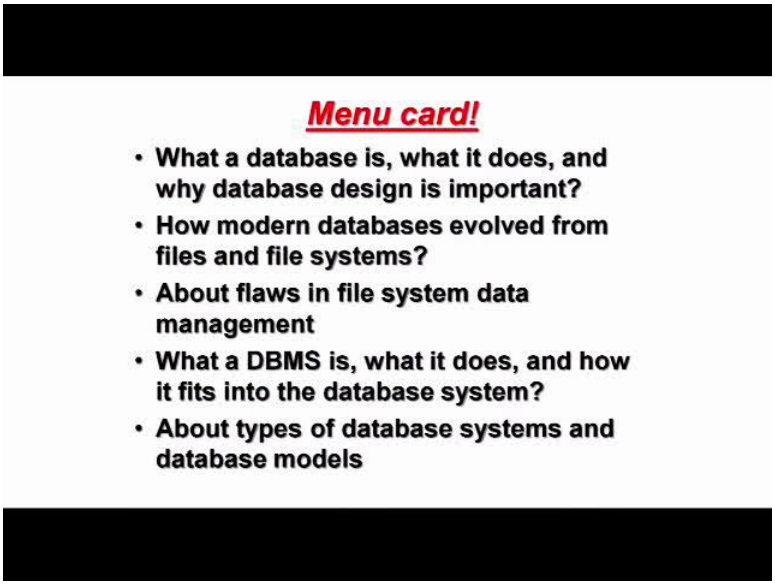
Introduction to Geographic Information Systems
Prof. Dr. Arun K Saraf
Department of Earth Sciences
Indian Institute of Technology, Roorkee

Lecture – 08
Spatial database system and their types

Hello everyone, in this lecture we will be talking about a spatial database systems and their types as you know that everything in GIS has to go inside a database. And the database which in GIS we use we call as a spatial database. The concept of a all this almost same as the other database which exist, and the most common database which is used in GIS is r d, RDBMS that is relational database management system , but there are other types of database is are also used in GIS. So, we will go through one by one, and a also as you know that a file systems we are being used and now we are going slowly-slowly or moving towards the database.

(Refer Slide Time: 01:23)

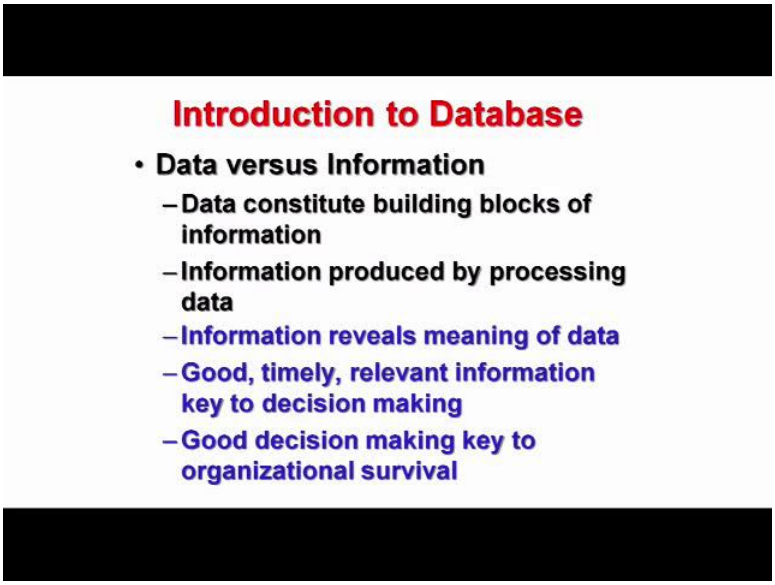


So, what we are going to discuss in this lecture is what is database? How what it does and a how it works and how to design a spatial data base? How it is important why, we should a take all care before we put we put start putting the data into the GIS data base.

And how modern database is evolved from files and file systems because, earlier systems when computer started coming we are keeping things in file systems still many people keeps in file system, but not in GIS and about flaws what are the disadvantages demerits

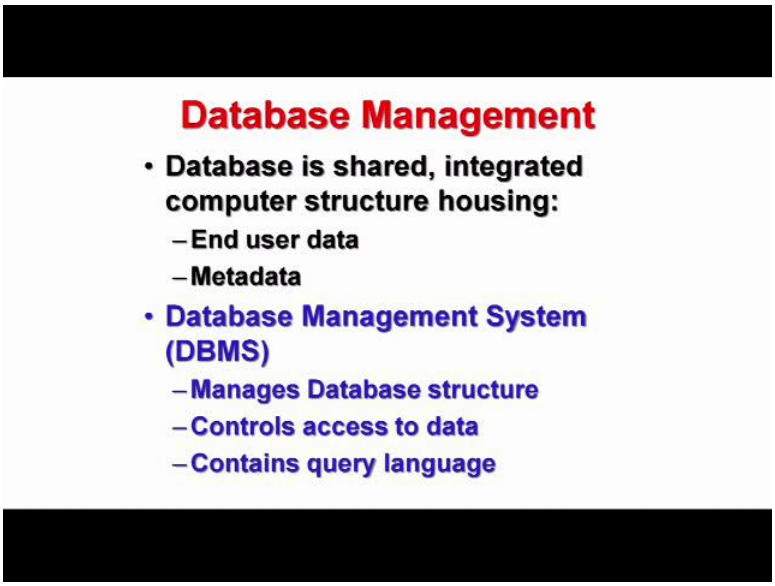
of file system, what are the advantages of DBMS data base management system and also about different types of databases, which have been implemented into GIS. So, all this we will discuss in this particular lecture as you know that our GIS, what GIS does is it converts your data into information and then ultimately information to knowledge. But before that we have to keep their data inside the system. So, a data as we know is the one of the major components of GIS out of 5 components one of the important components of GIS is data. So, we have to take care about that very carefully we have to organize our data into the GIS software and that because if data we organize correctly we should be able to retrieve it efficiently and during analysis and modelling this all will help if we have organized correctly.

(Refer Slide Time: 03:10)



And as you know that information reviles meaning of data and a good timely relevant information is key to decision makings and good decision making is key to organization survival. So, if it everything depends how data has been organized, in a organization in a set up or in a company.

(Refer Slide Time: 03:36)



Now, database as you know there might be a possibility of sharing as database, nowadays this is one of the very common thing, but to what extend the sharing is; that means, the security aspects access aspects we will also touch up on that, then a how the integration is done of the database and a who are the users and a 1 more important term which is mentioned here is metadata. Metadata in GIS or in other systems are very, very important because metadata means the data about the data or information about the data. So, for example, if I have gone in the field and collected ground water level data from different observation wells. Now i-i need to record the time when the data was collected, and who has collected the data and all kinds what kind of instrumentations i have used to measure the ground water level. All kinds of relevant information about the data has to be stored as metadata and the many times whenever you download a satellite image from internet free satellite images.

