Geographical Information Systems
Prof. A. K. Saraf
Department of Earth Sciences
Indian Institute of Technology – Roorkee

Lecture – 29
Spatial Database Systems and their types - 01

Hello everyone! and welcome to a new topic discussion which is spatial database systems and their types. So, this would be in two parts minimum. It is a big topic as you know that database management itself has become nowadays a big subject. But we will be mainly focusing on that part of database management systems which is related directly with GIS, not a generic database management system.

Otherwise, it would be a very long completely separate discussion also very. As you know that if we compare with simple file systems which generally, we follow in our computers, laptop or desktop, we go for file systems. But if we organise the data in databases then we can get lot of advantages. All in GIS as you know that two types of data; one is the spatial data, another one is the non spatial data.

So, non spatial data is organised in a database and overall, we say it's a spatial database. So, a spatial data when we mention, it means that we are having not only spatial information like point, line, polygon, raster, TIN etcetera but we are also having attribute information organised in our database. So, the conventional database management systems; though they are in the background for us but major difference is that we are having good connection or dynamic linkage between the conventional database systems and spatial information also.

And obviously, this is done through your geographic coordinates that allows us to handle this thing. So, in this discussion, we are going to have very briefly what is a database and how it is designed?

(Refer Slide Time: 02:30)

Why it is important for the GIS users and how modern databases in evolved from files to file systems? And then about flaws in file systems means lacunas and what is database management system and how it fits into our GIS database system and different types of databases models which exist so far? And mainly, we will discuss again as I am repeating related with GIS especially like relational database management system or RDBMS is the most common database system which has been implemented in GIS and most useful also.

But databases like network and others are also being used nowadays extensively. So, when I was discussing the network analysis, at that time, I mentioned that network database is having linkage with for network analysis. So, if you are having your data in that structure then network analysis can be performed in a much easier manner.

(Refer Slide Time: 03:44)

Anyway, so first as you know the data versus information and some might be wondering, what is basically data? So, data are the raw facts basically. And these data, we are organised into a computer or in a database and convert that data or raw facts into information. And GIS obviously goes one step ahead. Through GIS analysis, this information can be converted to knowledge.

So, GIS is capable of handling not only raw facts that is data, it also allows us to convert that raw facts or data into information and finally to knowledge; what is there, how it is there, how it will become after such changes? And that basically is nothing but the knowledge. And also, through the prediction or modelling; we get lot of insight about data. So basically, as you know that information is generated through data so it's a building blocks of information.

And information is produced by processing data, that can be done easily. If it is attribute data, it can be done very easily in a conventional database. But if it is a GIS data or spatial data then it can be only done on a GIS platform. And when we process the data then you know these raw facts can reveal quite good information. And of course, when we do this processing in GIS, we can convert finally data into knowledge.

So, good, timely and relevant information are the key for decision making. So, good and timely means you are having not only the quality of the data but very updated information; updated data and relevant data. And that will allow us to process it as quickly as possible; convert to information, convert to knowledge and when we present these scenarios to the decision makers, they can make quick decisions.

So, good decision-making as you know is the key to organizational survival or success and everyone would like to have a very nice database management system in their organization. So that whenever they want that information or that data immediately, they get it in an organised fashion. And also, this has to be updated; the timely and quality data is most important in this.

(Refer Slide Time: 06:02)

So, basically what we do in database management is data which is shared, integrated computer structure which houses the end user data and also a metadata. Metadata is the data about the data; that means the information about the data is called metadata. Even in case of satellite images, generally when you download a free satellite image, you will find a file which is .met.

That does not mean it is having meteorological data, that means it is having the metadata; the information about that particular satellite image. When it was acquired? Which was the satellite? Which was the sensor? Who has done what kind of processing so far? All that information is recorded in the metadata. And it is a very good practice of good organisations to record everything about the data.

Who has generated the data? How much processing has been done? When it was generated? So, all that information can go as metadata. Now in this database management system, what we do? We manage the database structure. How? Because there are different structure or different models are there. So, this is first task which is done in database management system. Also, it controls access to data.

Who will have the access? How much he will have the access? Or when he will have the access? All that can be controlled by the administrator of the database management system. Also, it should allow us or users to query in a very simple manner. I want to access the data to certain purposes. So, I should be allowed to get certain information from the database only

through this query. So, this query language has to be also very simple and should be allowed also.

