

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

(Week 14) Lecture 27-28

Objectives: Learning objectives of this lecture are

- **Motivation for DDBS**
- **Distributed Computing**
- **What is DDBS?**
 - **What is DDB?**
 - **What is D-DBMS?**
- **What is not DDBS?**
- **Centralized and Distributed Database Environment**
- **Data Delivery Alternatives**
 - **Delivery modes**
 - **Pull-only**
 - **Push-only**
 - **Hybrid**
 - **Frequency**
 - **Periodic**
 - **Conditional**
 - **Ad-hoc or irregular**
 - **Communication methods**
 - **Uni-cast**
 - **One-to-many**

Reference Material: Database Systems 4th edition Thomas Conolly (Chapter 22)

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

Overview:

In this lecture, we will discuss differences between Distributed Computing and Distributed Databases.

- We will discuss distributed database, Distributed Database Management System and Distributed Database System.
- What is centralized databases system and what is distributed database system with examples.
- Data delivery alternatives in different systems.
- What are different types of Delivery modes, types of frequencies and types of communication methods used in data delivery alternatives.

Recap:

- In previous lecture, we had discussed how recovery system of a DBMS works.
- How data is read and written to storage device.
- What is buffer and what it does.
- How DBMS would react at the event of failure.
- And in the end we had discussed what recovery facilities a DBMS would provide to minimize the effect of failure on the system.

Distributed Computing

A number of autonomous processing elements (not necessarily homogeneous) that are interconnected by a computer network and that cooperate in performing their assigned tasks.

- What is being distributed?
 - Processing logic
 - Function
 - Data
 - Control

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

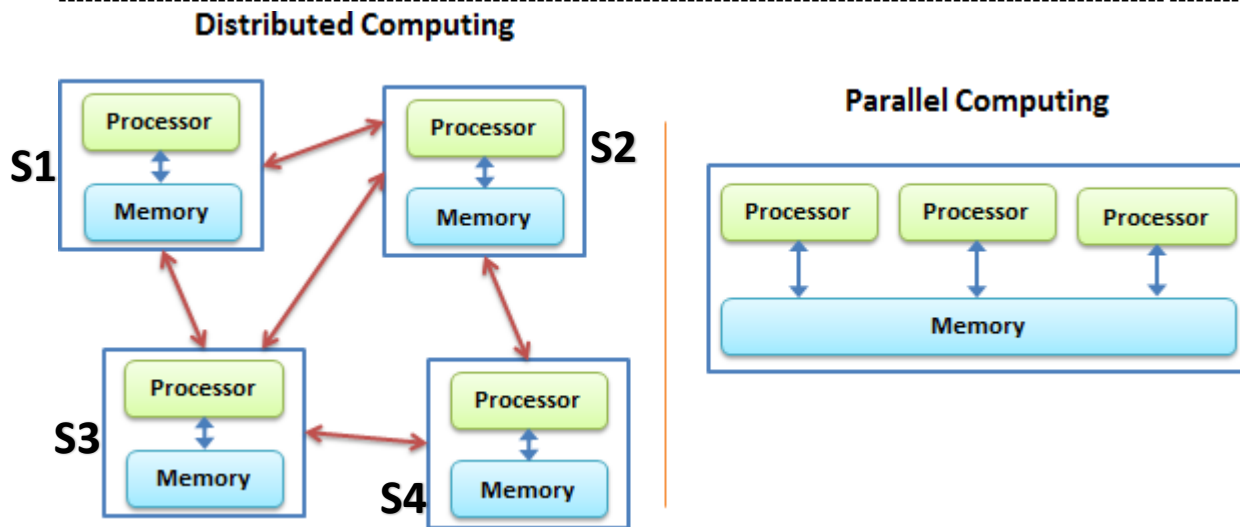


Figure 1

- As we can see in above diagram, that in distributed computing, four systems {S1, S2, S3, S4} are interconnected with each other and sharing their personal resources like processor and memory with each other to fulfill their needs.
- On the other hand, in parallel computing, there is only one system, which has multiple processing units operating at the same time.

