MARVEL
CLOUD COMPUTING

TOPIC: MAP-REDUCE
DATE: 4/1/2022
MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. The map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Secondly, reduce the task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. Map-reduce has 2 stages - map stage, reduce stage.
The major advantage of MapReduce is that it is easy to scale data processing over multiple computing nodes.
REFERENCE :
https://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm

TOPIC: GOSSIP
DATE: 4/1/2022
Gossip Protocol is a communication protocol, it is a process of computer to computer communication that works on the same principle as to how information is shared on social networks. Nowadays, most systems often use gossip protocols to solve problems that might be difficult to solve in other ways, either due to inconvenience in the structure, is extremely large, or because gossip solutions are the most efficient ones available. Because gossip conveys information just like how a virus spreads in a biological community, the phrase "epidemic protocol" is frequently used interchangeably with the gossip protocol.
REFERENCE:
https://www.youtube.com/watch?v=pvdDVsxXKcM

https://www.tutorialspoint.com/the-gossip-protocol-in-cloud-computing

TOPIC: MEMBERSHIP AND GRIDS
DATE:6/1/2022
We need a mechanism to detect failures and disseminate the failure information through the network. This is where Group Membership Protocols come into the picture. We give every process a membership list. A membership list contains the list

of all processes along with its last heartbeat. (Heartbeat refers to the last time the process P communicated with the current process.)
Grid computing is the practice of leveraging multiple computers, often geographically distributed but connected by networks, to work together to accomplish joint tasks. It is typically run on a "**data grid**" a set of computers that directly interact with each other to coordinate jobs.
REFERENCE:
https://sunitha98selvan.medium.com/group-membership-protocols-15c62a93eec7

https://hazelcast.com/glossary/grid-computing/#:~:text=Grid%20computing%20is%20the%20practice,each%20other%20to%20coordinate%20jobs.

TOPIC: P2P SYSTEMS
DATE: 6/1/2022
Peer-to-peer (P2P) computing or networking is a distributed application architecture that partitions tasks or workloads between peers. Peers are equally privileged, equipotent participants in the application. They are said to form a peer-to-peer network of nodes.
Peers make a portion of their resources, such as processing power, disk storage or network bandwidth, directly available to other network participants, without the need for central coordination by servers or stable hosts. Peers are both suppliers and consumers of resources, in contrast to the traditional client-server model in which the consumption and supply of resources are divided.
REFERENCE:
https://en.m.wikipedia.org/wiki/Peer-to-peer
Peer to peer methods
https://www.google.com/amp/s/www.geeksforgeeks.org/what-is-p2ppeer-to-peer-process/amp/
Peer to peer file sharing
https://www.geeksforgeeks.org/p2ppeer-to-peer-file-sharing/

TOPIC: KEY-VALUE STORES TIME AND ORDERING
DATE: 10/1/2022
A key-value store, or key-value database, is a type of data storage software program that stores data as a set of unique identifiers, each of which has an associated value. This data pairing is known as a "key-value pair." The unique identifier is the "key" for an item of data, and a value is either the data being identified or the location of that data.
REFERENCE:
https://hazelcast.com/glossary/key-value-store/

https://medium.com/coinmonks/time-and-clocks-and-ordering-of-events-in-a-distributed-system-cdd3f6075e73

Time and clocks and ordering of events in a distributed system | by Ameya | Coinmonks | Medium

TOPIC: CLASSICAL DISTRIBUTED ALGORITHMS
DATE:10/1/2022
Distributed computing is a field of computer science that studies distributed systems. A distributed system is a system whose components are located on different networked computers, which communicate and coordinate their actions by passing messages to one another from any system. The components interact with one another to achieve a common goal. Three significant challenges of distributed systems are: maintaining the concurrency of components, overcoming the lack of a global clock, and managing the independent failure of components. When a component of one system fails, the entire system does not fail. Examples of distributed systems vary from SOA-based systems to massively multiplayer online games to peer-to-peer applications.
A computer program that runs within a distributed system is called a distributed program, and distributed programming is the process of writing such programs.[There are many different types of implementations for the message passing mechanism, including pure HTTP, RPC-like connectors and message queues.
REFERENCE:
https://en.wikipedia.org/wiki/Distributed_computing

