

USN



Internal Assessment Test 5 – Feb 2022

Sub:	Cloud Computing and its Applications				Sub Code:	15CS742/ 17CS742	Branch :	CSE		
Date:	09/02/2022	Duration:	90 mins	Max Marks:	50	Sem / Sec:	7D	OBE		
<u>Answer any FIVE FULL Questions</u>								MARKS	CO	RB T
1	Explain three distributed File System						[10]	CO4	L4	
2	Discuss Aneka MapReduce infrastructure with a neat diagram						[10]	CO4	L2	
3	Explain Map Reduce computation workflow with a neat diagram						[10]	CO5	L4	
4	Describe core services of windows Azure platform						[10]	CO5	L2	
5	Explain CRM and ERP implementations with three examples and required diagrams						[10]	CO4	L4	
6	Write Short Notes on the Following: i. Amazon Dynamo ii. ECG Monitoring iii. Animoto media Applications						[10]	CO5	L3	

SOLUTIONS

USN



Internal Assessment Test 5 – Jan 2022

Sub:	Cloud Computing and its Applications				Sub Code:	15CS742/ 17CS742	Branch :	CSE		
Date:	09/02/2022	Duration:	90 mins	Max Marks:	50	Sem / Sec:	7D	OBE		
<u>Answer any FIVE FULL Questions</u>								MARKS	CO	RB T
1	Explain three distributed File System						[10]	CO4	L4	
	<ul style="list-style-type: none"> ▶ Scalable algorithms that can search and process massive datasets ▶ New metadata management technologies that can scale to handle complex, heterogeneous, and distributed data sources ▶ Advances in high-performance computing platforms aimed at providing a better support for accessing in-memory 									

