Unit - 3

#) Data Management issues in Mobile database

Data management technology that can support easy data access from and to mobile devices is among the main concerns in mobile information systems. Mobile computing may be considered a variation of distributed computing.. Here are some of the issues which arises in data management of the mobile databases:

1. **Mobile database design** – Because of the frequent shutdown and for handling the queries, the global name resolution problem is compounded.

2. **Security** – The data which is left at the fixed location is more secure as compared to mobile data. That is mobile data is less secure. Data are also becoming more volatile and techniques must be able to compensate for its loss.

3. **Data distribution and replication** – Uneven distribution of data among the mobile units and the base stations take place here. Higher data availability and low cost of remote access is there in data distribution and replication. The problem of Cache management is compounded by the consistency constraints.

4. **Replication issues** – There is increase of costs for updates and signalling due to increase in number of replicas. Mobile hosts can move anywhere and anytime.

5. **Division of labour** – There is a certain change in the division of labour in query processing because of certain characteristics of the mobile environment. There are some of the cases in which the client must function independently of the server.

6. **Transaction models** – In mobile environment, the issues of correctness of transactions and fault tolerance are aggravated. All transactions must satisfy the ACID properties, these are atomic, consistency, isolation, and durability. Depending upon the movement of the mobile unit, possibly on multiple data sets and through several base station, a mobile transaction is executed sequentially. When the mobile computers are disconnected, ACID properties gets hard to enforce. Because of the disconnection in mobile units, there is expectation that a mobile transaction will be lived long.

7. **Recovery and fault tolerance** – Fault tolerance is the ability of a system to perform its function correctly even in the presence of internal faults. Faults can be classified in two types: transient and permanent. The characterization of mobile computing is done by:
• Limiting resource availability • Frequent disconnection • High mobility • Low bandwidth

8. **Location based service** – One of the most challenging tasks which must be undertaken is determining the location of mobile users, which must be undertaken in order to enable a location based service. A cache information becomes sale when clients move location dependent. Issues that arises in location and services are:
• User Privacy
• Diverse mobile mapping standards
• Market capability
• Interoperability Updation of the location dependent queries and then applying spatial queries to refresh the cache causes a problem.

9. **Query processing** – Because of the mobility and rapid resource changes of mobile units, Query optimization becomes the most complicated. That is query processing is affected when mobility is considered. There is a need to returned a query response to mobile units that may be in transit.

#) Data Replication :

Data Replication is the process of storing data in more than one site or node. It is useful in improving the availability of data. It is simply copying data from a database from one server to another server so that all the users can share the same data without any inconsistency. The result is a distributed database in which users can access data relevant to their tasks without interfering with the work of others. Data replication encompasses duplication of transactions on an ongoing basis, so that the replicate is in a consistently updated state and synchronized with the source. However in data replication data is available at different locations, but a particular relation has to reside at only one location. There can be full replication, in which the whole database is stored at every site. There can also be partial replication, in which some frequently used fragment of the database are replicated and others are not replicated.

ADVANTAGES OF DATA REPLICATION – Data Replication is generally performed to:
• To provide a consistent copy of data across all the database nodes.
• To increase the availability of data.
• The reliability of data is increased through data replication.

• Data Replication supports multiple users and gives high performance.
• To perform faster execution of queries.

DISADVANTAGES OF DATA REPLICATION –

• More storage space is needed as storing the replicas of same data at different sites consumes more space.
• Data Replication becomes expensive when the replicas at all different sites need to be updated.
• Maintaining Data consistency at all different sites involves complex measures

Types of Data Replication –

1. **Transactional Replication** – In Transactional replication users receive full initial copies of the database and then receive updates as data changes. Data is copied in real time from the publisher to the receiving database(subscriber) in the same order as they occur with the publisher therefore in this type of replication, transactional consistency is guaranteed. Transactional replication is typically used in server-to-server environments. It does not simply copy the data changes, but rather consistently and accurately replicates each change.

2. **Snapshot Replication** – Snapshot replication distributes data exactly as it appears at a specific moment in time does not monitor for updates to the data. The entire snapshot is generated and sent to Users. Snapshot replication is generally used when data changes are infrequent. It is bit slower than transactional because on each attempt it moves multiple records from one end to the other end. Snapshot replication is a good way to perform initial synchronization between the publisher and the subscriber.

3. **Merge Replication** – Data from two or more databases is combined into a single database. Merge replication is the most complex type of replication because it allows both publisher and subscriber to independently make changes to the database. Merge replication is typically used in server-to-client environments. It allows changes to be sent from one publisher to multiple subscribers.