**(Refer Slide Time: 08:20)**
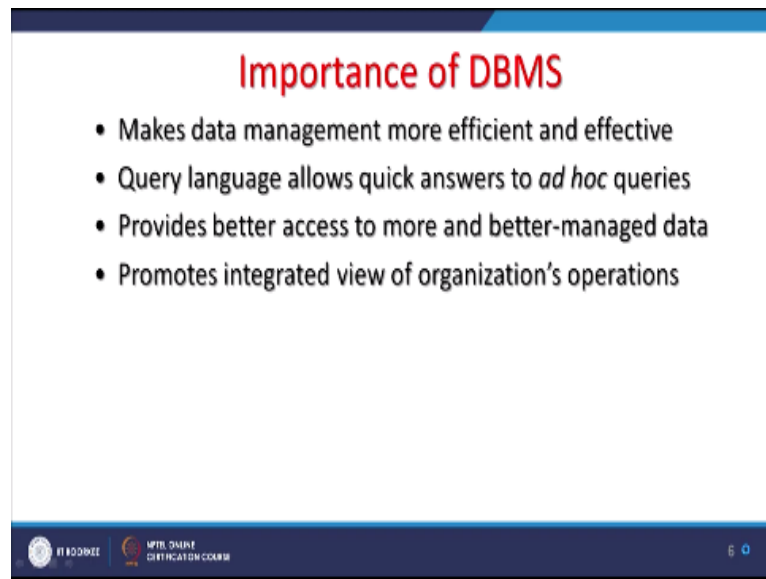
Now, when we discuss the importance of database that database management system makes data management more efficient and effective. Also, query language allows quick answers to ad hoc queries. Suddenly, you are having a question; you just raised that question through easy syntax and you get the answer from the database, if it is having that information obviously. And it provides the better access to more and better manage data.

Promotes integrated view of organisation's operations that how the organisation is working and you know, how it has been integrated within that. What it means basically that an organisation may be having good databases but lying at different places for different purposes. And there is no integration or there is no talking between two databases. And therefore, the query if I want, I cannot have an integrated or you know, nested query implying two different databases within that same organisation of two different divisions or sections.

So integrated way, it is also sometimes important. And if you are having a good database, it reduces the probability of inconsistent data. There will be a discussion also on islands of data. That means the data is lying but it is inaccessible or it is not getting updated; inconsistent. And then it becomes a confusion and maybe a useless information and unnecessary, it will occupy space in our computer systems.

(Refer Slide Time:10:14)
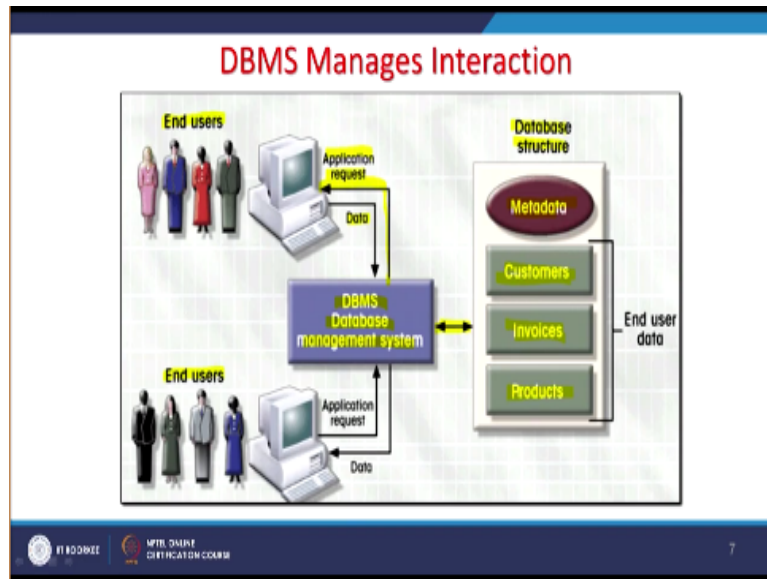
Now, when we discuss like database management system then what we find that there will be always end users maybe at different ends. And then they will have some way to interact with the database management system so application request. And then there is a flow of the data in 2 directions. The same way then this database management system will allow for the organisers and as per the chosen database structure.
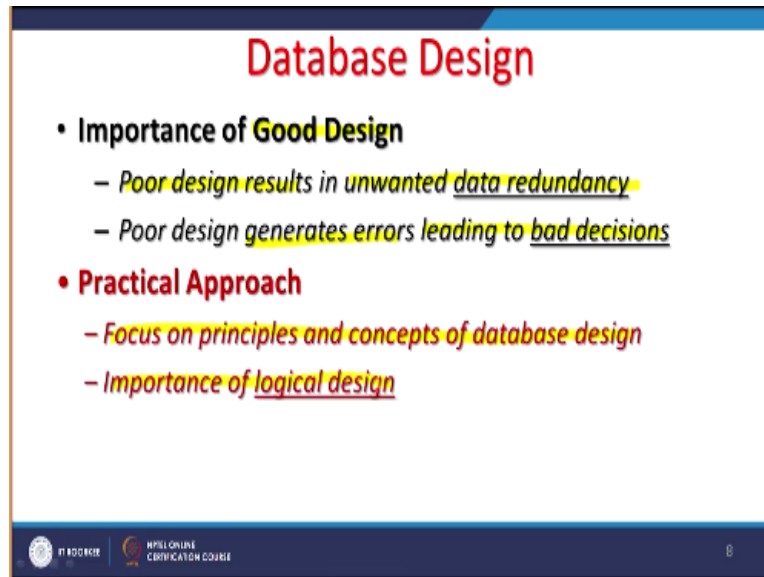
Of course, it will have metadata. The customers are being handled. Depending on the organisation, this example is from the sales organisation and then products. So, a company which is having say online company which is supplying certain items. So, likewise and again end user data is here. So, that is being maintained. It's a typical database management interaction which we see here.