What is Distributed Database (DDB)?

A distributed database (DDB) is a collection of multiple, *logically interrelated* databases distributed over a *computer network*.

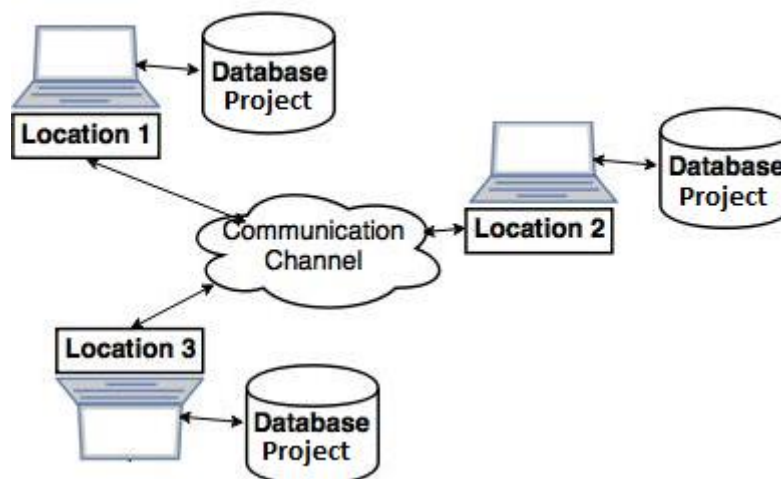


Figure 2

- As we can see that there are three locations at which a database 'Project' is placed. So this is collection of multiple database.

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

- By logically interrelated, it means that all these databases must have some logical link with each other. As in above diagram, all databases have same name, which leads us to a conclusion that these databases must be serving similar purpose. And schema of all these databases will be almost same with each other.
- As all these databases are connected via a computer network and interchanging information with each other which makes it a setup of distributed database.

What is Distributed Database Management System (D-DBMS)?

A distributed database management system (D-DBMS) is the software that manages the DDB and provides an access mechanism that makes this distribution transparent to the users.

Examples:

- 1- SQL server
- 2- Oracle Database
- 3- MySql
- 4- MongoDB
- 5- Postgre SQL

All above software provides distributed management of data over a computer network.

What is Distributed Database System (DDBS)?

DDBS is a combination of distributed database (DDB) and distributed database management system (D-DBMS).