#) File system

A file system is a process of managing how and where data on a storage disk which is also referred to as a file management . It is logical disk component that compresses file separated into groups which is known as directories . The File System Enables you to view a file in the current directory as files are often managed in a hierarchy.

Goals :

Efficient and transparent access to Stored Files within a mobile environment while manufacturing while maintaining data consistency.

#) Transaction Processing :

A Transaction Processing System or TPS is software that keeps track of transactions by processing the data in an online recording system.
It is a logical unit of database processing that includes one or more access operations. (read - retrieval, write-insert or update). It is a unit of program execution that accesses and if required updates various data items.
A transaction is a set of operations that can either be embedded within an application program or can be specified interactively via a high-quality language such as SQL.

Types of Transaction Processing Systems

- **Batch processing**
  Batch processing is when clusters of transactions are refined simultaneously using a computer system.
  This method, although designed to be efficient for breaking down bulky series of programs, has a drawback
  as there is a delay in the transaction result.
- **Real-time Processing**
  Real-time processing carries out its transactions exclusively; this method ensures a swift reply on the condition of the transaction result. It is an ideal technique for dealing with singular transactions.

Transaction Processing System Features

There are several features involved in a good transaction processing system. A few of these critical features are described below.

- **Performance**
  The concept behind the use of TPS is to efficiently generate timely results for transactions. Effectiveness is based on the number of transactions they can process at a particular time.
- **Continuous availability**
  The transaction processing system should be a very stable and reliable system that must not crash easily. Disruption of TPS in an organization can lead to work disturbance and financial loss.
- **Data integrity**
  The TPS must maintain the same method for all transactions processed, the system must

be designed to effectively protect data and overcome any hardware/ software issues.

- **Ease of use**
  The TPS should be user-friendly in order to encourage the use and also decrease errors from inputting data. It should be structured in such a way that it makes it easy to understand as well as guarding users against making errors during data-entry.
- **Modular growth**
  The TPS hardware and software components should be able to be upgraded individually without requiring a complete overhaul.
- **Controlled processing**
  Only authorized personnel, staff members, or employees should be able to access the system at a time.

Advantages :

- Batch processing or real-time processing available.
- Reduction in processing time, lead time and order cycle time.
- Reduction in inventory, personnel and ordering costs.
- Increase in productivity and customer satisfaction

Disadvantages :

- High setup costs.
- Lack of standard formats.

#) CLOUD ARCHITECTURE

Mobile cloud computing is a new platform combining the mobile devices and cloud computing to create a new infrastructure, whereby cloud performs the heavy lifting of computing-intensive tasks and storing massive amounts of data. In this new architecture, data processing and data storage happen outside of mobile devices.

Architectures of mobile cloud computing :

Mobile devices are connected to the mobile networks via base stations (e.g., base transceiver station, access point, or satellite) that establish and control the connections (air links) and functional interfaces between the networks and mobile devices. Mobile users' requests and information (e.g., ID and location) are transmitted to the central processors that are connected to servers providing mobile network services. Here, mobile network operators can provide services to mobile users as authentication, authorization, and accounting based on the home agent and subscribers' data stored in databases. After that, the subscribers' requests are delivered to a cloud through the Internet. In the cloud, cloud controllers process the requests to provide mobile users with the corresponding cloud services. These services are developed

with the concepts of utility computing, virtualization, and service-oriented architecture (e.g., web, application, and database servers). The details of cloud architecture could be different in different contexts.

Advantages :
Extending battery lifetime: Computation offloading migrates large computations and complex processing from resource-limited devices (i.e., mobile devices) to resourceful machines (i.e., servers in clouds).
Improving data storage capacity and processing power: MCC enables mobile users to store/access large data on the cloud. MCC helps reduce the running cost for computation intensive applications.
Improving reliability and availability : Keeping data and application in the clouds reduces the chance of lost on the mobile devices.
Scalability

#) Types of Clouds :

1. Public Cloud:

- Public Cloud provides a **shared platform** that is accessible to the **general public** through an Internet connection.
- Public cloud operated on the **pay-as-per-use model** and administrated by the **third party**, i.e., Cloud service provider.
- In the Public cloud, the same storage is being used by multiple users at the same time.
- Public cloud is **owned, managed,** and **operated** by businesses, universities, government organizations, or a combination of them.