I will give you an example at this stage and may discuss little later further. Many of us or let me say that lakhs of people are interacting with the IRCTC database system or IRCTC portal for booking our seat in a train on a particular day on a particular class. And you can think that how complicated that software is or that portal is. There are more than 6000 trains, may not be currently doing corona period.

But otherwise in normal conditions, more than 6000 passenger trains are running 24 hours having different classes, different categories and more than 2 crore people are travelling at a time in these trains. So, you can imagine that lakhs of people are booked or logged at the same time to book their seats on a train. So, that is also a very good example of a good design of a database.

Though, directly it is not related with GIS in that sense because geographic information is not handled in that IRCTC. But from database design and management, it is a very good product or a very good example. Because information is being updated; any change happens, it is implemented. Any category or any new train is introduced, that is implemented. So, it is a very good example of a good database system.

(Refer Slide Time: 12:53)

Now, that is why the good design is the key. Starting point has to be very good. If poorly designed database is there then you will have lot of data redundancy. Data redundancy means data is lying in database but it is not accessible to anyone and data is just occupying your storage space. And it is very bad example of poor design. Poor design also generates errors, sometimes leading to bad decisions.

Think that if IRCTC portal or software database is poorly designed or having some errors then that can create havoc in railway system but no. And if they find any error, obviously they go and immediately correct it. Otherwise, we will be booking a wrong seat in a wrong train on wrong date and that would be a disaster in the country because remember 2 crore people travelling all the time on 6000 trains all over the India.

So, good designs are very much important. Now what are the practical approaches? Focus is on the principles and concept of database design, that is the first task done. Then importance of logical design. Design has to be very logical. How data will flow?

(Refer Slide Time: 14:17)

Historical Roots of Database

- First applications focused on clerical tasks
- Requests for information quickly followed
- File systems developed to address needs
  - Data organized according to expected use
  - Data Processing (DP) specialists computerized manual file systems

Before that, we will also be discussed about the historical roots of the data. We started with the file system when we got introduced to computers. And slowly-2, we have started moving now on DBMS or database management systems. So, first application focused on the clerical tasks; simple task which a clerk used to do in the office. These were thought that let the computer do it.

Request for information quickly forwarded. And that was if somebody has asked information, let us share that information. So basically, we started with a file system which was developed to address the needs of like data organised according to expected use. And data processing specialists computerised manual file systems.