Distributed database system (DDBS) = DDB + D-DBMS

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

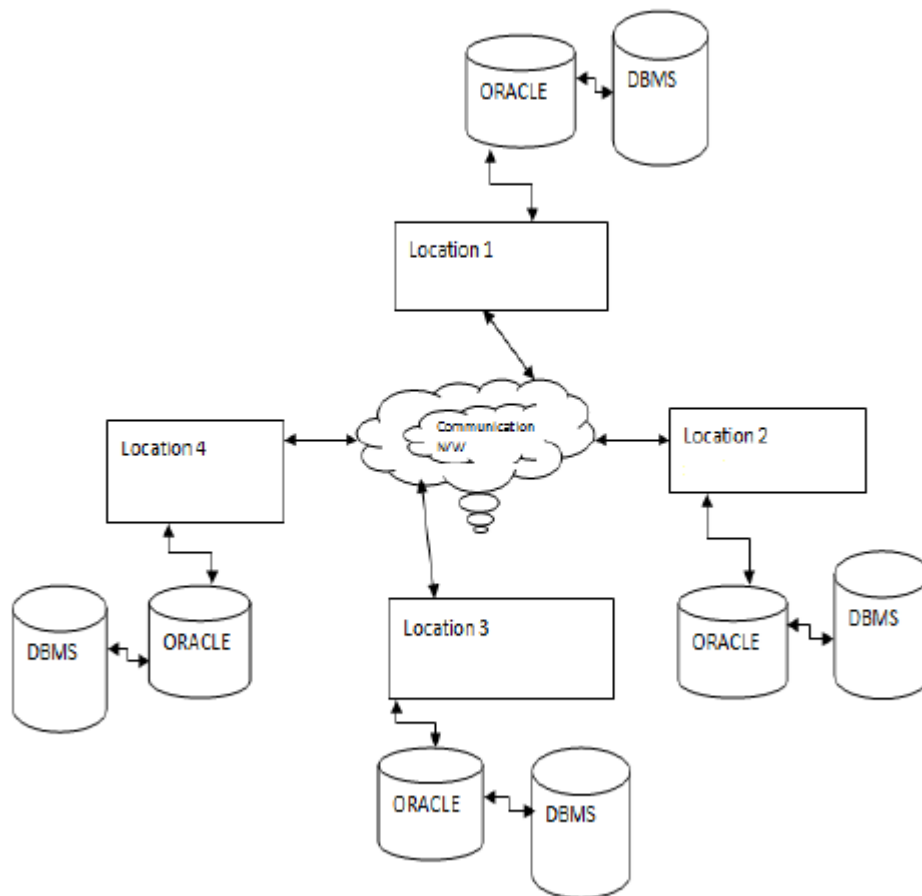


Figure 3

- In above figure, we can see that each location has a DBMS (Oracle) and a database connected to it.
- So above system is an example of “Distributed Database System”.

What is not Distributed Database Systems (DDBS)?

- A timesharing computer system. **“In computing, time-sharing is the sharing of a computing resource among many users at the same time by means of multiprogramming and multi-tasking.”**
- A loosely (separate primary memory and shared secondary memory) or tightly coupled (shared memory) multiprocessor system.
- A database system which resides at one of the nodes of a network of computers - this is a centralized database on a network node

Centralized Database Environment

Centralized database environment over network is given in figure below. In centralized database environment database is placed at one server and the remaining will be client and data will be accessed from single place for all clients.

