



Department of Information Technology Central University of Kashmir

Tullmulla, Ganderbal, J&K-191131
www.cukashmir.ac.in

MTIT C 203: Advanced Database Management System



Sub Topic(s):

EMERGING TRENDS IN DATABASES: MOTIVATION, TEMPORAL DATABASES, TIME SPECIFICATION IN SQL, MULTIMEDIA DATABASES, CONTINUOUS MEDIA DATA , DATA RETRIEVAL, CONTENT DELIVERY NETWORKS, SIMILARITY-BASED RETRIEVAL , MOBILITY AND PERSONAL DATABASES, MODEL OF MOBILE COMPUTING, MODEL OF MOBILE COMPUTING, ROUTING AND QUERY PROCESSING, BROADCAST DATA, DIS-CONNECTIVITY AND CONSISTENCY, AVOIDING INCONSISTENCY USING VERSION VECTOR SCHEME, VERSION VECTOR SCHEME, SPATIAL AND GEOGRAPHIC DATA, REPRESENTATION OF GEOMETRIC INFORMATION, NEARNESS QUERIES, GEOGRAPHIC DATA AND APPLICATIONS.

Course Title **ADBMS**
Course Code: **MTIT C 203**
Unit: **5**
Department: **Department of IT**
Year: **2020**

Compiled by: Aabiroo Bader
Email: baderabiro@gmail.com
Contact: 6005655573
Designation: Teaching Assistant
Department: Department of IT

Index

EMERGING TRENDS IN DATABASES

1) MOTIVATION.....	4
2) TEMPORAL DATABASES.....	5
VALID TIME.....	6
TRANSACTION TIME.....	7
3) TIME SPECIFICATION IN SQL.....	7
4) MULTIMEDIA DATABASES.....	8
5) CONTINUOUS MEDIA DATA.....	9
6) DATA RETRIEVAL.....	10
7) CONTENT DELIVERY NETWORKS.....	10
8) SIMILARITY-BASED RETRIEVAL.....	10
9) MOBILITY AND PERSONAL DATABASES.....	11
10) MODEL OF MOBILE COMPUTING.....	11
11) ROUTING AND QUERY PROCESSING.....	12
12) BROADCAST DATA.....	13
13) DIS-CONNECTIVITY AND CONSISTENCY.....	13
14) AVOIDING INCONSISTENCY USING VERSION VECTOR SCHEME.....	13
15) VERSION VECTOR SCHEME.....	14
16) SPATIAL AND GEOGRAPHIC DATA.....	14
17) REPRESENTATION OF GEOMETRIC INFORMATION.....	15
18) GEOGRAPHIC DATA AND APPLICATIONS.....	15
1) Raster Data	

2) Vector Data

19) NEARNESS QUERIES.....16

EMERGING TRENDS IN DATABASES

[1] MOTIVATION

Earlier Era - For most of databases, the type of data stored in databases were relatively simple.

–Mobility has changed everything – Given the growth of mobile computers, the dimension of data to be stored and retrieved from databases have changed.

–To cater the changing environments of having to store multidimensional and complex datatypes in databases, **Temporal Databases**, **Spatial Databases** and **Multimedia Databases** came to exist.

➤ TEMPORAL DATA

•Most of the databases model the current state of the world – for instance Current Customers, Current Students and Courses being offered.

•In many of the applications, it is very important to store and retrieve information about past states. This task is greatly simplified by a database that supports temporal data.

➤ SPATIAL DATA

•Spatial data includes geographic data, such as Maps, Computer-Aided design such as Integrated circuit design or building designs.

•Applications of spatial data were initially stored as files in a file system. But as the complexity of and volume of data plus the number of users that have grown, adhoc approaches to storing and retrieving data from a file system have proved insufficient for needs of many applications that use spatial data.

➤ MULTIMEDIA DATA

•Multimedia data are images, video and audio data. The distinguishing feature of video and audio data is that display of data requires retrieval at steady, predetermined rate – hence Such data is known as Continuous-Media data.

•By continuous-media data, we mean there has to be a continuous synchronization relationship between Source and Sink – plus Audio and Video.

[2] TEMPORAL DATABASES

A **temporal database** is a database with built-in-support for handling data involving time.

- Typically, databases model only one state – the current state of the real world and doesn't store information about past states.

- When state of the real world changes, the database gets updated and information about old state gets lost. However, it is also important to store and retrieve information about current and past states.

–Examples:

- Patient database must store information about the medical history of patient.
- Judicial records.
- Various sensory information.

So we define a Temporal database – “Database that stores the states of real world across time”.