Advantages

- **Controls:** Better controls for data, users and information assets.
- **Cost:** Initial investment for hardware is very high in case of an on-premise infrastructure.
- **Security:** The cloud belongs to a single client. Hence, the infrastructure and systems can be configured to provide high levels of security.
- **Superior Performance:** Normally private clouds are deployed inside the firewall of the organization's intranet which ensures efficiency and good network performance.
- **Easy Customization:** The hardware and other resources can be customized easily by the company.
- **Compliance:** Compliance is achieved easily in private clouds.

Disadvantages

- **Cost:** Costs are substantial in the case of building an on-premise private cloud. The running cost would include personnel cost and periodic hardware upgrade costs. In the case of outsourced private cloud, operating cost will include per resource usage and subject to change at the discretion of the service provider.
- **Under-utilization:** In some instances the resources subscribed can be under-utilized. Hence, optimizing the utilization of all resources is a challenge.
- **Capacity ceiling:** Due to physical hardware limitations with the service provider, there could be a capacity ceiling to handle only certain amount of servers or storage.
- **Vendor lock-in:** This can be a major impediment in private cloud adoption especially when the hardware and infrastructure is outsourced. This is a service delivery technique where the client company is forced to continue with the same service provider, thus preventing the client to migrate to another vendor.

2. Private Cloud:

- Private cloud is also known as an **internal cloud** or **corporate cloud**.
- Private cloud provides computing services to a **private internal network (within the organization)** and **selected users** instead of the general public.
- Private cloud provides a **high level of security** and **privacy** to data through firewalls and internal hosting. It also ensures that operational and sensitive data are not accessible to third-party providers.
- HP Data Centers, Microsoft, Elastra-private cloud, and Ubuntu are the example of a private cloud.

Advantages

In public clouds the resources are shared between multiple clients and all the services are controlled by services provider.

- **Simple and easy:** Public clouds are available as a service in the internet, they are easy to deploy.
- **Cost:** Initial investment is very low or nil.
- **Less time:** The IT resources and services are available immediately saving time for the company.
- **No maintenance:** The hardware and networks are maintained by the cloud services provider. Internal IT staffs have no responsibility in maintaining the infrastructure.
- **No contracts:** No long term commitment with service provider because public clouds are usually pay-as-you-go models.

Disadvantages

- **Performance:** The performance of the network depends on the speed of the internet connectivity.
- **Weak on Security:** Since the hardware resource is shared between multiple users, IT security issues are more profound and data is vulnerable to thefts.
- **Customization:** Customization of resources or services is not possible.

3) Hybrid Cloud :

- Hybrid cloud is a combination of **public and private** clouds.
  Hybrid cloud = public cloud + private cloud
- The main aim to combine these cloud (Public and Private) is to create a unified, automated, and well-managed computing environment.
- In the Hybrid cloud, **non-critical activities** are performed by the **public cloud** and **critical activities** are performed by the **private cloud**.
- Mainly, a hybrid cloud is used in finance, healthcare, and Universities.
- The best hybrid cloud provider companies are **Amazon, Microsoft, Google, Cisco,** and **NetApp**.

Advantages of Hybrid Cloud

There are the following advantages of Hybrid Cloud -

1) Flexible and secure

It provides flexible resources because of the public cloud and secure resources because of the private cloud.

2) Cost effective

Hybrid cloud costs less than the private cloud. It helps organizations to save costs for both infrastructure and application support.

3) Cost effective

It offers the features of both the public as well as the private cloud. A hybrid cloud is capable of adapting to the demands that each company needs for space, memory, and system.

4) Security

Hybrid cloud is secure because critical activities are performed by the private cloud.

5) Risk Management

Hybrid cloud provides an excellent way for companies to manage the risk.

Disadvantages of Hybrid Cloud

1) Networking issues

In the Hybrid Cloud, networking becomes complex because of the private and the public cloud.

2) Infrastructure Compatibility

Infrastructure compatibility is the major issue in a hybrid cloud. With dual-levels of infrastructure, a private cloud controls the company, and a public cloud does not, so there is a possibility that they are running in separate stacks.

3) Reliability

The reliability of the services depends on cloud service providers.

#) Data Processing

Data processing occurs when data is collected and translated into usable information. Usually performed by a data scientist or team of data scientists, it is important for data processing to be done correctly as not to negatively affect the end product, or data output.

Data processing starts with data in its raw form and converts it into a more readable format (graphs, documents, etc.), giving it the form and context necessary to be interpreted by computers and utilized by employees throughout an organization.