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

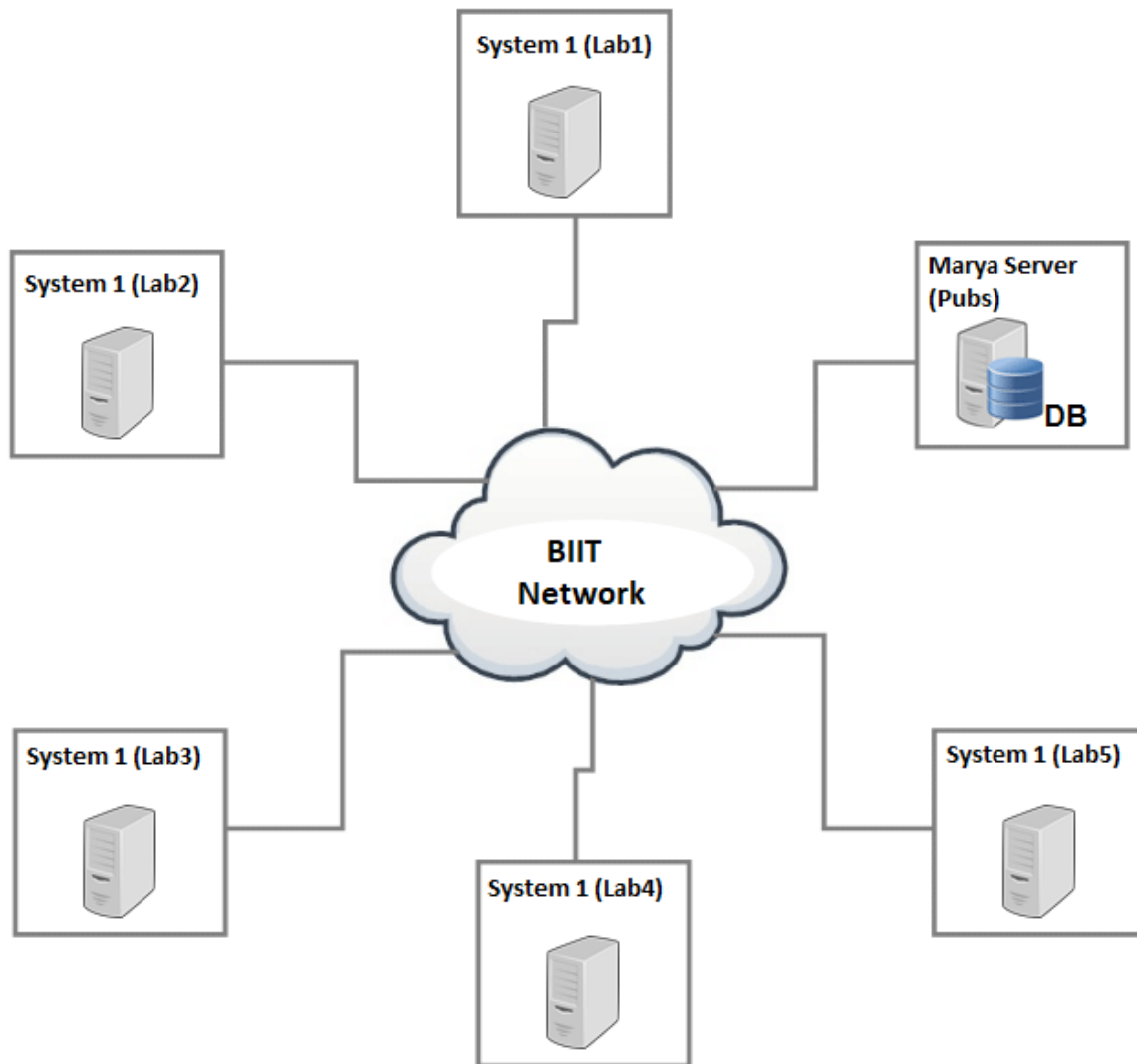


Figure 4

- Above figure shows an example of centralized database where database is placed at a single location and all users in system access data from the main site.
- Here we have discussed Database server of BIIT whose name is “Marya” and is placed at a system in Lab1.
- Marya server has a database named “Pubs” which is normally used by all students.
- We can access “Pubs” database from any of the Labs (Lab1, Lab2, Lab3, Lab4, Lab5, Lab6, Lab7) in BIIT

Advantages

Some advantages of Centralized Database Management System are –

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

- The data integrity is maximised as the whole database is stored at a single physical location. This means that it is easier to coordinate the data and it is as accurate and consistent as possible.
- The data redundancy is minimal in the centralised database. All the data is stored together and not scattered across different locations. So, it is easier to make sure there is no redundant data available.
- Since all the data is in one place, there can be stronger security measures around it. So, the centralised database is much more secure.
- Data is easily portable because it is stored at the same place.
- The centralized database is cheaper than other types of databases as it requires less power and maintenance.
- All the information in the centralized database can be easily accessed from the same location and at the same time.

Disadvantages

Some disadvantages of Centralized Database Management System are –

- Since all the data is at one location, it takes more time to search and access it. If the network is slow, this process takes even more time.
- There is a lot of data access traffic for the centralized database. This may create a bottleneck situation.
- Since all the data is at the same location, if multiple users try to access it simultaneously it creates a problem. This may reduce the efficiency of the system.
- If there are no database recovery measures in place and a system failure occurs, then all the data in the database will be destroyed.

Distributed DBMS Environment

Distributed database environment over network is given in figure below. In distributed database environment database is placed at more than one sites and data will be accessed from multiple places for all clients.