- Temporal aspects in databases usually include:

- Valid Time.
- Transaction Time.
- Bi-temporal Data.

Temporal Relation

<i>account-number</i>	<i>branch-name</i>	<i>balance</i>	<i>from</i>	<i>to</i>
A-101	Downtown	500	1999/1/1 9:00	1999/1/24 11:30
A-101	Downtown	100	1999/1/24 11:30	*
A-215	Mianus	700	2000/6/2 15:30	2000/8/8 10:00
A-215	Mianus	900	2000/8/8 10:00	2000/9/5 8:00
A-215	Mianus	700	2000/9/5 8:00	*
A-217	Brighton	750	1999/7/5 11:00	2000/5/1 16:00

Bi-temporal Relation

It is also true that we can store both Valid Time and Transaction Time in the databases. Such a relation involving both Transaction Time and Valid Time is known as Bi-Temporal Relation.

ID	Name	Dept.	Salary	Transaction Time	Valid S.T	Valid E.T
101	Asif	CSE	12345	2016	2009/1/1	2013/12/31
202	Amir	ECE	12346	2015	2008/1/1	2012/12/31
303	Asim	IT	12121	2016	2009/1/1	2014/12/31

❖ **STORY**

- Story of Mr. X....
- Born on 3rd April 1971 at Pulwama.
- Father of Mr. X registers D.O.B on 4th April , 1971 at Pulwama.
- Mr. X completes his graduation on August,1990.
- For Job purposes Mr. X goes to Srinagar on August, 1991, but forgets to register his new address officially.
- It was on December the 25, 1991 he registers the new address officially.
- Unfortunately, Mr. X was accidently hit by a speedy car on April 1 2001.
- The coroner reports the date of death on the very same day.

- Story to be continued.....

➤ **DIFFERENCE OF STORING DATA**

- If we store this information of Mr. X in a Current Database what will be the Relation like??
- And if we store same information in a Temporal database what will be the Relation like???

CURRENT DATABASE

Date	Event	Database Action	Database Output
3 April 1971	Mr. X was born	Nothing	No Person known as X
4 April 1971	Official reports d.o.b	Insert(X, Pulwama)	X lives in Pulwama
15 August 1991	X moves to SXR, but forgets to register	Nothing	X lives in Pulwama
25 December	X registers new address	Update(X, SXR)	X lives in SXR
1 April 2001	X dies	Delete (X)	No Person

I. VALID TIME

- A valid time is a time for which a fact is true in the real world.
- This time period may be in the past, or span the current time.

Since the official recording the birth doesn't know if X will move to some other place or when he will move. In this case Valid To is set to *;

- Person (X, Pulwama, 1971/04/03, *);
- Person (X, SXR, 1991/12/25, *);
- The original entry isn't deleted instead state of database is :
- Person (X, Pulwama, 1971/04/03, 1991/12/25);
- Person (X, SXR, 1991/12/25, 1991/12/25, *);
- When Mr. X dies, the database looks like:
- Person (X, Pulwama, 1971/04/03, 1991/12/25);
- Person (X,SXR, 1991/12/25, 2001/04/01);

II. TRANSACTION TIME

- Transaction Time records the time period during which a database entry is accepted as correct.
- Valid time doesn't record any event that wasn't made public even though at later stages, for some auditing purposes, transaction time will note the record if found true.
- Suppose Mr. X moved to Dubai from 1 June 1997 to 1 June 2000. But to avoid increased taxations, he didn't report it to authorities. On Feb. 2001 it was in fact discovered that Mr. X lived in Dubai for so many years.
- Transaction Time will allow capturing this changing information in the database.
- Transaction Time takes into account each entry record when it was entered and when it was superseded.

TEMPORAL DATABASE

Name	City	Valid From	Valid To	T.T Entered	T.T Supersede
X	Pulwama	1971/04/03	*	1971/04/03	1991/12/25
X	Pulwama	1971/04/03	1991/12/25	1991/12/25	*
X	SXR	1991/12/25	*	1991/12/25	2001/02/01
X	SXR	1991/12/25	1997/06/01	2001/02/01	*
X	Dubai	1997/06/01	2000/06/01	2001/02/01	*
X	SXR	2000/06/01	*	2001/02/01	2001/04/01
X	SXR	2000/06/01	2001/04/01	2001/04/01	*

[3] TIME SPECIFICATION IN SQL

- **date**: four digits for the year (1--9999), two digits for the month (1--12), and two digits for the date (1--31).
- **time**: two digits for the hour, two digits for the minute, and two digits for the second, plus optional fractional digits.
- **timestamp**: the fields of **date** and **time**, with six fractional digits for the seconds field.

