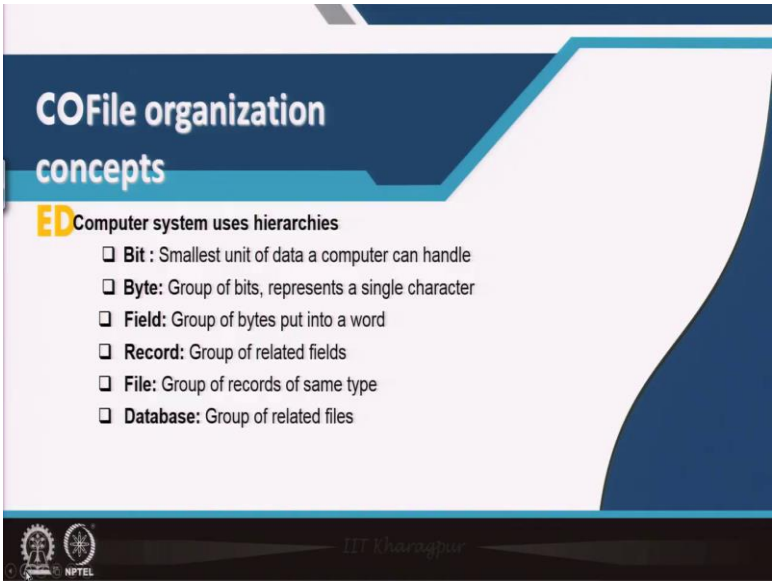
Management Information System
Prof. Saini Das
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Module - 02
Foundations of Business Analytics
Lecture – 06
Database & Information Management

Hello students! I am Professor Saini Das and I am a faculty at the Vinod Gupta School of Management, IIT Kharagpur. So, I will be taking you through the 2nd module of the course on Management Information Systems. So, here we will be talking about 'foundations of a business analytics' and in this particular session, we will be talking about 'databases and information management'.

(Refer Slide Time: 00:40)



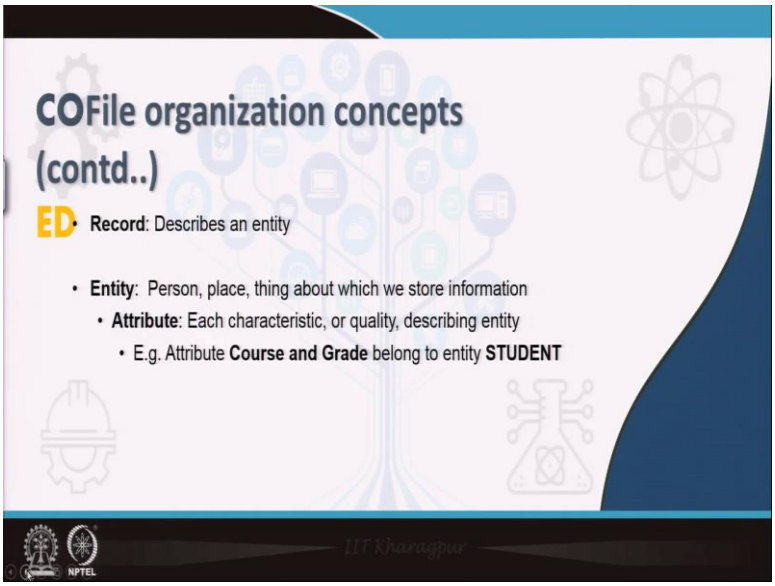
So, by now I am sure you would have understood that in order to make effective business decisions; organizations require accurate and timely data availability through information systems. So, in order so that data is available in a timely manner and in an accurate fashion data, needs to be organized in information systems. So, today we will be discussing the various ways in which data is organized in information systems.

So, we will begin with the traditional mode of data arrangement and then we will talk about the various stages of evolution that it underwent. Computer system uses

hierarchies; so the smallest unit of data; a computer can handle is called a bit and a group of bits represent a single character; so this is called a byte. So, a group of bits that represent a single character is called the byte and then a collection of bytes put into a word form a field.