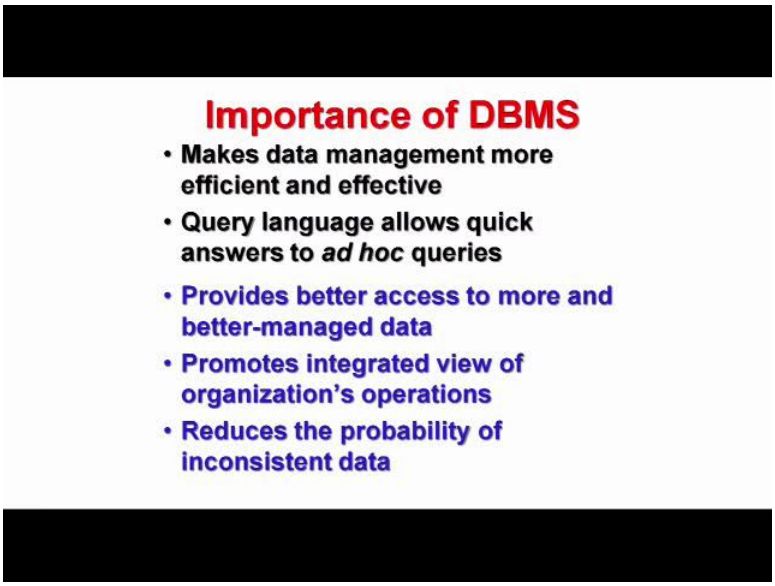
(Refer Slide Time: 04:47)

You will also see that a apart from tif file you also find a file which is MET. This is not metrological data this is this is the normal convention is used for metadata. So, this contains all the information about from which satellite which sensor which date what was the coordinates of sum and so and so forth. All that information will go in the MET data.

So, it is always the good practise whoever is developing data or analysing data into the GIS, modifying the data that all the information should be recorded in the metadata file. So, metadata is very, very important in GIS too. And data base management systems the purpose here to manage the data structure, database structure and also control the access to the data. Whether all the people can have all the access to the data or some people or there may be higher article system so on and so forth. And then a database should also support a query language because by putting certain questions to the system you should be able to retrieve the data. So, this query language is very much required, and this is the typical database management system which we are seeing here.

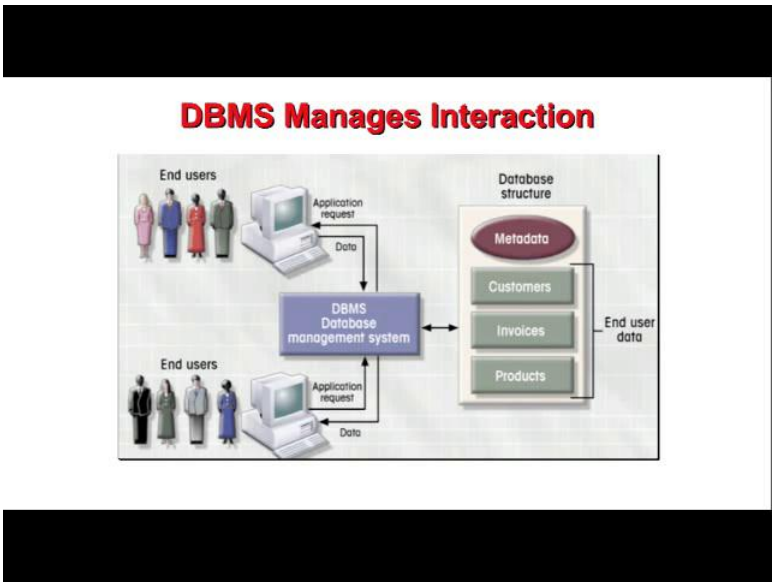
Now, make data managements are more efficient and effective if you organize the data. When you create the perfect structure and then put the data then you would be only able to efficiently retrieve the data. Query language allows such to allow query whenever there might be some queries which are systematic every time you are retrieving sometimes it is just certain information you want to retrieve from that database.

(Refer Slide Time: 06:25)



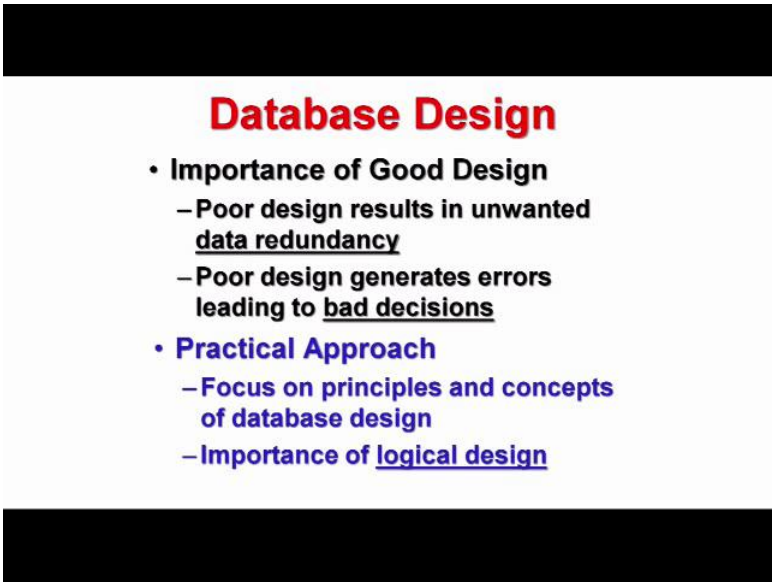
So, allow queries are also be allowed by your system. And then provide better access to more and better manage data promotes integrated view of organizations operations and reduce the probability of inconsistent data here also the purpose here is that we should not have redundancy in our data. And there should not be inconsistency data inconsistency means basically here is that a, if you are keeping the same data at two places within one database management system. You have a upgraded one you have not upgraded another one or automatically it has not been upgraded then there will be inconsistency and then it will lead to the wrong analysis and poor decisions. So, these things should be avoided in a in a normal practise.

(Refer Slide Time: 07:40)



How you know database manage the interactions basically you are having a database management since system in the centre you are having end users and you are having you know this kind of a structure database structure that you are having metadata, and if it is related with sales and business then, you would be having costumers invoice this products and again and user data and then there is a interaction of different people with the data base management system. So, this is what how the interactions within the data base happens, now what are the good practices which should be followed while designing a database, that you know the because, if you design poorly.

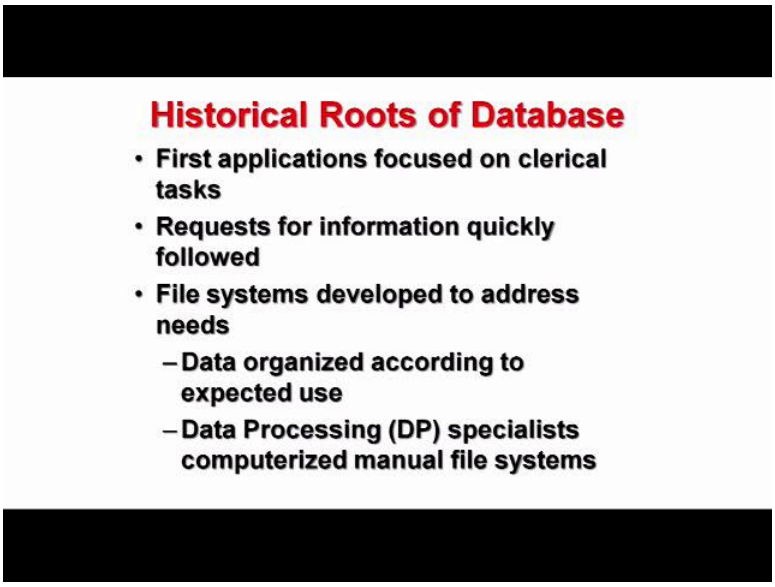
(Refer Slide Time: 08:12)



Then there will be data redundancy which i have just touched upon and we have also discussed this data redundancy specially in especially in case of raster data and therefore, we look for the data compression, but here the data redundancy will be increase of say textural data or some other numeric values, data redundancy may be that two places the same data is being stored and updated for no advantages, and poor design generates errors which may lead to the bad decisions.