[4] MULTIMEDIA DATABASES

Multimedia data are – Images, Audio and Video – They are the most popular form and increasingly preferred data these days.

•One approach of storing multimedia data is in **File Systems**.

–We know the size of this type of media is large, in fact very large, they were generally stored outside the database in file systems.

–General database features – Transactional Updates and Querying facilities are seriously dented when Multimedia object size is very large.

•**Second approach**

–Each media file has some descriptive attributes – File/Data Statistics that include, when a particular object is created, who created it and to which category it belongs.

–This approach tells that we will use a database for storing the descriptive attributes and keeping track of files in which the multimedia objects are stored.

❖ **Drawbacks**

• What if that object is missing or corrupted and database still points the location - multimedia data existed.

–The better approach is to store the multimedia data in database itself.

➤ **Issues that need to be addressed –**

•Databases must support Large Objects

–Given the size of multimedia data objects are in GBs, many database objects don't support objects larger than a few GBs.

–So larger objects must be split into smaller pieces and stored in databases.

–Alternately, the multimedia object may be stored in a file system but database may contain a pointer to the object – pointer being essentially a file name.

SQL/MED (– Management of External Data) is a standard that allows external data such as files to be treated as if they are part of database.

•Since the size of multimedia data is very large, it is essential that multimedia data is stored and transmitted in compressed form.

•Image is widely stored in **JPEG**

- a) **JPEG – Joint Pictures Expert Group.**
- b) **MPEG – Moving Pictures Expert Group**, developed the MPEG series of standards for encoding video and audio data.

MPEG format achieves high degree of compression.

❖ **MPEG-MOVING PICTURE EXPERTS GROUP**

- ✚ **MPEG-1** stores a minute of 30-frame per second video and audio in approximately 12.5 MBs; compared to **JPEG** that stores same length of data at 75 MBs.

This huge difference is due to the fact the successive frames in **JPEG** are nearly same whereas MPEG exploits commonalities among sequence of frames.

•**MPEG-1** experiences Loss of data – Quality of video isn't good. It is comparable to VHS videotape.

- ✚ **MPEG-2** standard is designed for digital broadcast systems and DVDs.

•**MPEG-2** experiences negligible loss of video quality.

•1 minute video is approximately compressed to 17 MBs.

- ✚ **MPEG-4** provides techniques for further compression of video with variable bandwidth to support delivery of video data over networks.

[5] **CONTINUOUS MEDIA DATA**

•Most important types of Continuous media data are Video and Audio data.

•Continuous Media Systems are characterised by their **Real Time Information – Delivery requirements.**

•By Continuous Media Systems we mean data must be delivered smoothly i.e No gaps in the Audio or Video.

•Data must be delivered at a rate that doesn't cause overflow of system buffers.

•Synchronization among distinct data streams must be maintained – Lip Synchronization of Audio and Video.

[6] DATA RETRIEVAL

- Usually data are fetched in periodic cycles. Each cycle may consist of some time period – say ‘n’ seconds.
- Data will be fetched from database and stored in **Memory Buffers**.
- This stored data will be sent to consumer’s display unit.
- Finally the display unit displays the content.
- Time Period
- If Time Period/Periodic Cycle is small – More disk arm movements are required i.e.. Wastage of resources.
- If Time Period/Periodic Cycle is large – Large memory buffer requirement plus large initial delay.
- **Admission Control:** When a new request arrives, admission control comes into play, i.e System checks if request can be satisfied with the available resources – if so it is admitted otherwise it is rejected.

[7] CONTENT DELIVERY NETWORKS

- Content Delivery Networks **provide content on demand**.
- It comprises of:
 - a. **Video Server:** Multimedia data are stored on several disks – RAID configuration or even Cloud Storage.
 - b. **Terminals:** People view multimedia through various devices known as Terminals – PCs, TVs, Setop-boxes.
 - c. **Network:** Transmission of multimedia data from a server to multiple terminals requires a high capacity network.

[8] SIMILARITY-BASED RETRIEVAL

a. **Pictorial Data:**

- Two pictures or images that are slightly different as represented in the database may be considered the same by a user. When a new trademark is to be registered, the system may need to first to identify all similar trademarks that were registered previously.

b. Audio Data:

–Speech-recognition interfaces have been developed that allow the user to give a command or identify a data item by speaking. The input from user must be then tested for similarity to the commands stored in the system.

c. Handwritten Data:

–Signatures are being created to authenticate a particular customer for Bank Account Validations.