Similarly, group of related fields together form what is called a record. So, I will be explaining this entire hierarchy using a very pertinent example soon, but prior to that let us talk about what a group of records are called. So, a group of records of the same type form a file and finally, a group of related files form a database. Before moving on to the example let us understand what a record is.

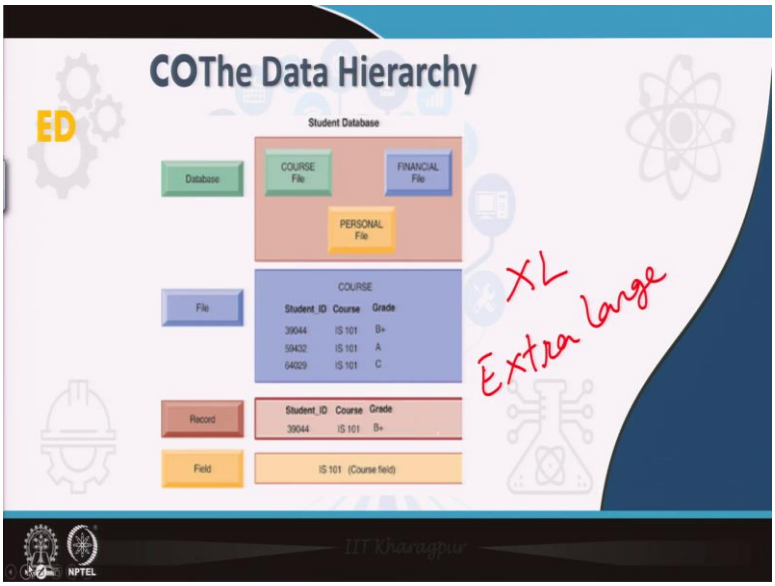
(Refer Slide Time: 02:11)



A record describes an entity; an entity could be anything, it could be a person, it could be a place, it could be a thing about which we want to store some information. An attribute or a column represents each characteristic or quality that describes that particular entity.

So, for example, if we are talking about a student entity a certain you know attributes that could represent or explain the student could be courses that the student has taken or could be the grades that the student has obtained. So, courses and grades could be attributes pertaining to the entity, student.

(Refer Slide Time: 03:01)



So, with this let us move on to the example. So, this particular slide talks about the example and explains whatever we have just discussed. So, first we will begin with the field; so here in this particular example, there is a field called course and the course field represents all the courses that the student has taken.

So, a group of fields such as; you know there could be a lot of other fields for the student entity such as grades that the student has obtained, the address of the student, the local guardian of the student. So, there could be a lot of data; age of the student a lot of other data pertaining to a student entity that our student different attributes; multiple attributes or fields.

So collection of these fields together form what is called a record. So, at the next higher level we see a record represents a student, each record represents student entity and record contains information about each student; such as you know the course, the grade as we can see in this particular example. So, at the next higher level when we go up; we see a group of related records form a file.

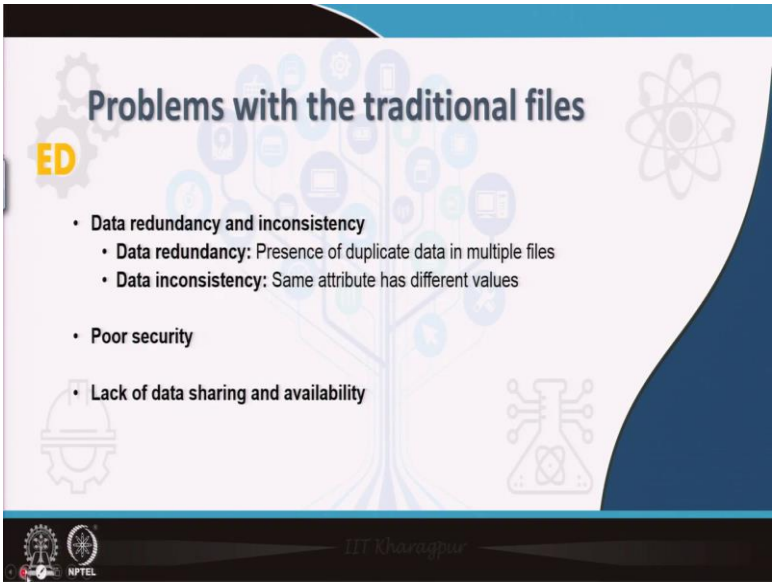
So, here we see there are multiple records; so student the first student, the second student, the third student and all their details here. So, multiple records are grouped together in this particular file; so this file is called the course file. At the next higher level or the highest level; we have groups of files which are related to one another and all of

them are of course related to the student entity. So, all of these files together form what is called the student database.