And the practical approaches are that we should follow the principles and concepts of database design. And also follow this logical design system, now there are as you know that this is not very new even before the invention of GIS the concepts of database management systems were there. So, first you know we thought that instead of using analogue files we started using digital files.

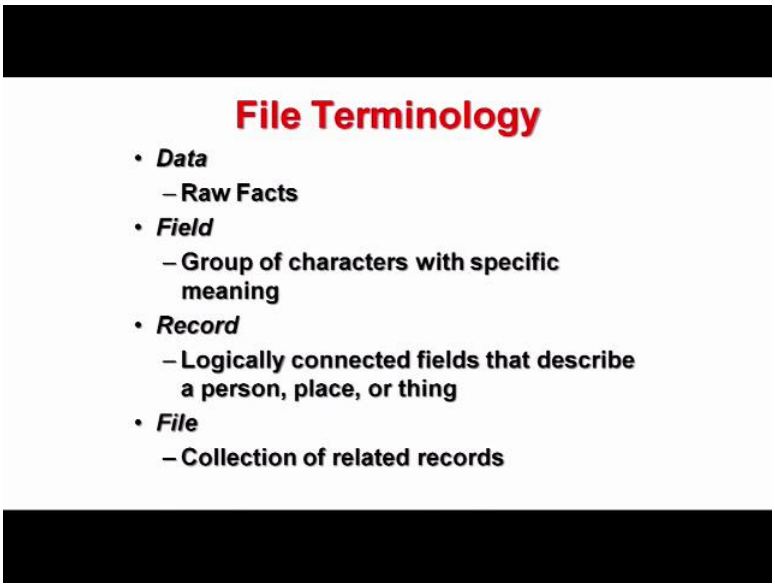
(Refer Slide Time: 09:29)



So, initially the concept was concentrated mainly on the clerical task and therefore, the file systems were developed initially later on the database management systems were developed. And request for information which we have quickly followed file systems to address needs data organize according to expected use and in file system data processing specialists computerized manual file system. So, this though there were these were the digital files, initially they were analogue files converted into digital files, but since there were no database management system and therefore, the query adobe queries or regular queries retrieving information was still not efficient. And a as a you will also realize that

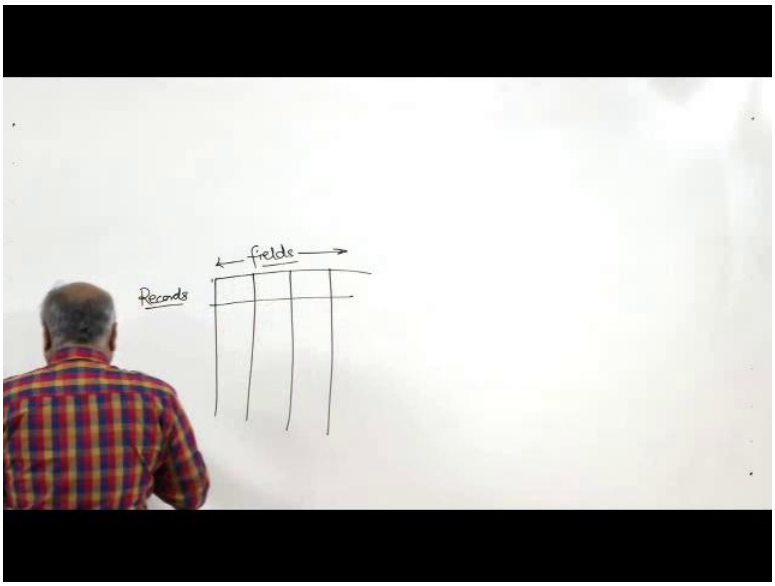
while using computers sometime we keep information in simple separate files and therefore, it becomes very difficult to retrieve appropriate information and all kinds of problem will come.

(Refer Slide Time: 10:25)



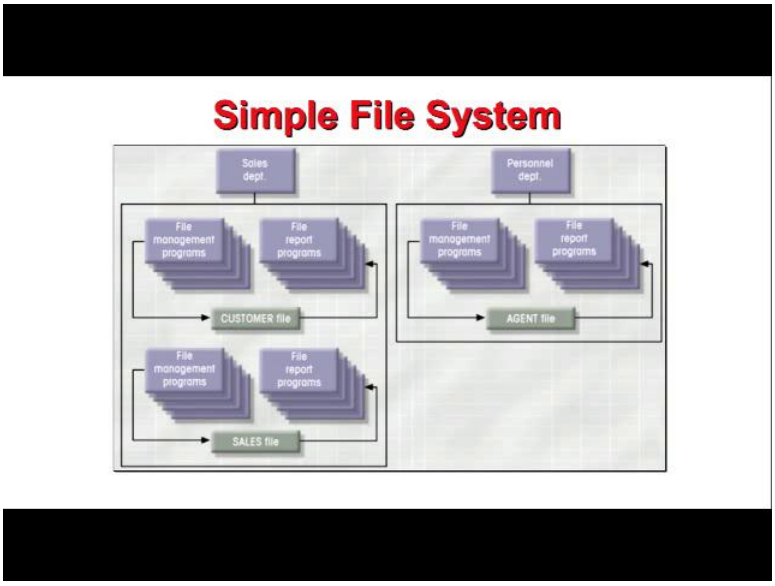
Now, whenever we use the database there are certain terms which we use in database. Like data which are nothing, but the raw facts, and a then we use a in GIS database as a i was discussing in attributes data i said that we can think as a tabular data.

(Refer Slide Time: 10:43)



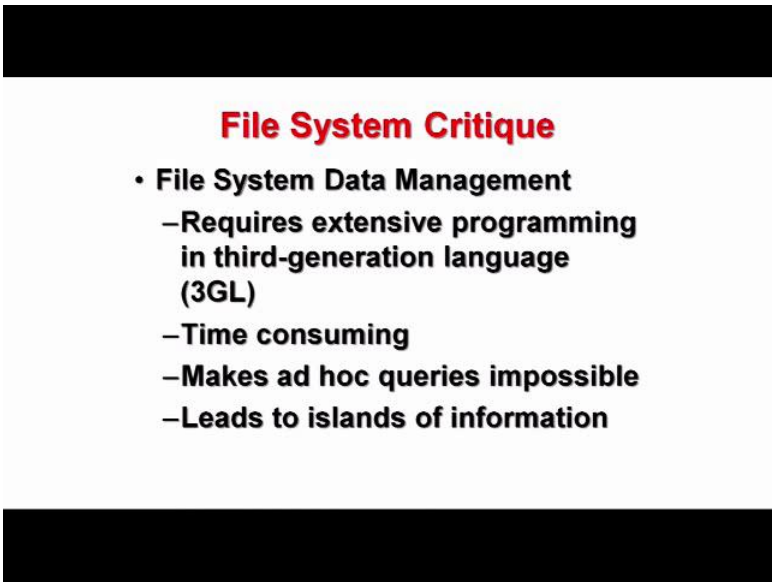
So, when we are having a tabular data, then we this these are the things which we call as fields or columns and these are which we call as rows we call as records in database languages. So, now, this a fields or group of correctors with specific meaning, these fields or columns in a tabular database and these we defined the properties we define according to our data which is going to come inside these columns. So, therefore, these are the very important thing the properties especially to define and then record as i have already mentioned that these rows or records, which are logically connected fields that describe a person place or a thing and then overall everything is stored in a file. So, we use this term file as well which is collection of related records. Now if we look a same sales scenario and it is file system then this is how things are kept completely separately and hardly there are linkages.