(Refer Slide Time: 15:09)

File Terminology

- *Data*
  - Raw Facts
- *Field*
  - Group of characters with specific meaning
- *Record*
  - Logically connected fields that describe a person, place, or thing
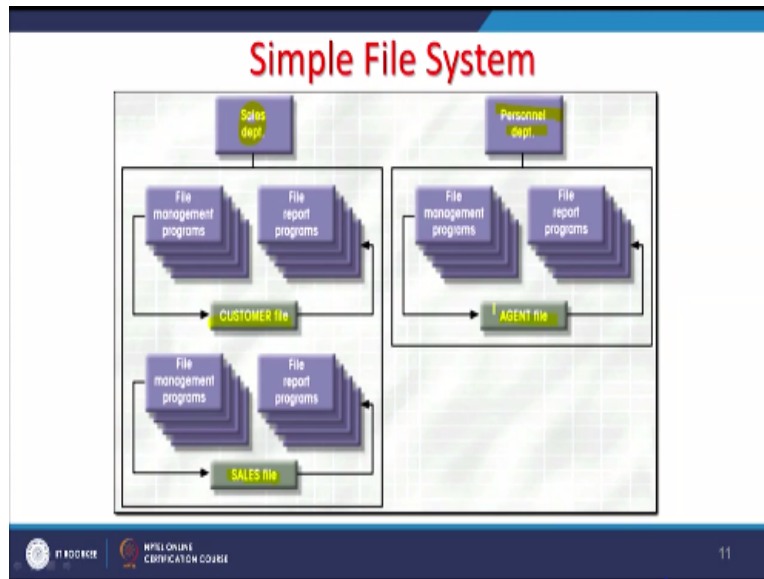- *File*
  - Collection of related records

Now, this terminology which started with the file system, still many terms continue in our database. So, it is necessary to know also. I have already mentioned data means raw facts and a field or column in a spreadsheet kind of concept that is the group of characters with specific meaning that like if you see you know a table then individual column becomes field in database.

But in display; it is displayed as a table. So, field is important and characteristics of the field are also important of the data. What type of data it is? And in GIS term; what type of attribute, it is having? And if it is numeric data then what is the precision of the data and so on? Because once the data has come there and you want to change anything in that field characteristics, you may have problem later on.

So, it is always in a better design to think that what kind of data will be coming in that particular field and accordingly, the characteristics would be decided. And record are the row numbers basically. So logically connected fields that describe a person, place or thing. So, that is also a record. And overall, this information is kept so a collection of related records; that is in file.

Now, these terms are still being used though the GIS file system database is not used. Now as you know the simple files, lot of us uses the computers and laptops and other things and many times, we keep the data in simple file systems. And you know because they are not related to each other and therefore, sometimes searching even or making access to other or multiple files together becomes very-2 difficult. So, some limited search can be made.

(Refer Slide Time: 17:18)

Simple File System

So, if the same scenario if I put about an organisation which is involved in sales and business then the sales department is having its own file system. And then you know, the personnel department is having its own file system. And every information; this is all lying in isolation. And of course, here the customer files are also there. So, sales department and customer; everything is one but the personnel department is completely detached.
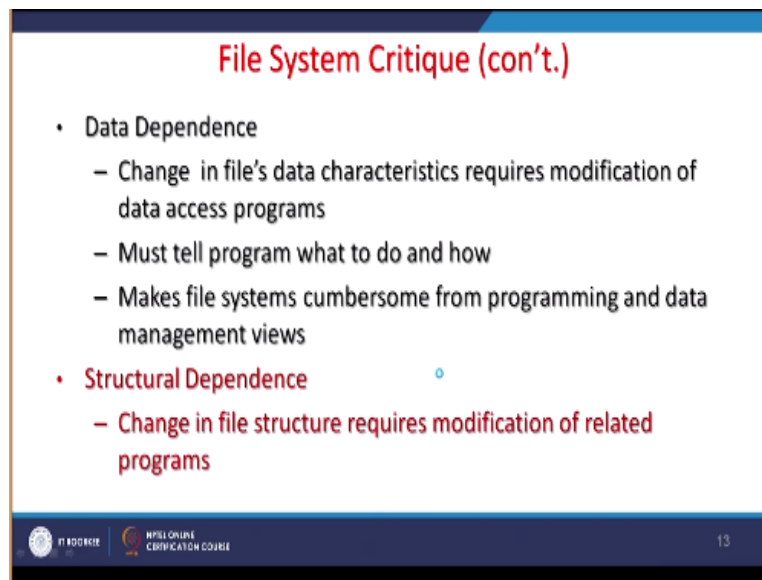
It means it is not having any connection. So, these two systems are lying completely in isolation but in a typical database management system everything should become integrated. Now, what are the bad points of a file system as you must have realised that in the file system, requires extensive programming and 3rd generation language to access the data and even make a search or make a query.