[9] MOBILITY AND PERSONAL DATABASES

➤ **Personal Databases** have come to exist due to the fact that:

–Widespread use of Laptop, Notebooks and PCs.

–Widespread use of Cell Phones with the capabilities of a computer.

➤ **Mobile Computing**

–A technology which we avail every second in today's world.

–Wireless Technology has created a situation where machines no longer have fixed locations and Network Addresses.

➤ **MOBILITY**

Location-dependent Queries are a different class of queries, in which location of the user (computer) is a parameter of the query.

•Value of the location parameter is provided by a **Global Positioning System (GPS)**.

–**Example** Makemytrip and Goibibo Hotels, that involve providing data on Hotels, Roadside Services to the users travelling a particular destination.

•Processing of queries about services ahead on the current route has certain constraints – like **Direction of Motion, Speed** of a particular user.

•In mobility the Energy (Battery Lifetime) is a scarce resource. This has a great influence on System Design Architecture, Impact on protocols used to communicate with mobile devices.

[10] MODEL OF MOBILE COMPUTING

•Mobile-Computing environment consists of mobile computers known as Mobile Hosts.

- The Mobile Hosts communicate with the wired network via computers known as Mobile Support Stations.
- Cell: is a particular geographical area that a particular Mobile Support Station cover and support.
- Handoffs are important to support mobility services among users.
- It is also possible for mobile hosts to communicate directly without the intervention of mobile support station. This is possible only for short range communications like **Bluetooth**. Range of 10 m and a speed of 721 Kbps.

Bluetooth, Wireless LANs and 2.5G and 3G cellular networks make it possible for a wide variety of devices to communicate at low cost. The accounting, monitoring and management data pertaining to this communication generate huge databases.

- To improve on the energy efficiency side of mobile devices it is always a preferable method to use low power Flash Memories and power down a component if that is not in use for a while.
- WAP – Wireless Application Protocol is a standard protocol for Wireless Internet access that take into measure constraints of mobile and Wireless Web browsing.

[11] ROUTING AND QUERY PROCESSING

- Routing** is a method for obtaining a best path between source and destination, but the major constraint is the location of users are no longer constant.
- Mobility results in dynamically changing communication costs, thus complicating the optimization process.
- Other factors that need to be considered:
 - User Time:** User Time is a highly valuable commodity in many business applications.
 - Connection Time:** is the unit by which monetary charges are assigned in some cellular systems.
 - Number of bytes, or packets, transferred:** is the unit by which charges are computed in some digital cellular systems.
 - Energy:** is a scarce resource whose use must be optimized. Use of Sleep Scheduling algorithms and data aggregation algorithms will help in improving battery life of mobile devices.

[12] BROADCAST DATA

- It is often a necessary to broadcast data periodically rather a set of users making explicit requests. Example is Stock Market Data and Cricket Live Scores.

- Broadcasting data has two advantages:

- Transmitting Cost of data request from Mobile User side is avoided.

- A group of users has a same requirement and they avail this facility either at minimum possible cost or totally Free.

- Another approach is to **cache** the broadcast data on a local inbuilt mobile storage and a user request may be fulfilled locally.

- If the cached data is insufficient to there are two options

- Wait for data to be broadcasted..

- An explicit request for new data.

[13] DIS-CONNECTIVITY AND CONSISTENCY

- To avoid an increased charges for **mobile connection time**, a general strategy we follow is, we keep our data connection OFF – Technically we are disconnected to the network.

- During this period of disconnection, the mobile host may remain in operation and will definitely issue queries and updates on the data that reside or are cached locally.

- Problems that result because of this:

- Recoverability**: Updates entered on a disconnected machine may be lost if the mobile host experiences a catastrophic failure. Since recoverability options being bleak, all local updates are lost.

- Consistency**: Being disconnected for a while, consistency issues are common as data in cache may become out of date, and host can only discover this fact, when it makes a reconnection.

[14] AVOIDING INCONSISTENCY USING VERSION VECTOR SCHEME

- The data entered by a mobile host on a local cache during the course of **disconnection** can be flooded to other hosts when mobile host makes a reconnection.

- But the problem arises if a particular host has entries in the cache are **Read only** and those entries are updated by other hosts.

–When a mobile host is reconnected a report known as Invalidation Reports may be sent to that particular host informing it of out-of-date cache entries.

–A more simple solution was to **invalidate the whole cache when a reconnection occurs** but it will be at higher costs.

–A more convenient solution is a Version Numbering known as Version Vector Scheme.