(Refer Slide Time: 12:00)



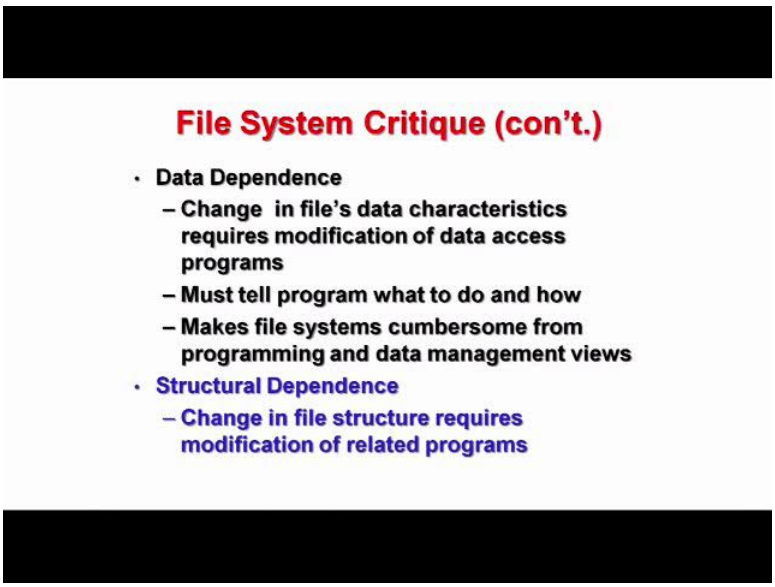
So, there is a customer then there that department then this sales department is having customer and sales files and then personal department is having agent; that means, the information is there within an organization, but information is lying in isolation. And it is sometimes when we want to retrieve the information from one section of a organization to another one department from another then this integration is not possible, retrieval of information from one department to another is not possible if it has not been organized in a typical database. If it is organized in a file system this is what it will happen.

(Refer Slide Time: 12:46)



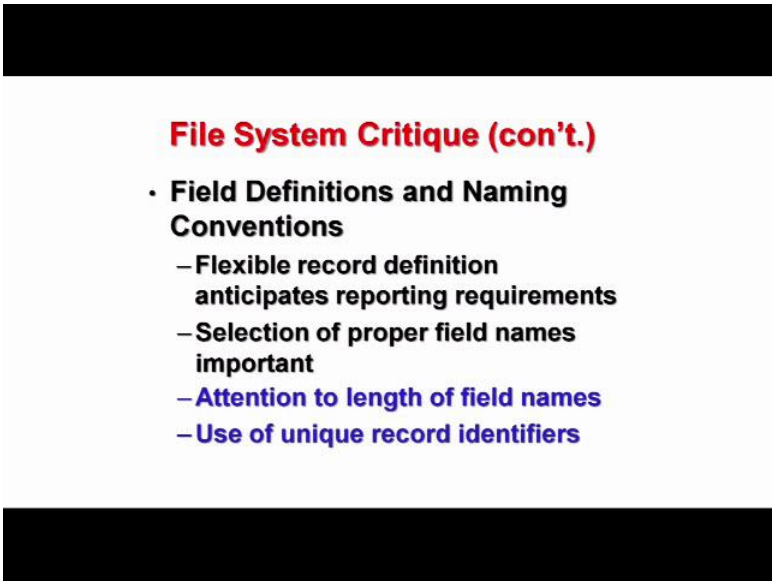
So, the what are the bad things about the file systems and that requires extensive programming in third generation language and a in order to access the information or data it is also time consuming and makes ad hoc queries impossible. You cannot just raise certain questions to the system or file system and you will get the answer or you will get the data no it is not possible and then i leads to island of information it is nothing, but the redundancy, that information is lying in your file system, but i am unable to retrieve it and this is what we call island of information.

(Refer Slide Time: 13:32)



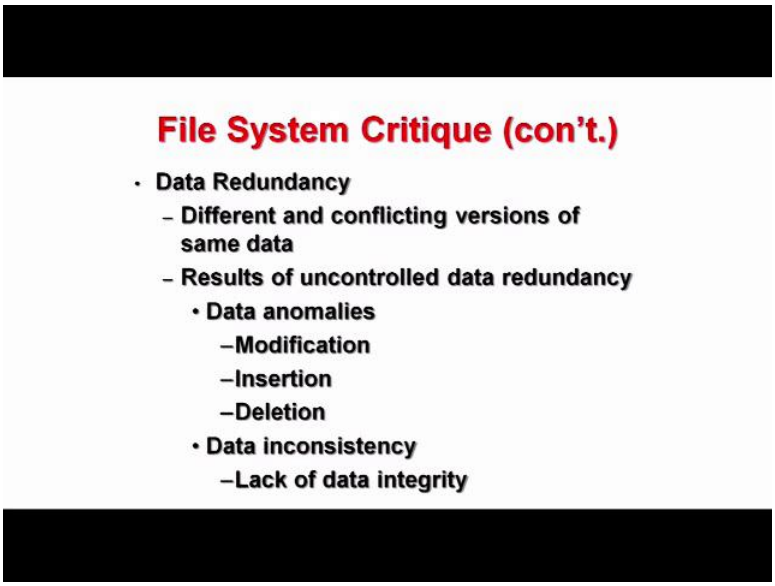
So, the these are the problems in a file system this continues that a this a file system is also dependent on the data which if i change the file name then everything i have to take care, because that file name might be being used elsewhere as well. So, accordingly then modifications in the program must take place and make file system cumbersome from programming and data management views. This is structural dependence is also there in the file system the change in file system requires modification of related programs. So, every time you change something in the file or naming in the file or you add some new file then again in the programs you have to change. So, this is not a good practise, but these are the earlier things and what is the best thing, which has happened in case of data management system which we will see.

(Refer Slide Time: 14:20)



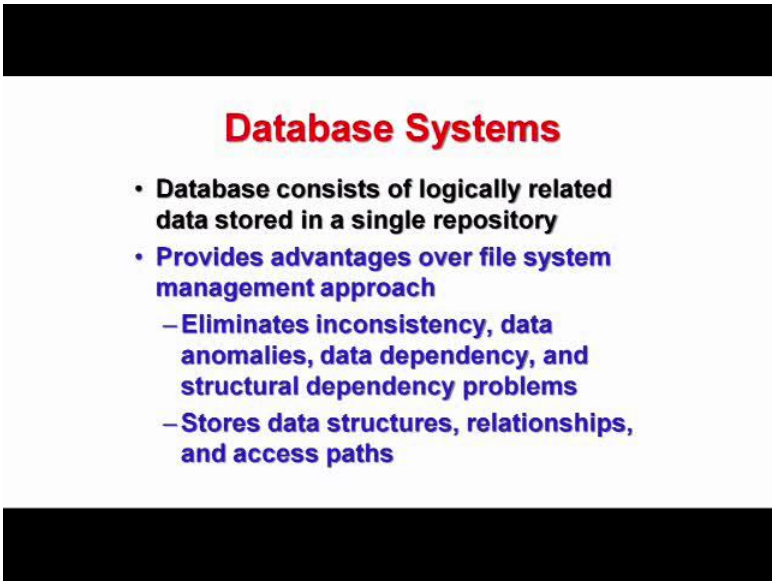
So, this is field definitions and naming conventions that is also limited and flexible it is not flexible and a selection of proper field name are very important in file system. And attention to length of fields and use of unique recorded in; that means, there are lot of things lot of things have to be taking take care while if we are using a file system .

