLOGY

*(An Autonomous Institution)*

*Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai*

*Accredited by NAAC-UGC with ‘A++’ Grade (Cycle III) &*

*Accredited by NBA (B.E - CSE, EEE, ECE, Mech & B.Tech.IT)*

 COIMBATORE-641 035, TAMIL NADU

#  TWO MARKS QS and Answers

# UNIT IV - PROGRAMMING MODEL

1. **What is Hadoop development?**

Apache **Hadoop** is an open-source software **framework** written in Java for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware.

# Define- GT4.

Globus Toolkit 4 is a open – source toolkit developed to build grids. It provides full capabilities for sharing computing power and databases. Usage of Globus is extensive throughout the scientific community within NSF, DOE, DARPHA, IBM, NASA, and Microsoft projects.

# Define- Map Reduce Computation.

MapReduce is designed to continue to work in the face of system failures. When a job is running, MapReduce monitors progress of each of the servers participating in the job. If one of them is slow in returning an answer or fails before completing its work, MapReduce automatically starts another instance of that task on another server that has a copy of the data. The complexity of the error handling mechanism is completely hidden from the programmer



# What are the characteristics of Cloud Programming Model?

* + Cost model
	+ Scalability
	+ Fault-tolerance
	+ Support for specific services
	+ Control model
	+ Data model
	+ Synchronization mode

# What are the phases in MapReduce Programming Model?

**Map Phase:**

Processes input key/value pair Produces set of intermediate pair

map (in\_key, in\_value) -> list(out\_key, interm\_value)

# Reduce Phase:

Combines all intermediate values for a given key Produces a set of merged output values reduce(out\_key, list(interm\_value)) -> list(out\_value)

# Define- Hadoop Scheduler.

Job divided into several independent tasks executed in parallel

* + The input file is split into chunks of 64 / 128 MB
	+ Each chunk is assigned to a map task
	+ Reduce task aggregate the output of the map tasks

# Define- HDFS.

Hadoop File System was developed using distributed file system design. It is run on commodity hardware. Unlike other distributed systems, HDFS is highly fault tolerant and designed using low-cost hardware. HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored across multiple machines.

# What are the features of HDFS?

* + It is suitable for the distributed storage and processing.
	+ Hadoop provides a command interface to interact with HDFS.
	+ The built-in servers of name-node and data-node help users to easily check the status of cluster.
	+ Streaming access to data in the file system.
	+ HDFS provides file permissions and authentication.

# Sketch the HDFS Architecture.

****

1. **What is Cloud Dataflow Programming Model?**

The Dataflow programming model is designed to simplify the mechanics of large-scale data processing. When you program with a Dataflow SDK, you are essentially creating a data processing job to be executed by one of the Cloud Dataflow runner services. This model lets you concentrate on the logical composition of your data processing job, rather than the physical orchestration of parallel processing. You can focus on what you need your job to do instead of exactly how that job gets executed.

# What is Java Cloud service?

Oracle Java Cloud Service is a subscription-based, self-service, reliable, scalable, and elastic enterprise-grade cloud platform that enables businesses to securely develop and deploy Java applications.

* + Dedicated virtual machines for running your entire WebLogic Server cluster.
	+ Pre-configured WebLogic Server software, with your choice of the 11g or 12c version.
	+ Choice of virtual machine size (virtual cores, memory), as well as the size of the WebLogic cluster.
	+ Self-managed, with fully automated cloud tooling for administrative and lifecycle operations, such as patching, scaling, and backup.
	+ Fully automated, one-click, point-in-time restore for the entire service.
1. What is AIM?

The most-used instant messaging program is AOL Instant Messenger ([www.aim.com](http://www.aim.com/)), also known as AIM. AIM supports all manner of special features in addition to basic text messaging. The users get file sharing, RSS feeds, group chats, ability to text message to and from mobile phones, voice chat, video chat, and even a mobile client and can also enhance the basic AIM experience with a variety of official and user-created plug-ins.

1. Define- Multi-tenancy.

Multi-tenancy can be defined as a principle in software architecture, where a single instance of a vendor’s offering runs on the vendor’s servers, serving multiple client organizations (tenants). Often these tenants will pay a fee for this.

In practice, multi-tenancy allows a cloud provider to provide a service to organizations that have users of their own. Of course, in certain cases the tenant could have only 1 user; the important point is that the cloud provider has taken the tenant concept into account and provided e.g. access based on the tenant concept, billing based on the tenant concept, etc.

1. Define- GFS.

Google File System (**GFS** or GoogleFS) is a proprietary **distributed file system** developed by Google for its own use. It is designed to provide efficient, reliable access to data using large clusters of commodity hardware. A new version of the Google File System is codenamed Colossus which was released in 2010.

1. Define- OGF.

**OGF** is an open global community committed to driving the rapid evolution and adoption of modern advanced applied distributed computing, including cloud, grid and associated storage, networking and workflow methods. OGF is focused on developing and promoting innovative scalable techniques, applications and infrastructures to improve productivity in the enterprise and within the international research, science and business communities