So, here in this particular example we see; course as a file, then we see personal details of the student as a file and we see financial details of the student in another file. So, these files are related to one another; for example, the course file could have information about the student such as the courses and the grades obtained by the student. Personal file could have data pertaining to the student such as his demographic background, his age, his say corresponding address, his permanent address and other details.

Financial record could have data pertaining to the student such as his average annual family income and other financial details. So, all of these files together form what is called the student database. So, this is all about the data hierarchy in which data is stored in a computer system.

(Refer Slide Time: 05:57)



But there are certain problems with traditional files; as we see in the previous example we have seen that there could be a problem of data redundancy. So, if we go back to the previous slide; we see that you know there could be same data residing in multiple locations; so that could give rise to something called a data redundancy.

For example, the personal file could have information, could have you know financial information about the student such as his or her annual average family income right; the same detail could also be present in the financial file.

So, there is unnecessarily overlap of data and data; same data is stored in two different locations, this also leads to a wastage of a lot of storage space. So, that is one major problem and the next problem related to traditional file system is called the problem of data inconsistency. So, what do we mean by this? So, if one particular data; the same data is stored in multiple locations, there is a problem of inconsistency or discrepancy.

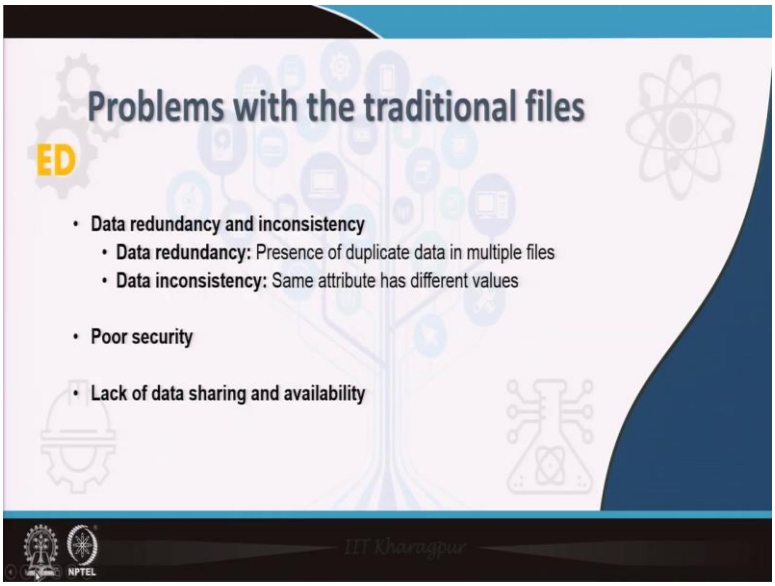
We will understand this through to a very simple examples. In this particular example of a student you know database what could happen is; so the course file could have a particular column or attribute which contains the role number of the student. So, the header of that particular column would be roll number; whereas the same student or the same details, the roll numbers of the students could also be there in the personal file as well as in the financial file.

But in the personal file; the header could be roll ID. So, though we see the same data present in two different files; in one file, it is mentioned as roll number, in the other file that is roll ID. So, there is a clear case of discrepancy here or inconsistency; so this is a major problem.

Another example where we see a scenario of inconsistency is you know in case of say; we take the example of a garment store or a departmental store which deals in garments. So, this particular store would be having a multiple files in its database; there could be a file related to sales, there could be another file related to distribution, there could be a third file related to manufacturing related information about different garments and in these files you know there could be a difference in nomenclature between say different garment sizes.