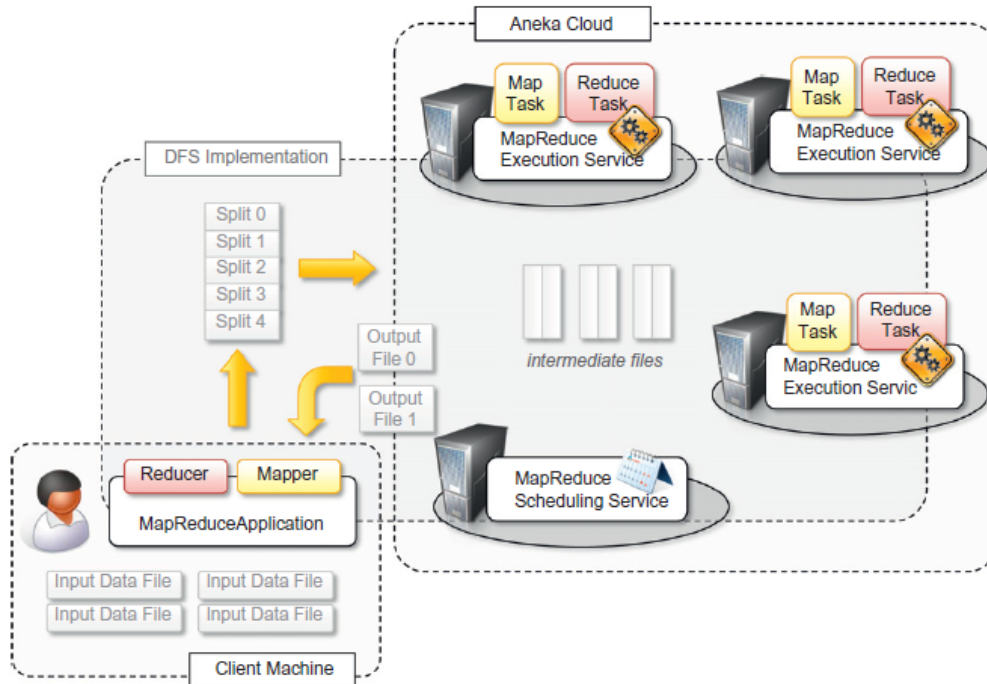
multiterabyte data structures

- ▶ **High-performance, highly reliable, petascale distributed file systems**
- ▶ **Data signature-generation** techniques for data reduction and rapid processing
- ▶ New approaches to **software mobility** for delivering algorithms that are able to move the computation to where the data are located
- ▶ Specialized **hybrid interconnection architectures** that provide better support for **filtering multigigabyte datastreams** coming from high-speed networks and scientific instruments
- ▶ Flexible and high-performance software integration techniques.
- ▶ **Distributed file systems constitute the primary support for data management**
- ▶ Provide an interface to store information in the form of files and later access them for read or write.
- ▶ **Lustre:** Lustre file system is a **massively parallel distributed file system** that covers the needs of a small workgroup of clusters to a large-scale computing cluster
 - provide access to petabytes (PBs) of storage
 - Users access the file system via a POSIX-compliant client
 - Mounted as a module through a kernel or library
 - 3 nodes - client, metaserver, object storage
 - <https://www.youtube.com/watch?v=oNeYpRDbB2M>
- ▶ **IBM General Parallel File System (GPFS)** (6 min):
<https://www.youtube.com/watch?v=eKxDcubEw8> –
- ▶ Google File System (GFS)

2 Discuss Aneka MapReduce infrastructure with a neat diagram

[10]

CO4 L2



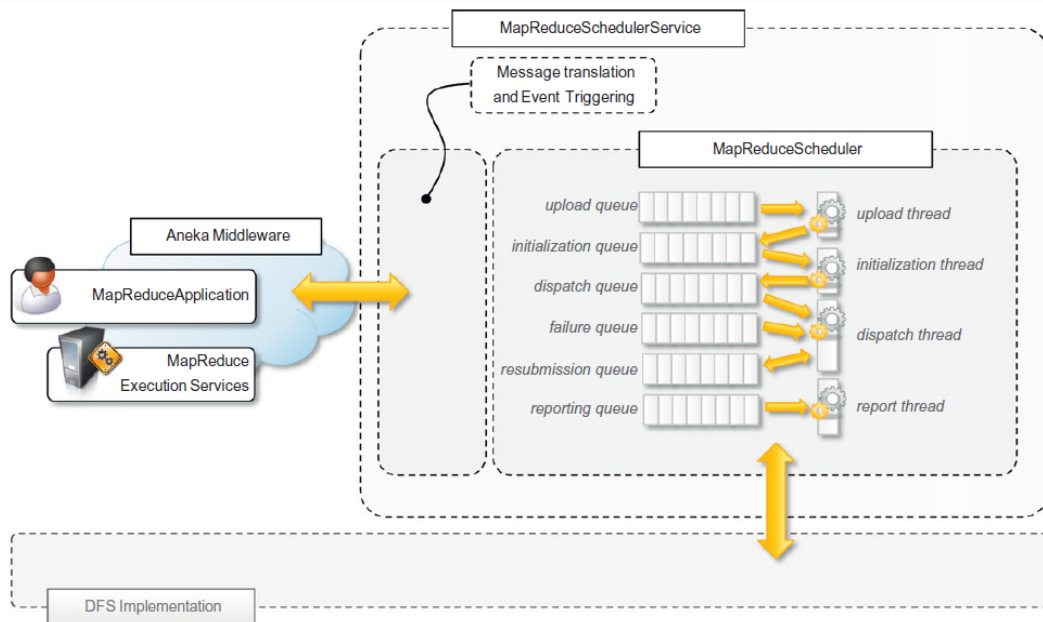
the abstractions and runtime support for developing MapReduce applications on top of Aneka

- ▶ A **MapReduce job** in Google MapReduce or Hadoop corresponds to the execution of a **MapReduce application** in Aneka
- ▶ The application instance has components identify map and reduce functions
- ▶ Expressed as **Mapper** and **Reducer** classes
- ▶ 3 main elements :
 - ▶ **MapReduce Scheduling Service**
 - ▶ **MapReduce Execution Service**
 - ▶ specialized **distributed file system**
- ▶ **Client components** : MapReduceApplication is used to submit execution of a MapReduce job