TOPIC: Concurrency and Replication Control
DATE:10/1/2022
Data distribution is commonly used in high-performance computing (HPC). There are two fundamental data distribution topologies: replication and partition.
With data partitioning, you can achieve parallel processing of a large amount of data. With data replication, you can achieve load balancing (LB) and high availability (HA). This article focuses on the requirements for data replication.
Even though a data item has several replicas in a data replication environment, it should have some degree of data coherence by only appearing to be one virtual global item to the end-users. The biggest challenge in using data replication is the proper trade-off between data coherence and performance based on your business requirements.
REFERENCE:
https://www.infoq.com/articles/Concurrency-Control-Data-Replication/

https://www.youtube.com/watch?v=HB7dD2NnEVM

TOPIC: Emerging Paradigm
DATE:14/1/2022
Cloud computing has recently emerged as a new paradigm for hosting and delivering utility-oriented IT services to users over the Internet. Cloud computing is lucrative to enterprise owners as it eliminates the requirement for users to plan for resource provisioning and allows organizations to initiate from the small and increase resources only when there is a hike in service demand. Cloud computing has emerged as a new technology that will lead the next generation of the Internet. It tries to provide optimal and efficient computing through collaboration, agility, availability and scalability. In this paper, the cloud computing paradigm is illustrated in a variety of aspects and certain open challenges of research and recent development trends are being discussed. This research paper aims to provide a comprehensive overview of the cloud computing paradigm and identify its major research challenges and recent development trends taking place in this increasingly important field.
REFERENCE:
https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.672.8792&rep=rep1&type=pdf#:~:text=According%20the%20definition%20given%20by,%2C%20rapid%20elasticity%2C%20measured%20service.

TOPIC: Real-life behaviours
DATE:14/1/2022
In the cloud computing environment, users can directly access various cloud resources provided by Cloud Service Providers (CSP). User(s) with malicious intent can affect and/or destroy software and hardware resources in the cloud. The damage can occur from a variety of sources like competitors, hackers, etc. For example, in the PaaS service, users can develop and deploy a program to the cloud servers. The malicious user may submit code which attacks other users and occupy CPU time, memory space and other resources [3]. In a cloud environment, the traditional way of authorization is not enough for many reasons. User identity could be stolen, user may behave maliciously to destroy cloud servers or other resources on the cloud. To enhance the security in cloud computing, user behaviour trust plays an essential role.
REFERENCE:
https://easychair.org/publications/open/vD8J

TOPIC: Containers
DATE:17/1/2022
Containers are packages of software that contain all of the necessary elements to run in any environment. In this way, containers virtualize the operating system and run anywhere, from a private data centre to the public cloud or even on a developer's

laptop. From Gmail to YouTube to Search, everything at Google runs in containers. Containerization allows our development teams to move fast, deploy software efficiently, and operate at an unprecedented scale. We've learned a lot about running containerized workloads and we've shared this knowledge with the community along the way: from the early days of contributing groups to the Linux kernel, to taking designs from our internal tools and open-sourcing them as the project.
REFERENCE:
https://cloud.google.com/learn/what-are-containers

TOPIC: Virtual Machine

DATE:21/1/2022

A virtual machine, commonly shortened to just VM, is no different than any other physical computer like a laptop, smartphone or server. It has a CPU, memory, and disks to store your files and can connect to the internet if needed. While the parts that make up your computer (called hardware) are physical and tangible, VMs are often thought of as virtual computers or software-defined computers within physical servers, existing only as code.

REFERENCE:

https://azure.microsoft.com/en-in/overview/what-is-a-virtual-machine/#overview

TOPIC: JVM
DATE:21/1/2022
A **Java virtual machine** (**JVM**) is a virtual machine that enables a computer to run Java programs as well as programs written in other languages that are also compiled to Java bytecode. The JVM is detailed by a specification that formally describes what is required in a JVM implementation. Having a specification ensures interoperability of Java programs across different implementations so that program authors using the JDK need not worry about the idiosyncrasies of the underlying hardware platform.
REFERENCE:
https://www.geeksforgeeks.org/jvm-works-jvm-architecture/

TOPIC: MAAS

DATE:24/1/2022

MaaS is a security solution based on cloud computing. MaaS lets early detection of threats and reports it to their customer via emails. MaaS automates the detection and management of threats. MaaS also provides continuous system patching to update the security level and support the newer version of the application installed on the system. MaaS provides the forensic analysis of the vulnerability to identify the time effort and cost required to resolve it. MaaS provide 24/7/365 days of assistance to their customers.