(Refer Slide Time: 18:22)

File System Critique

- File System Data Management
  - Requires extensive programming in third-generation language (3GL)
  - Time consuming
  - Makes ad hoc queries impossible
  - Leads to islands of information

Of course, it is time consuming and also makes these ad hoc queries impossible that if we are searching certain things, we may not get that information though that information might be residing in some of the files which are lying in complete isolation. And if that is there then these islands of informations is becomes very common. And that means unnecessary occupancy in our storage because of file system.

(Refer Slide Time: 18:50)

Now, this file system is also data dependent. And if any change occurs in the characteristics of data that means also, modifications are required in data access program which are in the 3rd generation. And also, everything must tell to the program that what to do and how to do? And file systems may cumbersome from programming and data management.

So, file systems basically each file is lying in isolation; though maybe in different folders. But still, it is not well organised database management system. Also, it is having a structural dependence and modifications cannot be done. If changes in file system structure requires, modifications of related programs will immediately be required.

(Refer Slide Time: 19:44)

Now, field definitions that means the column definition or characteristics, is also one can face lot of problems. Now flexible record definition anticipates reporting requirements. Selection of proper field names important in file system. Attention to length of field names is again important because later on, you may change but it may bring some errors also. So, before even bringing data into a particular field of our database, we must be careful about these characteristics.

The length, the width and what kind of data is going to come. And use of unique record identifier; that is very much required here. For each record, there has to be a unit identifier. Now when we go for file systems, the data redundancy is a very common thing. Sometimes different and conflicting versions of same data is there. And sometimes if we do not have proper recording of metadata then one may get confused that which is the latest and which is the oldest and so on.

(Refer Slide Time: 21:02)

And results of uncontrolled data redundancy; data anomalies might be there. Modifications are difficult, insertion problems, deletion might be also problem and data inconsistency might be there also with the file system which brings the lack of data integrity.

(Refer Slide Time: 21:17)

When we go for database management system; it consists of logically related data stored in a single repository. And that is the advantage, not in different files. And they are related with logics and organise in a fashion that our access to the data or retrieval of information becomes much easier. If you recall the definition of GIS; you know, this data is coming from variety of sources and it allows the easy retrieval of the data.

So, if it is organised in a proper database or well-designed database management system then only efficient storage and retrieval is possible, otherwise not. So obviously, database systems

are advantages over the file system management approach. This will eliminate the inconsistency in the data. It will also eliminate the data anomaly, data dependency, structure dependency problems also.

And it stores data structures, relationships and access paths. This access path is very-2 important from the security point of the data which we will discuss a little later.

(Refer Slide Time: 22:36)
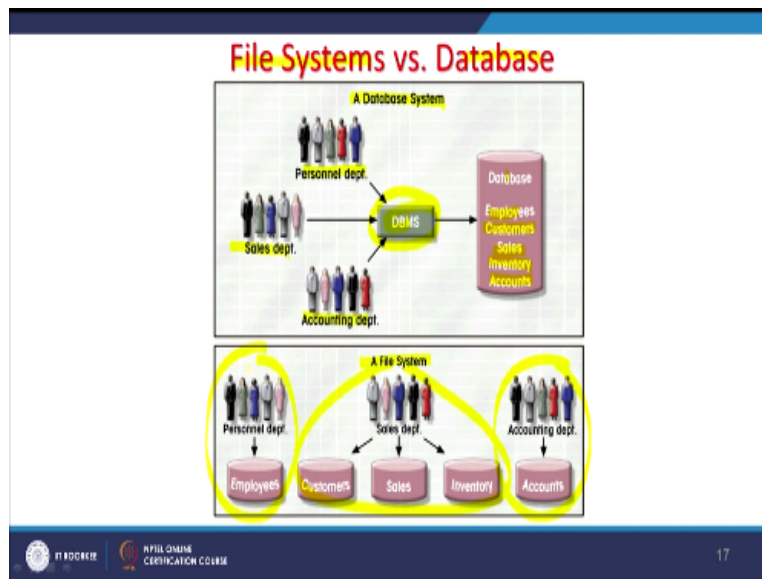
So, in nutshell, if we discussed file system versus database, this is what. In a database system, we are having a centralised database management system here and every inventory or related data is there in the database. Different departments may work independently but they are all connected with one database management system. Where as in case of file system, all people are working for the same organisation but their informations are independent to each other and every information is lying in isolation.