3 Explain Map Reduce computation workflow with a neat diagram

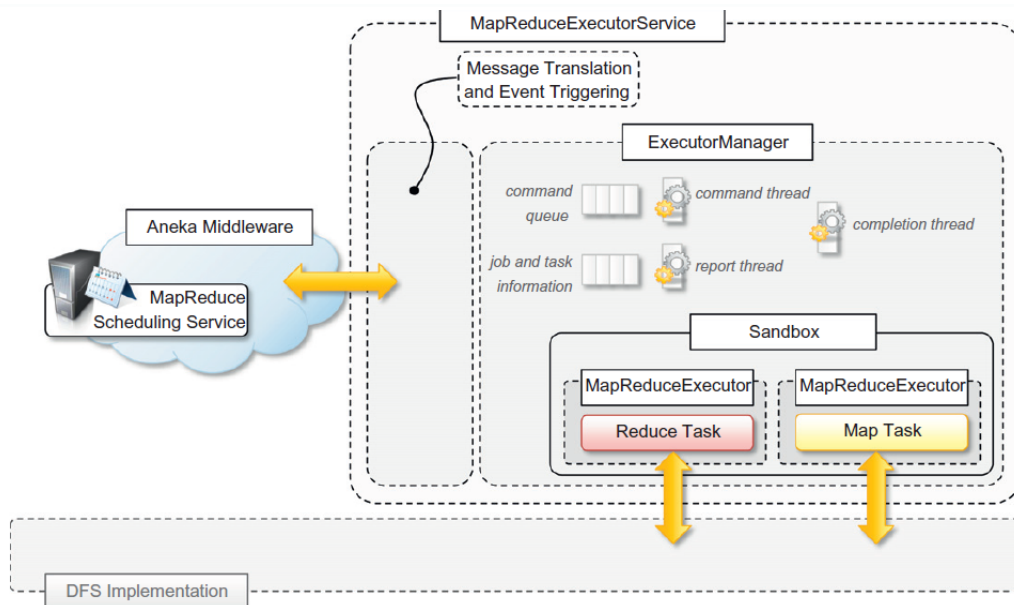
[10]

CO5 L4



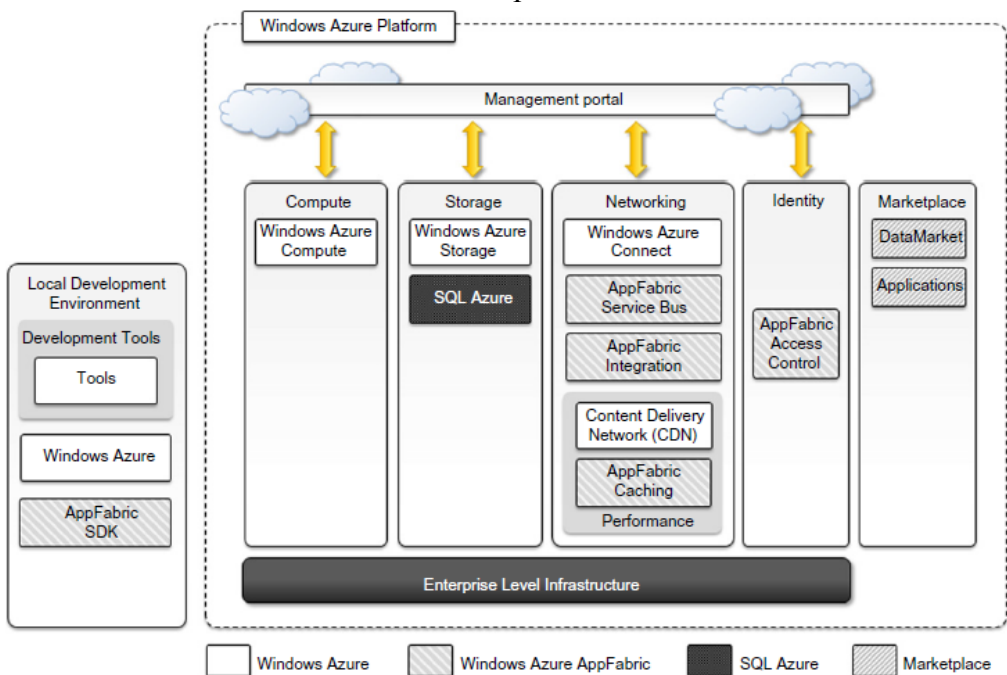
Comprises set of services that deal with scheduling and executing MapReduce tasks.