REFERENCE:

https://binaryterms.com/monitoring-as-a-service-maas.html

TOPIC: PASS

DATE:

Platform as a Service (PaaS) provides a runtime environment. It allows programmers to easily create, test, run, and deploy web applications. You can purchase these applications from a cloud service provider on a pay-as-per use basis and access them using the Internet connection. In PaaS, back end scalability is managed by the cloud service provider, so end-users do not need to worry about managing the infrastructure. PaaS includes infrastructure (servers, storage, and networking) and platform (middleware, development tools, database management systems, business intelligence, and more) to support the web application life cycle.

REFERENCE:

https://www.javatpoint.com/platform-as-a-service

https://www.tutorialspoint.com/cloud_computing/cloud_computing_platform_as_a_service.htm

TOPIC: Web Services

DATE:27/1/2022

A web service is a set of open protocols and standards that allow data exchange between different applications or systems. Web services can be used by software programs written in different programming languages and on different platforms to

exchange data through computer networks such as the Internet. In the same way, communication on a computer can be inter-processed.

REFERENCE:
https://www.javatpoint.com/web-services-in-cloud-computing

TOPIC: NAAS
DATE:1/2/2022
Network-as-a-service (NaaS) is a cloud servicing model in which customers rent networking services from cloud providers. NaaS allows customers to operate their networks without maintaining their networking infrastructure. Like other cloud services, NaaS vendors run networking functions using software, essentially allowing companies to set up their networks entirely without hardware. All they need is internet connectivity. and can replace VPNs, MLPs, connections, or other legacy network configurations. It can also replace on-premise networking hardware such as firewalls appliances and load balancers. A newer model for routing traffic and applying security policies, and has had a major impact on enterprise networking architecture.
REFERENCE:
https://www.cloudflare.com/en-in/learning/network-layer/network-as-a-service-naas/#:~:text=service%20(NaaS)%3F-,Network%2Das%2Da%2Dservice%20(NaaS)%20is%20a,maintaining%20their%20own%20networking%20infrastructure

TOPIC: Ceph
DATE:3/2/2022
Ceph is an open-source software-defined storage solution designed to address the block, file and object storage needs of modern enterprises. Its highly scalable architecture sees it being adopted as the new norm for high-growth block storage, object stores, and data lakes. Ceph provides reliable and scalable storage while keeping CAPEX and OPEX costs in line with underlying commodity hardware prices.
REFERENCE:
https://ubuntu.com/ceph/what-is-ceph

TOPIC: SWIFT
DATE:8/2/2022
The OpenStack Object Store project, known as Swift, offers cloud storage software so that you can store and retrieve lots of data with a simple API. It's built for scale and optimized for durability, availability, and concurrency across the entire data set. Swift is ideal for storing unstructured data that can grow without bound.

REFERENCE:
https://wiki.openstack.org/wiki/Swift

TOPIC: HDFS
DATE:8/2/2022
The Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from other distributed file systems are significant. HDFS is highly fault-tolerant and is designed to be deployed on low-cost hardware. HDFS provides high throughput access to application data and is suitable for applications that have large data sets. HDFS relaxes a few POSIX requirements to enable streaming access to file system data. HDFS was originally built as infrastructure for the Apache Nutch web search engine project. HDFS is now an Apache Hadoop subproject
REFERENCE:
https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

TOPIC: SAN
DATE:10/2/2022

A SAN is a specialized, high-speed network that provides block-level network access to storage. SANs are typically composed of hosts, switches, storage elements, and storage devices that are interconnected using a variety of technologies, topologies, and protocols. SANs may also span multiple sites. A SAN presents storage devices to a host such that the storage appears to be locally attached. This simplified presentation of storage to a host is accomplished through the use of different types of virtualisation.

REFERENCE:

https://www.snia.org/education/storage_networking_primer/san/what_san

TOPIC: Zookeeper.
DATE:15/2/2022

ZooKeeper is a distributed coordination service to manage a large set of hosts. Co-ordinating and managing a service in a distributed environment is a complicated process. ZooKeeper solves this issue with its simple architecture and API. ZooKeeper allows developers to focus on core application logic without worrying about the distributed nature of the application. The ZooKeeper framework was

originally built at “Yahoo!” for accessing their applications easily and robustly. Later, Apache ZooKeeper became a standard for organized service used by Hadoop, HBase, and other distributed frameworks. For example, Apache HBase uses ZooKeeper to track the status of distributed data.

REFERENCE:

https://www.tutorialspoint.com/zookeeper/zookeeper_overview.htm

BY,
NAME: SUJATHA BHAT