And they are not logically either related or connected. And therefore, getting information from one department to another may be very-2 difficult. But here, all personnel sales and accounting department can have because their data is residing in the same database management system and therefore, access and controls becomes much easier.

(Refer Slide Time: 23:34)
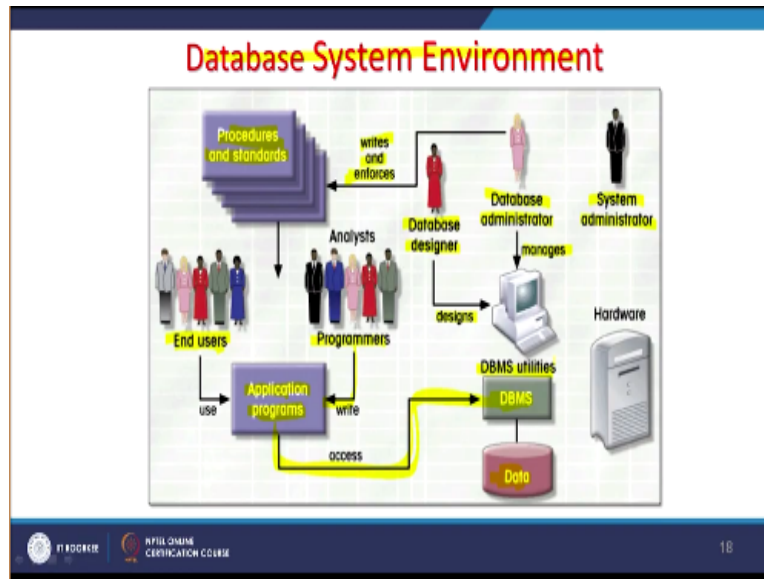
Now, if we go further to that same database system environment then we also have the procedures and standards. Nowadays, you know this is very important to talk about standards also of the database management system. Because tomorrow, somebody might be changing to some other database management system or allowing access to your data. So, if it is following certain standards, all these paths will be very easy to have access.

Otherwise, you may say no! our format is different, your format is different. I am sorry! you cannot access our database. So, if standards are followed then these problems will not come. And then there is an application programs which allows the end users to access the data. Programmers; they write so this is one way information. The end users through application programs can access the data; DBMS.
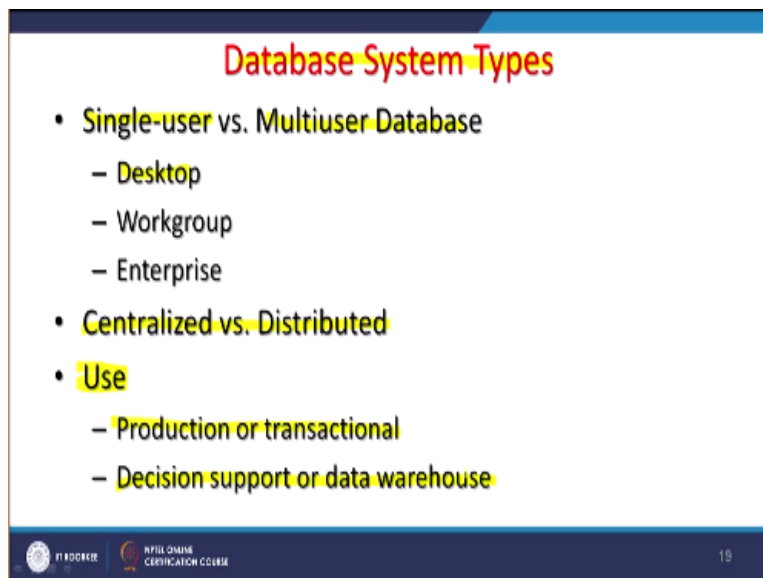
So, if I give the example like IRCTC or Railway, we can put ourselves as end users through that interface program on net; when we log in, we go into this program like this. And then we start accessing the database management which is very updated, almost every second. So, we know in which train now, how many seats in particular class of that date are vacant.

There is of course overall system administrator to control everything. There will be a database administrator which will come focusing mainly or managing the database. There will be a database designer because you know very frequently, they have to change something in the design or a treat within these designs so that they can accommodate new train, new class, new route; all kinds of thing.