(Refer Slide Time: 14:49)



Now, there is a data redundancy i have already discussed, that data redundancy islands of information is the big problem in file system. Results of uncontrolled data redundancy and whenever we modify the data, insert the data, deletion of the data, may bring the anomalies in the data and data inconsistency is also there because the data has not been properly integrated.

(Refer Slide Time: 15:16)

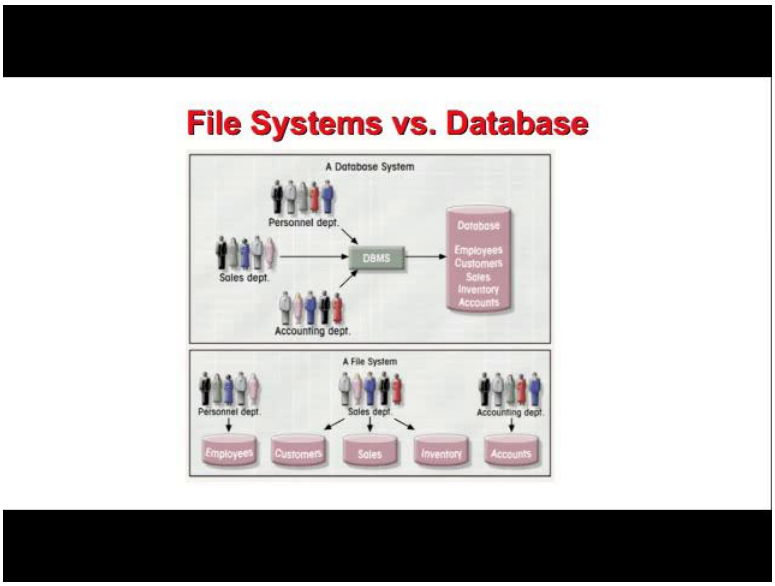


So, lack of data integrity whereas, in database management system which is based on the logical related data stored in a single repository. So, in a one system or one repository it

has been stored may not be exactly one system. But may be nowadays we go for clouds and other systems so, but a everything is controlled in a manner the data can be retrieved very efficiently. One of the best examples of nowadays online database management system which we see access is a like database of IRCTC for Indian railway reservations. Which is really in my opinion it is very complex database, but it is quite user friendly and person who know who knows about little bit about the computers can use the database and can book the trains. So, that kind of products when are created then they become very popular and highly useful as IRCTC programme.

The same the because database management provides advantages over file systems, what are those advantages? That eliminates inconsistency of data it also eliminates data anomalies, islands of data, it data dependency is also eliminated and structural dependency problem is also gone, if we go for database management system. And its two data structures relationship and access paths that is another advantage with database management.

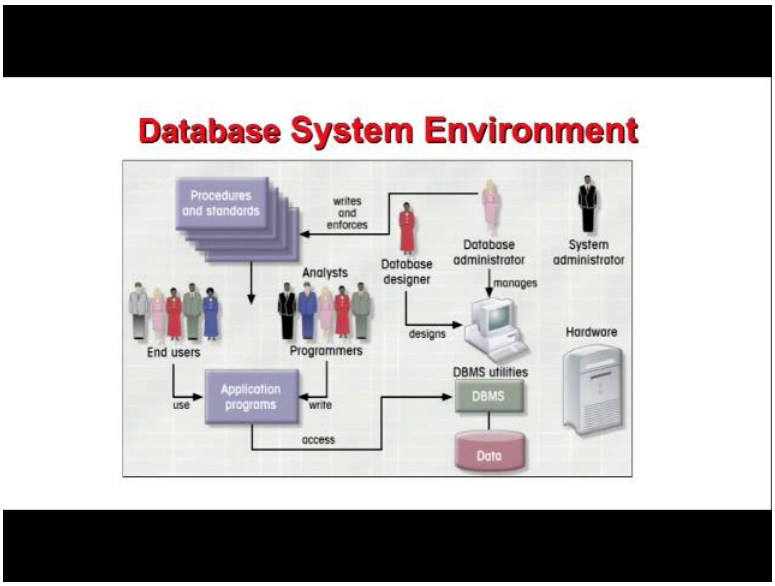
(Refer Slide Time: 16:56)



So, if we take the same example of a organization, which is in business for sales and other things then here with the database management system every other section or department is connected. And then every information is kept in a single repository and therefore, then compared to this database system you see in the file systems they are the information is lying in isolation. And therefore, the information retrieval data retrieval

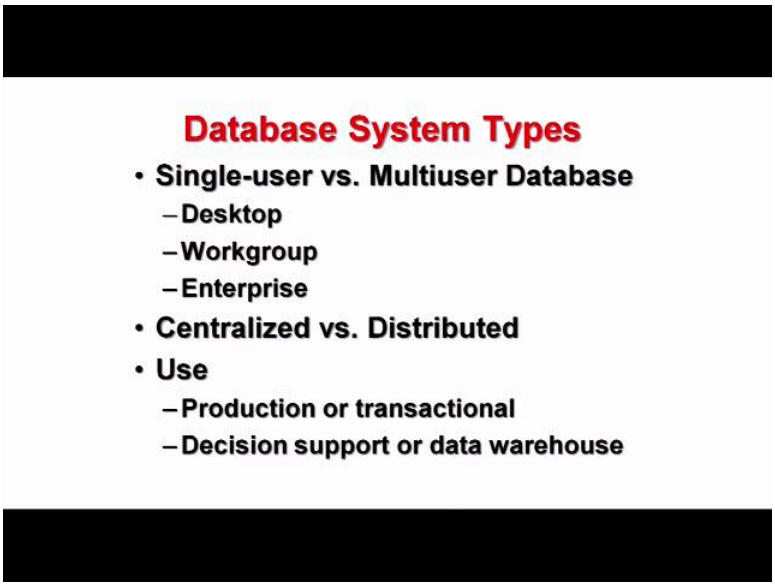
becomes difficult. Now in generally in database management system what are the hardware or software different component that we are having hardware you are having your data hard disc and computers and users will have their own computers and you follow certain procedures and standards.

(Refer Slide Time: 17:32)



Which all large enterprises organizations go for such a procedures and standards and then you are having application programmes which will allow you to retrieve the data interact with the data base and update the database by different people. Then there will be control or the access through a system administrator; that means, end user may not have access to the full data, but analyst might be having access to certain type of data and so on and so forth.

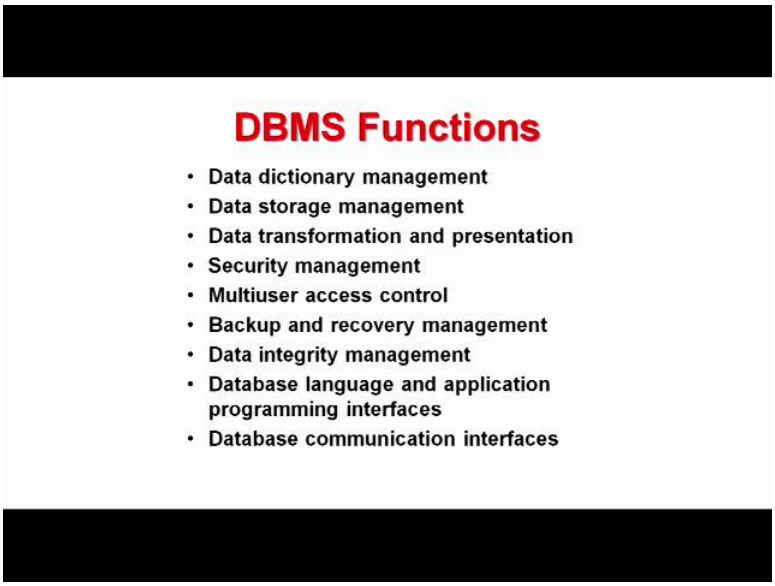
(Refer Slide Time: 18:17)



Database management systems there are different types of concepts exist with the database management system that is the single user verses multi user data base multi user database systems are now most popular example i have given already of IRCTC or some banking organizations database like a SBI online, that is also a multiuser database very user friendly and it is having various options various you know capabilities are available for us for transaction or we now using ha that a database for various purposes. Now the system if we go on hardware basis then, you might have a database on desktop maybe in a workstation, workgroup, enterprise, online centralized, distributed all kinds of options nowadays are available with database, and it is not necessary even there is, entire reposiditary is being kept in one country, there might be multiple countries it is connected and you are accessing the data.