- ▶ **MapReduce Scheduling Service**
- ▶ **MapReduce Execution Service**
- ▶ Integrate with existing services of the framework to provide persistence, application accounting
- ▶ Job and Task Scheduling :
 - ▶ 2 major components : **MapReduceSchedulerService** : wrapper around the scheduler
 - ▶ **MapReduceScheduler** : controls execution of jobs and schedules tasks.
- ▶ Controlled by **MapReduceExecutionService**.
- ▶ plays the role of the worker process in Google MapReduce implementation.
- ▶ manages the **execution** of **map** and **reduce** tasks
- ▶ It also performs **sorting** and **merging** intermediate files.
- ▶ It is organized internally.



4 Describe core services of windows Azure platform

[10] CO5 L2



Web Role – hosted in IIS 7 web server.

- ▶ designed to implement scalable Web applications.
- ▶ Web roles represent the units of deployment of Web applications within the Azure infrastructure
- ▶ NET technology natively supports Web roles
- ▶ It is possible to develop ASP.NET (ASP.NET Web Role and ASP.NET MVC 2 Web Role) and WCF (WCF Service Web Role) applications.
- ▶ Worker role - designed to host compute services in Azure.
- ▶ They can be used to quickly provide compute power or to host services that do not communicate with the external world through HTTP
- ▶ Developing a worker role is like a developing a service.

- ▶ a Worker role runs continuously from the creation of its instance until it is shut down
- ▶ The Azure SDK provides developers with convenient APIs and libraries that allow connecting the role with the service provided by the runtime and easily controlling its startup as well as being notified of changes in the hosting environment

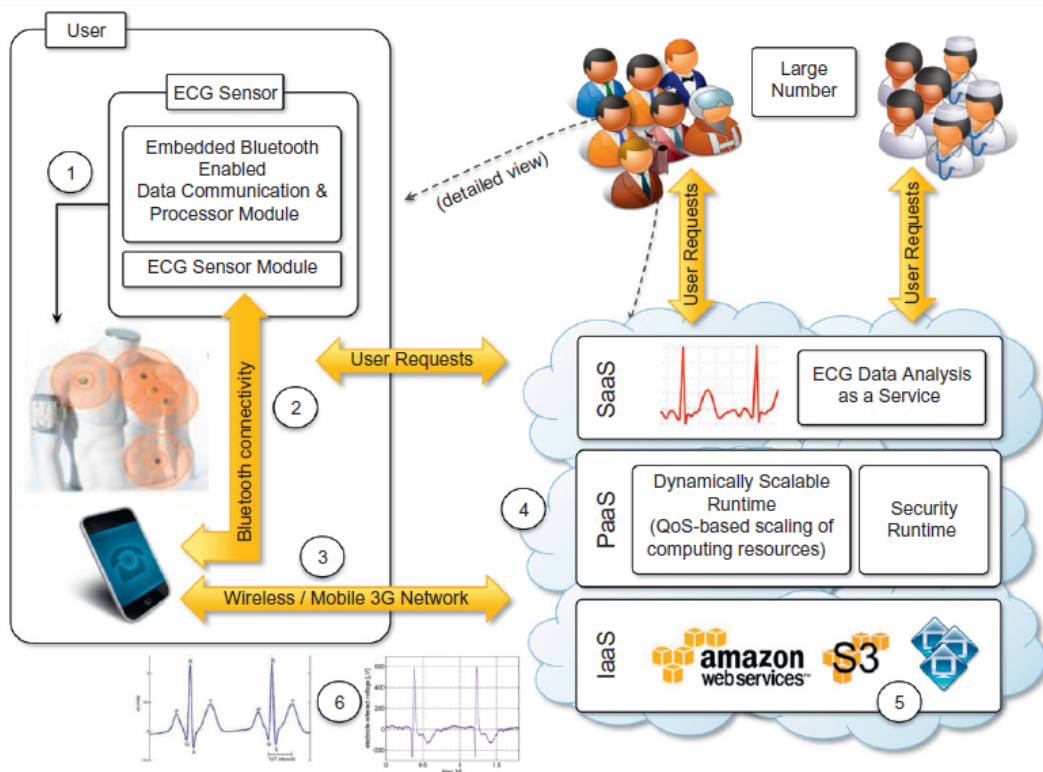
Windows Azure provides different types of storage solutions that