Six stages of data processing

1. Data collection

Collecting data is the first step in data processing. Data is pulled from available sources, including data lakes and data warehouses. It is important that the data sources available are trustworthy and well-built so the data collected (and later used as information) is of the highest possible quality.

2. Data preparation

Once the data is collected, it then enters the data preparation stage. Data preparation, often referred to as "pre-processing" is the stage at which raw data is cleaned up and organized for the following stage of data processing. During preparation, raw data is diligently checked for any errors. The purpose of this step is to eliminate bad data (redundant, incomplete, or incorrect data) and begin to create high-quality data for the best business intelligence.

3. Data input

The clean data is then entered into its destination (perhaps a CRM like Salesforce or a data warehouse like Redshift), and translated into a language that it can understand. Data input is the first stage in which raw data begins to take the form of usable information.

4. Processing

During this stage, the data inputted to the computer in the previous stage is actually processed for interpretation. Processing is done using machine learning algorithms, though the process itself may vary slightly depending on the source of data being processed (data lakes, social networks, connected devices etc.) and its intended use (examining advertising patterns, medical diagnosis from connected devices, determining customer needs, etc.).

5. Data output/interpretation

The output/interpretation stage is the stage at which data is finally usable to non-data scientists. It is translated, readable, and often in the form of graphs, videos, images, plain text, etc.). Members of the company or institution can now begin to self-serve the data for their own data analytics projects.

6. Data storage

The final stage of data processing is storage. After all of the data is processed, it is then stored for future use. While some information may be put to use immediately, much of it will serve a purpose later on. Plus, properly stored data is a necessity for compliance with data protection legislation like GDPR. When data is properly stored, it can be quickly and easily accessed by members of the organization when needed.

#) Clustering

A computer clustering is a collection of interconnected stand-alone computers which can work together collectively and cooperatively as a single integrated computing resource pool. Clustering explores massive parallelism at the job level and achieves high availability- HA through stand-alone operations.
Benefits : Scalable, Performance, HA, Fault Tolerance, Modular Growth, use of commodity component

Uses:

Clustering has a myriad of uses in a variety of industries. Some common applications for clustering include the following:

- market segmentation
- social network analysis
- search result grouping
- medical imaging
- image segmentation
- anomaly detection

What are the advantages of Clustering?

Clustering is said to be more effective than random sampling of the given data due to several reasons. The two major advantages of clustering are:

1. Requires fewer resources

A cluster creates a group of fewer resources from the entire sample. Due to this, there is a lesser requirement of resources as compared to random sampling. Random sampling will require travel and administrative expenses, but this is not the case over here.

2. Feasible option

Here, every cluster determines an entire set of the population as homogeneous groups are created from the entire population. With this, it becomes easy to include more subjects in a single study.

#) Fault tolerance :

Fault tolerance refers to the ability of a system (computer, network, cloud cluster, etc.) to continue operating without interruption when one or more of its components fail.

The objective of creating a fault-tolerant system is to prevent disruptions arising from a single point of failure, ensuring the high availability and business continuity of mission-critical applications or systems.

Fault-tolerant systems use backup components that automatically take the place of failed components, ensuring no loss of service. These include:

- **Hardware systems** that are backed up by identical or equivalent systems. For example, a server can be made fault tolerant by using an identical server running in parallel, with all operations mirrored to the backup server.
- **Software systems** that are backed up by other software instances. For example, a database with customer information can be continuously replicated to another machine. If the primary database goes down, operations can be automatically redirected to the second database.
- **Power sources** that are made fault tolerant using alternative sources. For example, many organizations have power generators that can take over in case main line electricity fails.

#) Resource management and scheduling

■ Critical function of any man-made system.

■ It affects the three basic criteria for the evaluation of a system:

- Functionality.
- Performance.
- Cost.

■ Scheduling in a computing system – deciding how to allocate resources of a system, such as CPU cycles, memory, secondary storage space, I/O and network bandwidth, between users and tasks.

■ Policies and mechanisms for resource allocation.
- Policy – principles guiding decisions.
- Mechanisms – the means to implement policies.

#) Adaptive Clustering

In the proposed network architecture, nodes are organized into non overlapping clusters. The clusters are independently controlled and are dynamically reconfigured as nodes move. This network architecture has three main advantages.

- First, it provides spatial reuse of the bandwidth due to node clustering.
- Secondly, bandwidth can be shared or reserved in a controlled fashion in each cluster.
- Finally, the cluster algorithm is robust in the face of topological changes caused by node motion, node failure and node insertion/removal.