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

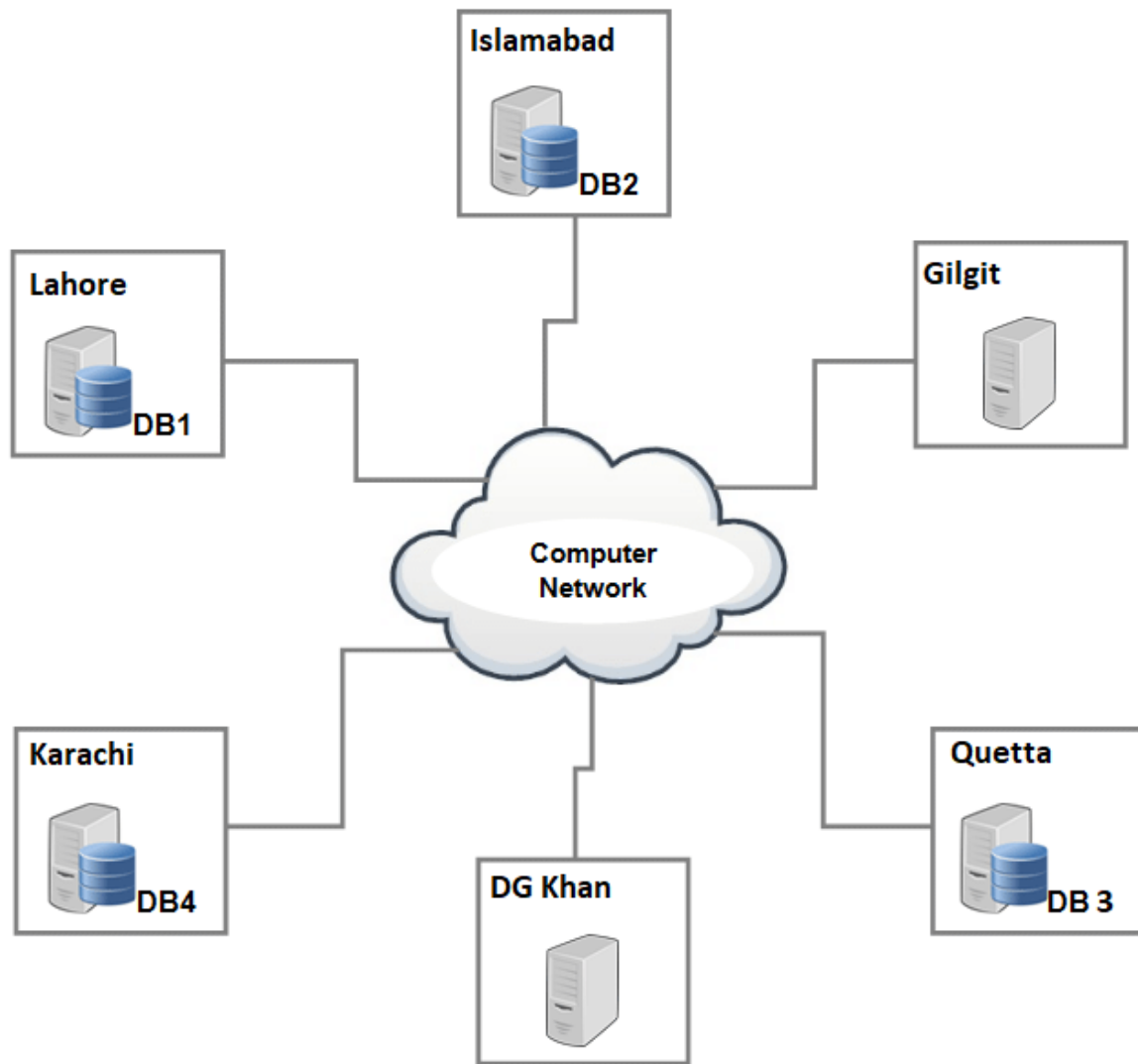


Figure 5

- In above example, we can see that there is a telecom organization which have servers at multiple locations.
- Important thing to notice here is there is not a single database. There are multiple databases placed at different locations (Islamabad, Lahore, Karachi, Quetta) over network.
- So that data can be accessed from multiple sites if one of the site is not available.
- We can see that all sites do not have Database placed.