So, generally people are going for multiuser database and distributed database because there are certain advantages with distributed database, and production and or transactional use like transactional i have given the example of banking organisation database and also database arrows and other things are there.

(Refer Slide Time: 19:52)

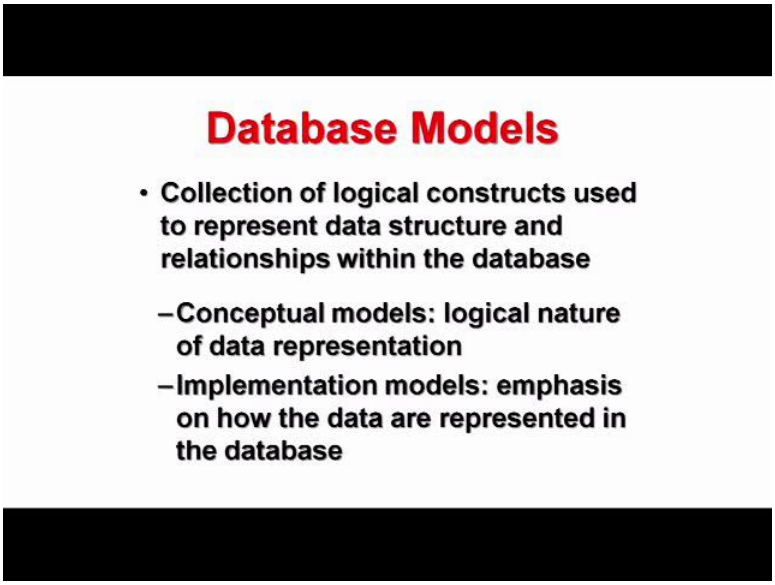


Now, what are the functions of database management systems, that data dictionary management data storage management data transportation presentation, security aspects nowadays has become very big issue. So, that has to be taken care multi user access control who would use what who would access what how to what depth they can access and so on, another very important thing to all kind of data base is the backup and recovery management, because anything can happen anytime there might be some accident, there might be some breakage or problem with the system. And may be fire or some flooding or other scenario so, big organization nowadays they keep you know this backup system very nicely may be two parallel systems one may be in say in example in Delhi another one might for in Chennai. So, that is something happens in Delhi at least the Chennai system will work and this is how this big database which we are using as a simple end user these are how this is how these are being maintained.

And the backup has to be taken most of the time it is simultaneously and there should be a very good recovery management, something goes wrong. So, that we should be able to recover the data because the millions of people information is especially in banking system database is there which is very, very important related with finance data integrity management data language and application programming interfaces, end user may not be aware of this thing, but on the administrator and these are the thing which has to be taken care. Then database communication interfaces clear on whether you are accessing the data on your mobile or on your computer or even in the field or through internet or a

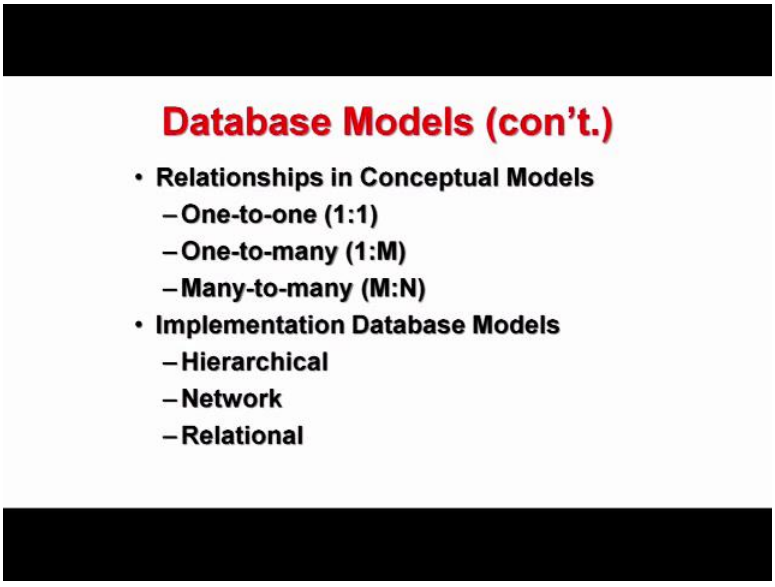
staying abroad all kinds of these accesses communication interfaces have to be developed.

(Refer Slide Time: 21:55)



Now, as you know that in database management system it is a collection of logical construct, which is used to represent data structure and relationship within the data database. And these conceptual models may be logical nature of data representation or may be implementation models, which emphasis on how the data are represented in data base. So, various ways of you know going in this database.

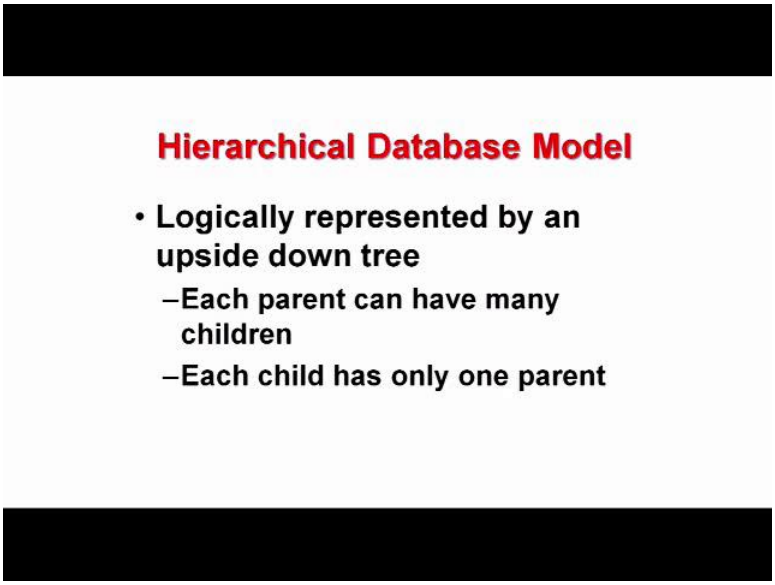
(Refer Slide Time: 22:34)



Now, from GIS specially, from GIS prospective particularly about how in GIS we are linked and related. Then this is very important that there might be a based on relationship the conceptual models which we can think, of about DBMS is one to one, one too many and many to many. So, these kind of as per our requirements we can have different options available in terms of relationships especially the relational data base which is just coming. So, this are the implement implementation database models which have been implemented in GIS and a extensively specially the relational one, and the network one. Which are having lot of advantages and therefore, these are the most 2 very common database management system which has become sort of special database management system and have been implemented in most of the GIS softwares.