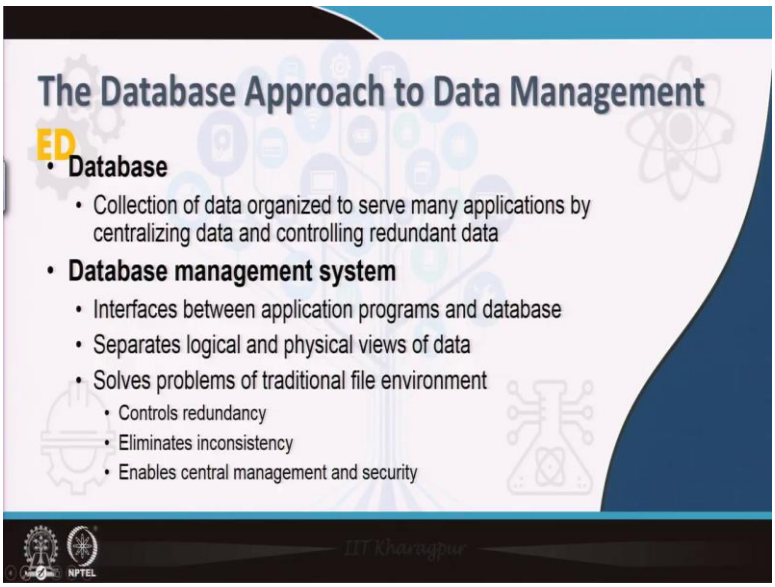
So, it could; the garment size XL could be represented at as XL somewhere and it could be represented as extra large in another place. So, that is again a clear case of; so it could be excel somewhere and it could be extra large somewhere else. So, this is again an example of a data discrepancy or data inconsistency which is present in traditional file system.

(Refer Slide Time: 09:03)



So, all of these happen because there is poor security as there is no central management of databases; what can happen is there is absolute lack of security, anybody can access the files, anybody unauthorized can use and can manipulate the files and all of these again lead to problem of data sharing and data availability. So, these are the problems with traditional file systems.

(Refer Slide Time: 09:31)

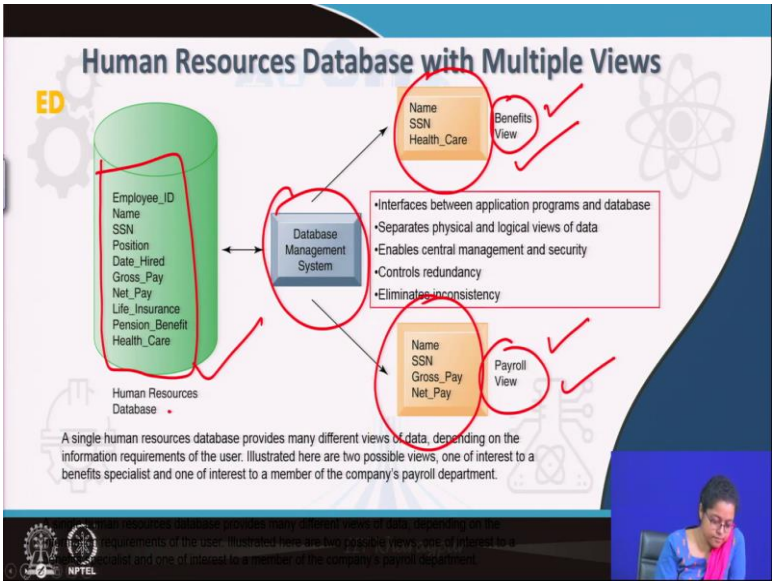


Now, all of these problems or many of these problems were overcome with the database approach to data management. So, we will quickly understand what this means. To begin

with what is a database? A database is a collection of data organized to serve many applications by centralizing data and controlling redundant data. So, that is all about the database and then a database management system interfaces between application programs and the database; at the same time it separates the logical view and the physical views of data.

So, what are these two? What do these two terms mean; logical view and physical view of data? Logical view of data means the way the data is actually visible or the way the data appears to a business person or an end user, but physical view means the way in which the data is actually stored right; so this is the difference between the two. Now, we will see this with the help of a very simple example.