Advantages of DDBMS

- The database is easier to expand as it is already spread across multiple systems and it is not too complicated to add a system.

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

- The distributed database can have the data arranged according to different levels of transparency i.e data with different transparency levels can be stored at different locations.
- The database can be stored according to the departmental information in an organization. In that case, it is easier for an organizational hierarchical access.
- If there were a natural catastrophe such as fire or an earthquake all the data would not be destroyed it is stored at different locations.
- It is cheaper to create a network of systems containing a part of the database. This database can also be easily increased or decreased.
- Even if some of the data nodes go offline, the rest of the database can continue its normal functions.

Disadvantages of DDBMS

- i. The distributed database is quite complex and it is difficult to make sure that a user gets a uniform view of the database because it is spread across multiple locations.
- ii. This database is more expensive as it is complex and hence, difficult to maintain.
- iii. It is difficult to provide security in a distributed database as the database needs to be secured at all the locations it is stored. Moreover, the infrastructure connecting all the nodes in a distributed database also needs to be secured.
- iv. It is difficult to maintain data integrity in the distributed database because of its nature. There can also be data redundancy in the database as it is stored at multiple locations.
- v. The distributed database is complicated and it is difficult to find people with the necessary experience who can manage and maintain it.

Data Delivery Alternatives

- In distributed databases, data are “delivered” from the sites where they are stored to where the query is posed.
- By defining Data Delivery Alternatives, it creates an overview of the system to be designed.
- We characterize the data delivery alternatives along three orthogonal dimensions.

- 1. Delivery Modes**
- 2. Frequency**
- 3. Communication Methods.**

The combinations of alternatives along each of these dimensions provide a rich design space.

1- Delivery Modes

- Delivery modes defines how data will travel from client to server.
- Will client request for the data?

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

- Will server share data with client without client's request?
- Or client can request data as well as server will also share data with clients without client's request.
- There are three types of delivery modes which are
 - i. **Pull-Only**
 - ii. **Push-Only**
 - iii. **Hybrid**

i. Pull-Only

In the pull-only mode of data delivery,

- the transfer of data from servers to clients is initiated by a client pull.
- When a client request is received at a server, the server responds by locating the requested information.
- The main characteristic of pull-based delivery is that
 - the arrival of new data items or updates to existing data items are carried out at a server without notification to clients unless clients explicitly pull the server.
 - servers must be interrupted continuously to deal with requests from clients.
 - the information that clients can obtain from a server is limited to when and what clients know to ask for.
 - Conventional DBMSs offer primarily pull-based data delivery

Examples

- Checking balance by customer of telecom company like Ufone
 - In this example clients initiate request and server responds
- Querying on Google
- Let us say a School system has a database. Parents are given a mobile application by which parent can check attendance status of their children by clicking a button. This is an example of Pull-Only where attendance will be only displayed when client will request for it.

ii. Push-Only

In the push-only mode of data delivery,

- the transfer of data from servers to clients is initiated by a server push in the absence of any specific request from clients.
- The main difficulty of the push-based approach is in deciding which data would be of common interest, and when to send them to clients – alternatives are periodic, irregular, or conditional.
- Thus, the usefulness of server push depends heavily upon the accuracy of a server to predict the needs of clients.

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

- In push-based mode, servers disseminate information to either an unbounded set of clients (random broadcast) who can listen to a medium or selective set of clients (multicast), who belong to some categories of recipients that may receive the data

Examples

- Updates of windows to users
- Updates of anti-viruses to users
- Packages updates to customers from telecom companies

iii. Hybrid

The hybrid mode of data delivery combines the client-pull and server-push mechanisms

- one possible way of combining the pull and push modes: namely, the transfer of information from servers to clients is first initiated by a client pull (by posing the query), and the subsequent transfer of updated information to clients is initiated by a server push.