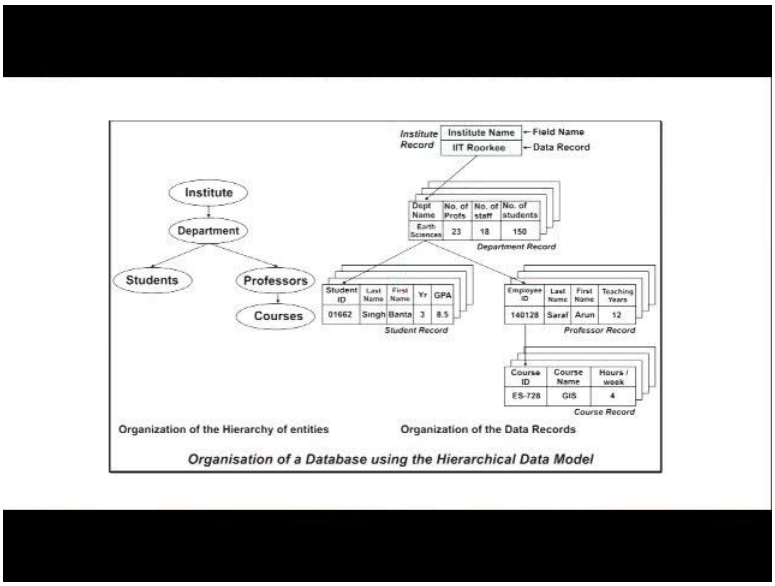
Another thing is a with the GIS nowadays, that all our modern GIS softwares or this concept also support access to the external database management. So, you not only the data base which is within your organization which you are having access, but you can also have access to other database, if the access has been provided by the administrator. So, those are the things which have to be.

(Refer Slide Time: 23:55)



So, we will go one by one on this that first is hierarchical database which is logically represented by an upside down tree as name implies there is a hierarchical system each parent can have many children. So, it is one too many and each child has one only one parent.

(Refer Slide Time: 24:19)



So, if we go upward then only one parent and a i have taken example of a typical department and some names i have put there, but it is having just a these are name no other meaning here, and what we see when we put this ah in hierarchical database data model then this is how, the flowchart would be that at the top we are having institute then each institute might be having several departments and then these departments are having students professors and courses, but here as you can see there is a hierarchical system one type of system which exist, but there is no linkage between students and courses and therefore, if a student of one department would like to attend the course of another department, then there is no linkage is available.

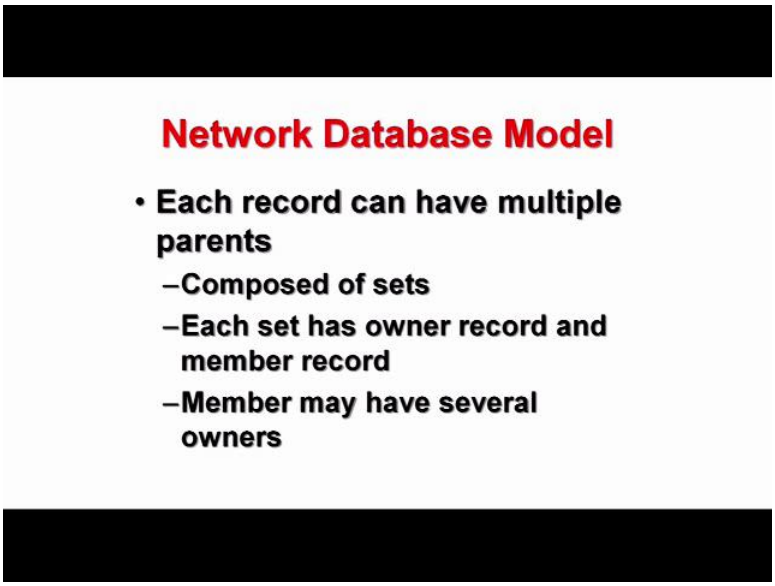
If we see in a you know in a proper structure form then the field name here is institute name then it is comes a Roorkee, IIT Roorkee then, department taken example i have come from earth sciences. So, i have taken example of earth science department and you know number of student's professors and staff and other things. Then you are having a student records you are having professor records you are having courses records.

(Refer Slide Time: 25:48)

But as i have said if a student of earth science would like to take a course of civil engineering or say mathematics then there are no information can be inbuilt in this type of hierarchical database. So, what are the advantages with hierarchical data base that it is conceptually it is very simple, database security and integrity is there and data independence. So, one data in one file is completely independent of another file, and a efficiency is also there because hardly there are any linkages or anything disadvantages that we would like to implement then the implementation point is complex. Difficult to manage and lack of standards because once new data new data set will come or something new as to be required then you will have problems. Lack of structural dependence, application programming, and use complexity are there and then implementations limitations are there complex implementation and that too is limited.

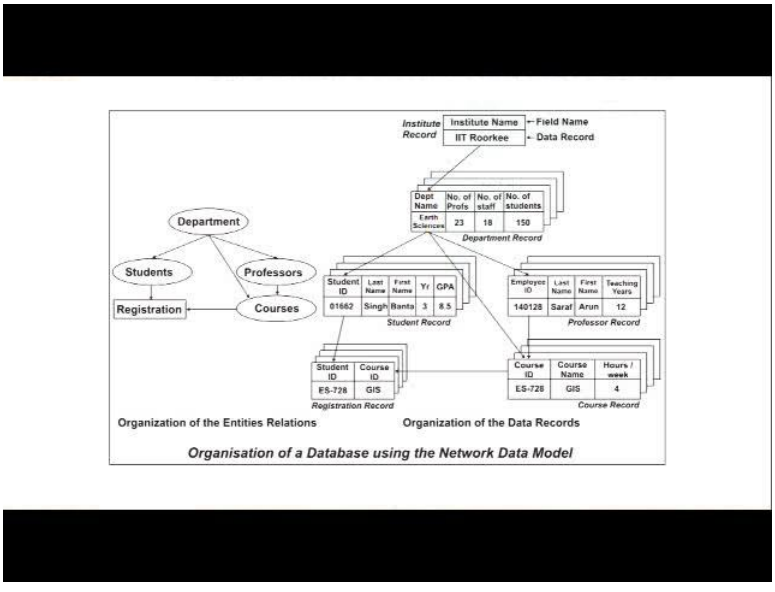
Now, the second type of database management system which is one of the very popular and a specially for the vector data, line data which is used for different kind of networks if we recall the history of GIS and the invention of GIS by Roger Tomlinson. He started initially thinking about the networks, because the task was given him for the network and therefore, the most of the GIS softwares are quite rich on this in these concept or especially about the network related because there might be management of a network through which the recourses are flowing may be natural network like stream network may be transport or may be railways or all kinds of things are there.

(Refer Slide Time: 27:26)



In network database each record can have a multiple period. So, that is the difference in previous example a parent can have a multiple children, but here each record can have multiple parents, and composed of sets each set has owner records and member record and member have several owners and the same example.