(Refer Slide Time: 10:37)



So, this is an example of a human resources database which has multiple views. So, in this particular human resources database; we see that here there is a human resources database which has details about employees such as employee ID, name, social security number, position, date hired, gross pay and a lot of other details.

Now, there are two clear applications here; one is the benefits application and the second is the payroll application. So, the person who is involved or the end user who is involved with the benefits application need not have access to pay details or other details of the employee.

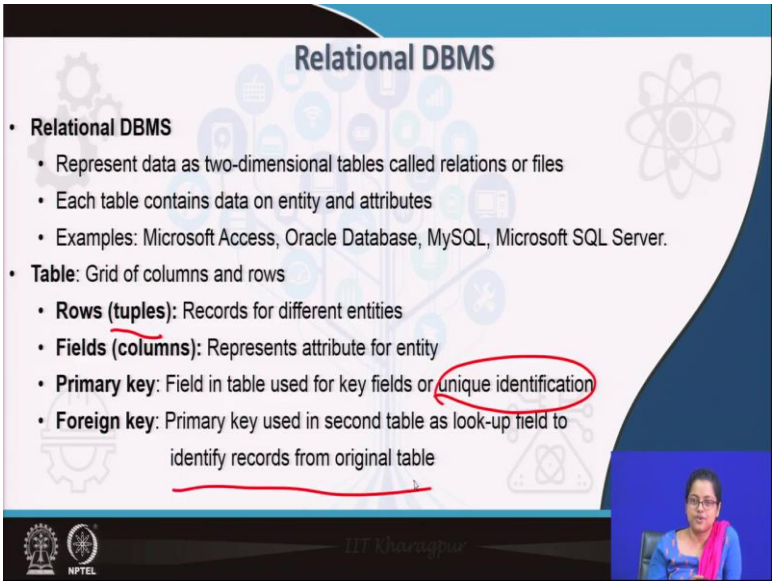
So, this person; the benefits person need not have a view of all the different you know records or different attributes that are present in the human resources database. Similarly, a person who is involved with payroll need not have much idea about what is there in the healthcare system of the particular employee right.

So, these are two distinct views; the benefits view and the payrolls view or we could call them two different applications and these two applications are visible to two different end users. Now, in between we have the database management system which plays the role of an interface between the human resources database on one hand and the applications on the other hand.

Now, here we see that the database management system that is there in the center; interfaces between the application programs and the database. Second, separates the physical and the logical views of data; as we see, the logical views are visible to the benefits personal and to the payroll personnel, but they need not see how the data is actually stored. So, now this DBMS enable central management and security, it also controls redundancy and eliminates inconsistency because there is hardly any overlap of data.

So, this is how a database management system helps eliminate many of the problems that are present within traditional file systems.

(Refer Slide Time: 13:08)



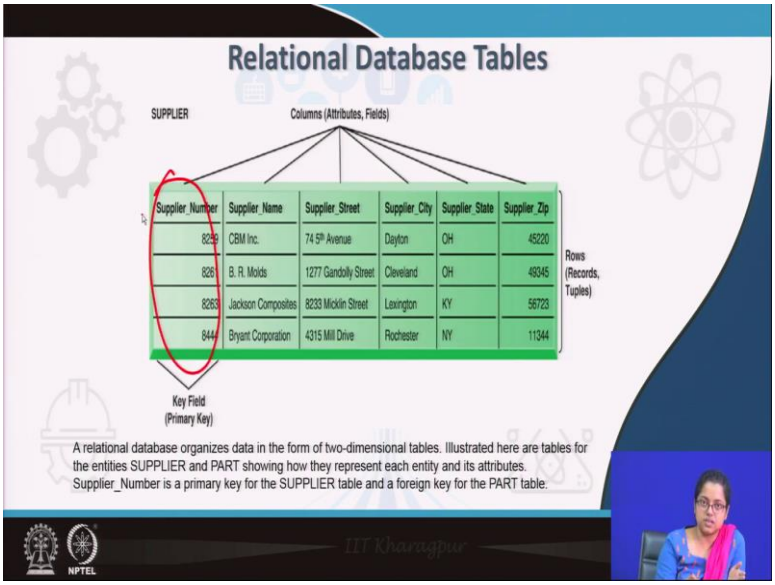
So, moving ahead; you know database management systems underwent another period of evolution and then we saw relational database management systems. So, relational database management systems as we see; represent data as two dimensional tables called relations or files and these files or relations have some sort of a relationship between them as the name suggests; so, each table contains data on entity and attributes.