[15] VERSION VECTOR SCHEME

•Version-Vector scheme detects inconsistencies when copies are independently updated. This scheme allows copies of a document to be stored at multiple hosts.

•Basic idea: each host i stores, with its copy of each document d , a **version vector** - a set of version numbers, with an element $V_{d,i}[k]$ for every other host k

•When a host i updates a document d , it increments the version number $V_{d,i}[k]$ by 1

• When two hosts i and j connect to each other they check if the copies of all documents d that they share are consistent:

1. If the version vectors are the same on both hosts (that is, for each k , $V_{d,i}[k] = V_{d,j}[k]$) then the copies of d are identical.
2. If, for each k , $V_{d,i}[k] < V_{d,j}[k]$, and the version vectors are not identical, then the copy of document d at host i is older than the one at host j

•That is, the copy of document d at host j was obtained by one or more modifications of the copy of d at host i .

•Host i replaces its copy of d , as well as its copy of the version vector for d , with the copies from host j .

[16] SPATIAL AND GEOGRAPHIC DATA

•Spatial data support in databases is important for efficiently storing, indexing and querying of data on the basis of spatial locations.

•Two types of spatial data are particularly important:

- i. **Computer-aided-design (CAD) data:** includes spatial information about how objects - such as buildings, cars or aircraft – are constructed. Other examples that include computer-aided-design databases are integrated-circuit and electronic-device layouts.

- ii. **Geographic Data:** Such as road maps, land-usage maps, topographic elevation maps, political maps showing boundaries, land ownership maps, and so on.

Geographic information systems are special purpose databases tailored for storing geographic data.

[17] REPRESENTATION OF GEOMETRIC INFORMATION

Geometric Information can be represented in several ways:

–A **line Segment** – represented by the coordinates of its endpoints. Example – For a map database the two coordinates of a point would be latitude and longitude.

–A **polyline** consists of a connected sequence of line segments in sequence and can be represented by a list containing the coordinates of the endpoints of the segments, in sequence.

•This representation of curves by polylines, when we partition the curve into sequence of segments is useful for two-dimensional features such as Roads. At the same time some systems use circular arcs to represent arcs or curvatures of Roads on the maps.

– We can represent a **polygon** by listing its vertices in order. A polygon can be divided into a set of triangles by a process known as **Triangulation**.

[18] GEOGRAPHIC DATA AND APPLICATIONS

Geographic data are spatial in nature, and are of the form of Maps and Satellite images.

Maps in particular not only provide location information like boundaries, rivers and roads but also a detailed information like Locations, Elevations, Soil Type, Land Usage and Annual Rainfall.

•Applications

–Use in Online Map Services, Vehicle Navigation System.

–Distributed-Network Information for Public-service utilities such as

•Electric Power and Water Supply Systems.

•Land Usage information for ecologists and planners.

Geographic data is categorized into two types:

3) Raster Data

Raster data consists of bit maps or pixel maps, in two or more dimensions. **Best example – Satellite image of an area** that not only includes the actual image but information like latitude and longitude of its corners.

- Raster data is often represented as tiles, each covering a fixed size area. A larger area can be displayed by displaying all the tiles overlapping that area.
- To allow the display of a data at different zoom levels, a separate set of tiles is created for each zoom level. Once the zoom level is set by the user, tiles at that zoom level, which overlap the area being displayed are retrieved and displayed.
- Raster data can be 3-dimensional.

4) Vector Data

– Vector data are constructed from basic geometric objects such as Points, Line Segments, Polylines, Triangles and other polygons in 2-D and Cylinders, Spheres, Cuboids are 3-D data.

– Map data are often represented in vector format. Roads being represented as Polylines.

– Geographic features such as large lakes, states and countries are represented as complex polygons., while as rivers may be represented as complex curves.

- Topographical information, i.e information about the elevation of each point on surface can be represented in raster form.

[19] Nearness Queries

A nearness query request objects that lie near a specified location. A query to find all ATMs that lie within a given distance is an example of a nearness query.

- The nearest neighbour query requests the object that is nearest to a specified point.

– Example – We want to find a nearest Recharge Point.

- Region Queries: Region queries deal with spatial regions. Such a query can ask for objects that lie partially or fully inside a specified region.

– Example – A query to find all retail shops within the geographic boundaries of a given town.

References & Bibliography

- Silberschatz–Korth–Sudarshan, “Database System Concepts”, The McGraw–Hill Companies.
- Ramez Elmasri, Shamkant B. Navathe, “Fundamental OF Database Systems”, Pearson.