- ▶ complement compute services with a more durable and redundant option compared to local storage.
- ▶ **Blobs**
- ▶ Store large amount of data in the form of binary large objects (BLOBs).
- ▶ Two types of blobs are available:
 - ▶ **Block blobs.** composed of blocks optimized for sequential access;
 - ▶ Blocks are of 4 MB, and a single block blob can reach 200 GB.
 - ▶ **Page blobs.** pages that are identified by an offset from the beginning of blob.
 - ▶ Split into multiple pages or constituted of single page.
 - ▶ Optimized for random access.
 - ▶ Maximum dimension of a page blob can be 1 TB.

5 Explain CRM and ERP implementations with three examples and required diagrams [10]

- ▶ **Major motivating factor :** infinite availability of **computing resources** and storage at **sustainable prices** compared to a complete in-house deployment.
- ▶ Scientific applications that are harnessing power of the cloud
 - ▶ high-performance computing (**HPC**) applications
 - ▶ high-throughput computing (**HTC**) applications
 - ▶ **data-intensive applications.**
- ▶ most relevant option is **IaaS** solutions : offers the optimal environment for running **bag-of-tasks applications and workflows**

CO4	L4



Wearable computing devices equipped with ECG sensors constantly monitor the patient's heartbeat.

- ▶ transmitted to the patient's **mobile device**, eventually forwarded to **cloud-hosted Web service** for analysis
- ▶ Web service forms the front-end that is entirely hosted in the cloud
- ▶ Leverages SaaS, PaaS and IaaS.
- ▶ Web service constitute the SaaS application : **store ECG data in the Amazon S3 service**
 - ▶ issue a **processing request** to the scalable cloud platform

runtime platform is composed of a dynamically sizable number of instances running the workflow engine and Aneka.

6 Write Short Notes on the Following:
 i. Amazon Dynamo ii. ECG Monitoring iii. Animoto media Applications
 i. Amazon Dynamo

- ▶ is composed of a collection of storage peers organized in a ring that share same key space.
- ▶ **Key space** : partitioned among the storage peers
- ▶ Each peer is configured with access to a local storage facility where original objects and replicas are stored
- ▶ each node provides facilities for distributing the updates among the rings and to detect failures and unreachable nodes

[10] CO5 L3

ii. Apache Cassandra

- ▶ is a distributed object store for managing large amounts of **structured data** spread across many **commodity servers**
- ▶ The system is designed to **avoid a single point of failure**
- ▶ offer a highly **reliable service**
- ▶ initially developed by **Facebook**
- ▶ It is now part of the Apache incubator initiative
- ▶ Used by **Facebook, Digg, and Twitter**
- ▶ 2nd gen distributed database that builds on concept of
 - ▶ **AmazonDynamo** : fully distributed
 - ▶ **Google BigTable** : inherits the column family concept

iii. Amazon S3

- ▶ Amazon S3 is the online storage service provided by **Amazon**
- ▶ the system is claimed to support high availability, reliability, scalability, infinite storage, and low latency at commodity cost
- ▶ offers a flat storage space organized into buckets, which are attached to an **Amazon Web Services (AWS)** account
- ▶ Each **bucket** can store multiple objects, each identified by a unique key
- ▶ Objects are identified by **unique URLs and exposed through HTTP**
- ▶ Allows very simple **get-put** semantics
- ▶ Because of use of **HTTP**, there is no need for any specific library for accessing the storage system
- ▶ a **POSIX-like client library** has been developed to **mount S3 buckets** as part of the local file system.

- ▶ Since the buckets are linked to AWS accounts, the owner of a bucket can decide to make it visible to other accounts or the public

--	--