Some very popular examples of relational database management systems are Microsoft Access which is extremely popular, it is used for desktop applications, Oracle Database, MySQL which is an open source DBMS and we have Microsoft SQL Servers; these are some of the very popular ones.

Now, what is a table? Very important; a table is nothing, but a grid of columns and rows. What are rows? As we saw with files; rows represent records for different entities and rows are also called tuples technically. Fields also called columns or attributes represent an attribute for an entity. Then we come to something very important which is called a primary key.

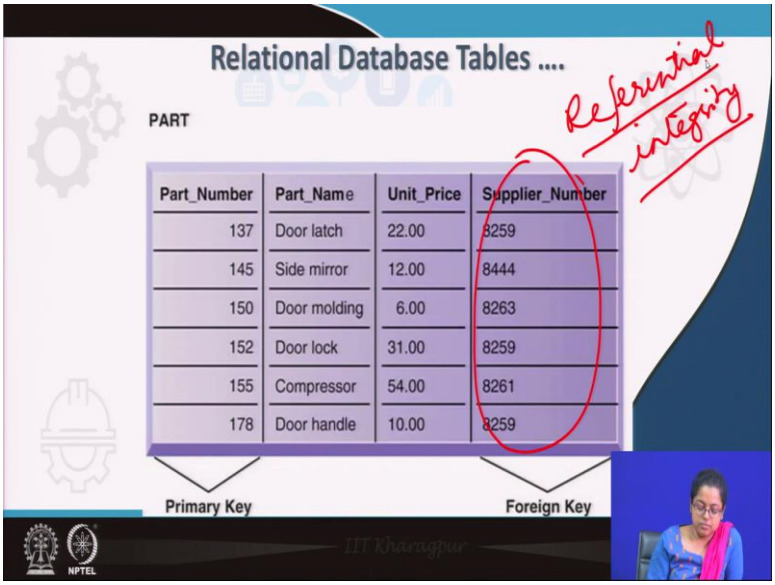
A primary key is a field in a table used for key fields or unique identification; what this means is a primary key is used as a unique identifier of a particular table. So, if there are multiple students in the student table; then roll number could be a primary key because it will uniquely identify each student. I will come to foreign key, but prior to that let us see this entire thing that I just discussed with the help of an example.

(Refer Slide Time: 15:00)



So, here we see a table; this particular table pertains to entity supplier and here we have various attributes. So, here we see different columns or attributes for multiple suppliers. So, we have supplier number, name, street, supplier city, supplier state and supplier zip and these are all records pertaining to multiple suppliers. Now, this field called the supplier number is the primary key because it uniquely identifies each supplier.

(Refer Slide Time: 15:45)



So, along with the supplier table; we also have another table called the parts table. So, the part table has information about different parts that are supplied by the suppliers. So, for the part table; part number is the unique identifier therefore, it is the primary key and supplier number here; this attribute supplier number is present in the previous table here; it was the primary key; this is the parent table the supplier.