So, they are designers are also involved. And of course, then these procedures and standards; they are connected with the rights and these standards are enforces through data administrator. So likewise, you are there. And one important thing is about of course, the data is there but important point about the backup because such important system has to have a very good back things in their system.

And your backup may not be at the same location. Backup maybe at a remote location in case of some accident or some fire in the data centre then the entire Indian railway system will collapse. So, there has to be frequent backup or live backup, continuous backup and maybe at different location. So, this is how it is followed in many large database systems.

(Refer Slide Time: 26:33)

Now when we design the database, the concept; either it is going to be used by a single user or one user at a time or maybe multiple user database or millions of users at a time like your IRCTC. Again, people going to use on desktop, workgroups, enterprise or even through mobiles. So, that is another latest thing that every such databases are being allowed to use through smart mobiles.

So that way, the database designers should think in the beginning nowadays. Because that is the most common devices which is being used. And again, they have to decide about the centralised or distributed systems. People talk about cloud and other systems, fantastic. Can we have a centralised system or should we go for a distributed system? So there, it has to be also waited. And then who are going to use? What is the purpose or other things?

Because the database type/system type will depend on all these parameters and of course production or transactional like banking; State Bank of India site that database, it's a transactional. Or some factory might be having so, it is production related. And decision support or data warehouses. Now, whether my data is related or database management system is being designed for making good timely decision or it's a data warehouse.

So, the type will depend on all these things. Now, what basically different functions within the DBMS are?

(Refer Slide Time: 28:26)
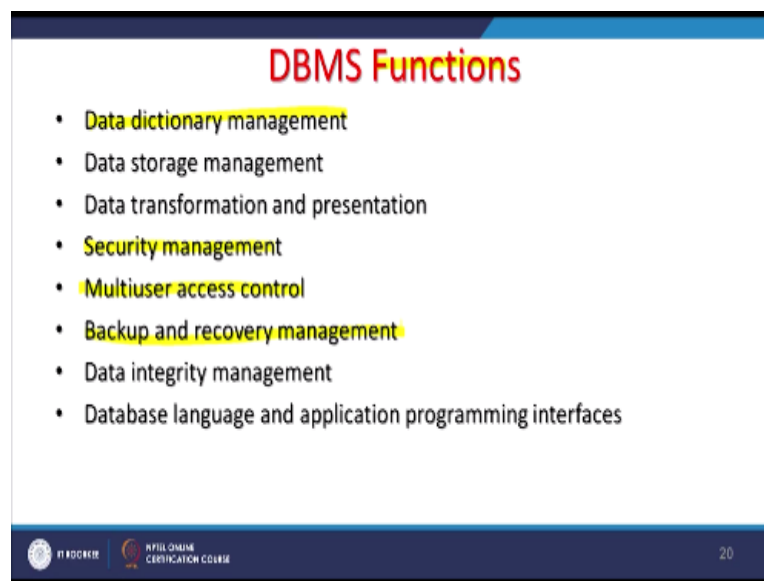
It maintains the dictionary management, data storage management, data transformation and presentation, security management. Nowadays it has become very-2 important because people are hacking these large systems differently and sometimes, they can create havoc. So, the security management is also very-2 important. Multi user access control; how users will have access, that can be controlled.

And of course, this I have already mentioned that backup and recovery management systems are very much required. Data integrity management and database language and application programming interface. So, you might see that in every few years time or almost every year, we are seeing some changes in IRCTC site. And then database communication interfaces; how it will and which kind of platforms, it will be used?

Which are going to be the interfaces or browsers, how do we be communicated that has to be also thought, designed and then finally checked.

Now when we talk about these database models, it's a basically collection of logical constructs used to represent data structure and relationship within the database. Now these conceptual models in which logical nature of data represented and whereas implementation models which emphasises on how the data are represented in database?

Now, the most common one in GIS is related with, is the relationships or relational database in conceptual model so we start thinking from that. Now this relationship either one to one (1:1) relation or can we one to many(1:M) relation or can we many to many (M: N) relations. So, 3 scenarios are possible; one to one, one to many and many to many. Now in case of implementation database models, we can have hierarchical model, quite common in many organisations.

Network based, like if I have to do lot of analysis which requires network analysis then network type of database model is most appropriate. And finally, of course the relational database. I want to relate one database with another and things have to be organised accordingly. So, implementational database models are hierarchical, network and relational. Now in the second part of this discussion, we will go in details about these concepts with some examples also. So, with this, I end this discussion. Thank you very much