Examples

- User subscribes some services from Telecom Company (news alert) and then server sends in future without any query from user.
- Let us say School system has enhanced their database and now it performs following operations.
 - a. Send monthly report to parents (without parent's request)
 - b. Parents can check attendance status of their children.

Here we can see, a Hybrid mode of Data delivery is used where Push-Only and Pull-Only are both integrated in the system.

2- Frequency

There are three typical frequency measurements that can be used to classify the regularity of data delivery.

- Periodic**
- Conditional**
- Ad-hoc or irregular**

i. Periodic Frequency

In periodic delivery,

- Data are sent from the server to clients at regular intervals.
- The intervals can be defined by system default or by clients using their profiles.
- Both pull and push can be performed in periodic fashion.
- Periodic delivery is carried out on a regular and pre-specified repeating schedule.

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

- Periodic push is particularly useful for situations in which clients might not be available at all times, or might be unable to react to what has been sent, such as in the mobile setting where clients can become disconnected.

Examples

- A Telecom Company's customer request for daily Hadith e Pak ﷺ.
- A client request for IBM's stock price every week is an example of a periodic pull.
- An example of periodic push is when an application can send out stock price listing on a regular basis, say every morning.

ii. Conditional Frequency

In conditional delivery,

- data are sent from servers whenever certain conditions installed by clients in their profiles are satisfied.
- Such conditions can be as simple as a given time span or as complicated as event-condition-action rules.
- Conditional delivery is mostly used in the hybrid or push-only delivery systems.
- Using conditional push, data are sent out according to a pre-specified condition, rather than any particular repeating schedule.
- Conditional push assumes that changes are critical to the clients, and that clients are always listening and need to respond to what is being sent. Hybrid conditional push further assumes that missing some update information is not crucial to the clients.

Examples

- An application that sends out stock prices only when they change is an example of conditional push.
- An application that sends out a balance statement only when the total balance is 5% below the pre-defined balance threshold is an example of hybrid conditional push.
- Message from Telecom Company to customer on low balance.
- An email from Google when user login from a new device.
- A message from our bank whenever amount is credited or debited in our account.

iii. Ad-hoc or Irregular

- Ad-hoc delivery is irregular and is performed mostly in a pure pull-based system and often in push-based system.
- Data are pulled from servers to clients in an ad-hoc fashion whenever clients request it.
- In contrast, periodic pull arises when a client uses pulling to obtain data from servers based on a regular period (schedule).

Distributed Database System (CS-600)

Mr. Nauman Iqbal
Mr. Ahsan

email id: nauman@biit.edu.pk Whatsapp# 0321-5821151
email id: ahsan@biit.edu.pk, Whatsapp# 0333-5189656

Examples

- User query on Google in irregular fashion (pull model)
- Telecom Company send updates in irregular fashion (push model)
- Telecom Company's send news alert on user's request in irregular fashion (hybrid model)

3. Communication Methods

These methods determine the various ways in which servers and clients communicate for delivering information to clients. The alternatives are

- i. **Uni-Cast**
- ii. **One-To-Many**

i. **Uni-cast**

In unicast,

- the communication from a server to a client is one-to-one;
- the server sends data to one client using a particular delivery mode with some frequency.

Example:

- A message from bank to client in case of amount credit or debit. This message is send to a specific account holder.
- A message from a mobile company in case of balance is low.
- A message from university to a student whose fee is not paid yet.

ii. **One-to-many**

In one-to-many, as the name implies,

- the server sends data to a number of clients.
- Note that we are not referring here to a specific protocol; one-to-many communication may use a multicast or broadcast protocol.

Example:

- A message from a mobile company to many users informing about new package.
- A message from a School system to all parents of 10th grade about farewell party.

X-----X