(Refer Slide Time: 27:50)



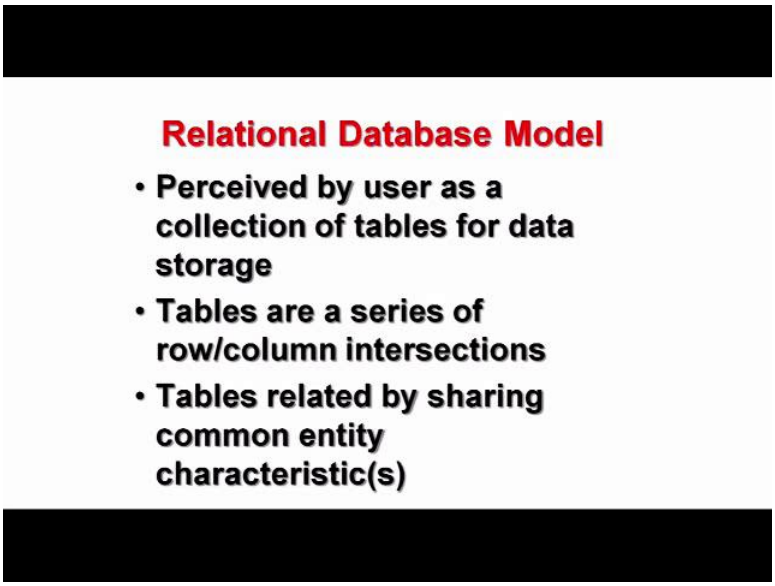
But a in a network data model this is how it looks, that departments are now having linkage with the courses and then there is another you know section or records have been created which is the registration of a students. So, when the students register for a

particular subject then for course of another department can be taken by those students and this is what the normal practises being followed in institute like ours.

(Refer Slide Time: 28:29)

So, they the advantage here that now there is a linkage and therefore, a complete network is available through which the data can flow, now let us see the advantages is conceptually it is also simple like hierarchical handles more relationship types it can handle more relationship data access flexibility. There promotes database integrity in case of file based system or hierarchical system this was lacking data independence is there and conformance of standard. At least some standards can be followed, disadvantage is that the system is a is complex and lack of structural independence. Now the third database or last one in this one is the relational database or RDBMS which is the most popular one and extensively it has been implemented in GIS.

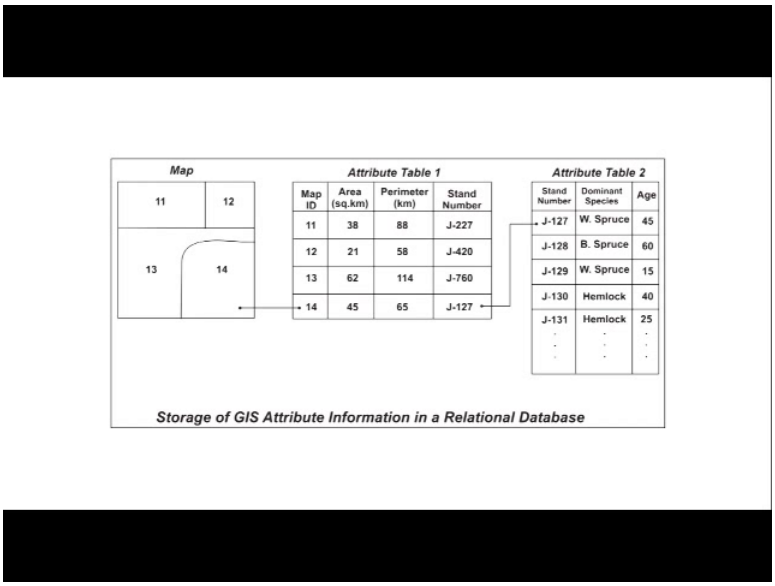
(Refer Slide Time: 29:14)



So, what it is it is perceived by the user as a collection of tables for data storage i give this example, that most of the time we keep in this tables and then tables are series of rows and column interaction as you can see and tables related by sharing common entity and characteristics, this is on this point I would like to emphasis little more, that whenever you would like to relate one table with another that is why it is called relational database. So, there is a one table there is another table, now i want to relate and i will be showing through view graph some example as well. So, for this relationship between 2 tables or 2 data bases there has to be a common entity; that means, there has to be a common field which is present in both the files or both the tables and having the identical properties.

So, if there is a common entity or common column is there having say for example, i d. So, here i might be having a i d field here, another table i might be having for the same area i might be having i d field. So, when this is common and same system of identifying different objects has been followed then i can relate this table with another table. So, this is what it is meaning is that the tables related by sharing common entity characteristic. Characteristics here means the properties of that field has to be the same.

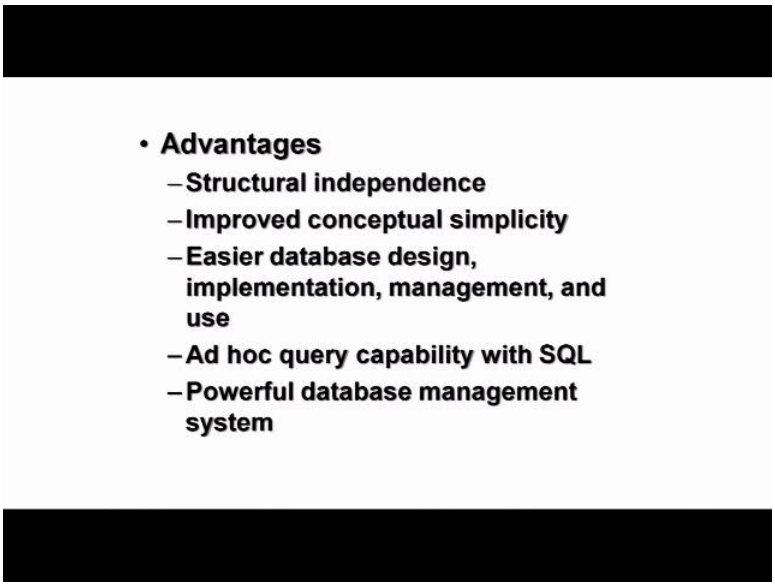
(Refer Slide Time: 30:47)



Now, let us take the example is a that i am having a map which is having four polygons here and i am having corresponding attribute table for each polygon there are lot some information like map i d is there, area is there, perimeter there, extend number is there. Now i want to relate this table of this map with another table which is which is also having same polygons, but attributes are different, here attributes are that stand number is common, but i am having a species information dominant species and their age. So, now, what we are seeing here in this two tables that at least one field is common, and once it is there then i can relate it if it is not there then relational database may not work in that way.

So, what we are seeing here in this particular example, that the stand number j 127 is also common here and elsewhere there the same records are not being reflected. Does not matter, but if i relate with this that data will whatever at the common may come in a new table or in the same table and then later on we can store the information. So, that is the advantage of relational database that anytime information may be lying in different tables you can think in this way. In an in digital table I am talking, but whenever i want to relate if one field is common i should be able to relate very successfully. So, this gives a lot of advantages in also in the GIS this kind of concept of database management give lot of advantage in GIS because we will be using in real GIS operations many maps which are having polygons or other vector features. So, there we it is very much required to have a relational database management system.

(Refer Slide Time: 32:52)



Now, what are the advantages structural independence, improved conceptual simplicity, easier database design, implementation and management use? So, these are the easier things ad hoc query capability with s q l and powerful database management system that is why it has become one of the most popular database management system one of the best examples like oracle is RDBMS earlier we used to have d base i base, all kinds of database management system, which are all more fall in this category of relational database management system.

(Refer Slide Time: 33:28)

Now, disadvantages substantial hard and system software overhead expenses are there, the design may be poor design implementation; however, implementation is easy and a may promote island of information problems. Why again because in several tables the same information might be there. So, in one table you have updated in another table you have not updated and therefore, you may encounter this kind of problem which is island of information.

Thank you very much.