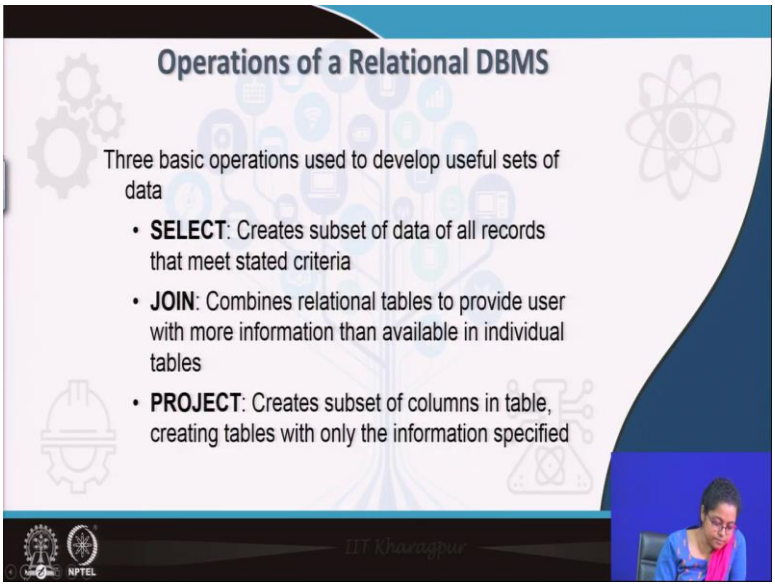
And in the child table that the related table that is part; supplier number is called a foreign key. Why this is called a foreign key is because this particular attribute that is supplier number in the part table is used to lookup or match the supplier number in the supplier table.

So, this particular lookup column is called the foreign key. So, foreign key here is the primary key used in the first table that is used in the second table as a lookup field to identify records from the original table; as you see here, I hope this concept is clear.

Now, here I have also very briefly talk about something called referential integrity. What this means is; in this particular you know these two tables that is the supplier and the part; the supplier number is the foreign key for the part table , but as you know the primary table that is the; that is the parent table here supplier number is the primary key.

So, there cannot be any supplier number here in the part table that does not have a; you know that cannot be matched with a record or an entry in the previous table, the parent table that is the supplier table. So, every supplier number that is there in the part table should be present in the supplier table that is the concept of referential integrity constraint. I hope this concept is clear.

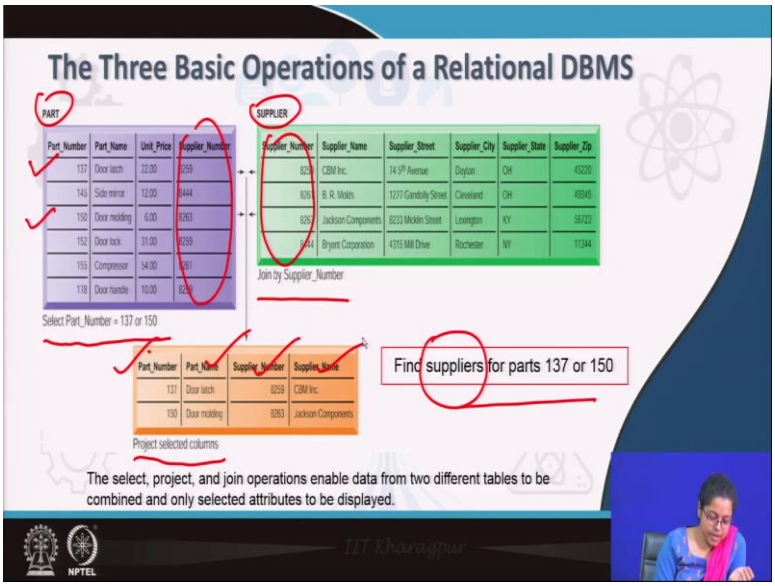
(Refer Slide Time: 18:05)



So, let us move ahead; operations of a relational database management system, there are three primary operations that are used to develop useful sets of data. So, we have the select operation, then we have the join operation and thirdly we have the project operation. So, select operation creates subset of data of all records that meet certain state and criteria or conditions. So, I will explain each of these with an example to help you understand better.

The next one; the next second operation is called the join operation. So, the join operation combines relational tables to provide users with more information than is available in the individual tables. And the third operation called the project operation create subsets of columns in table; creating tables with only the information specified.

(Refer Slide Time: 19:01)



So, now let us go ahead and see with an example. So, here we see first we have the part table here; so here we see the first operation that is the select operation. So, from the part table; we would want to select only part number all the details from the part table for only the part numbers 137 or 150; not for the others. So, what will happen is, this select operation here will select only the part number 137 and 150 and display their details, fine.

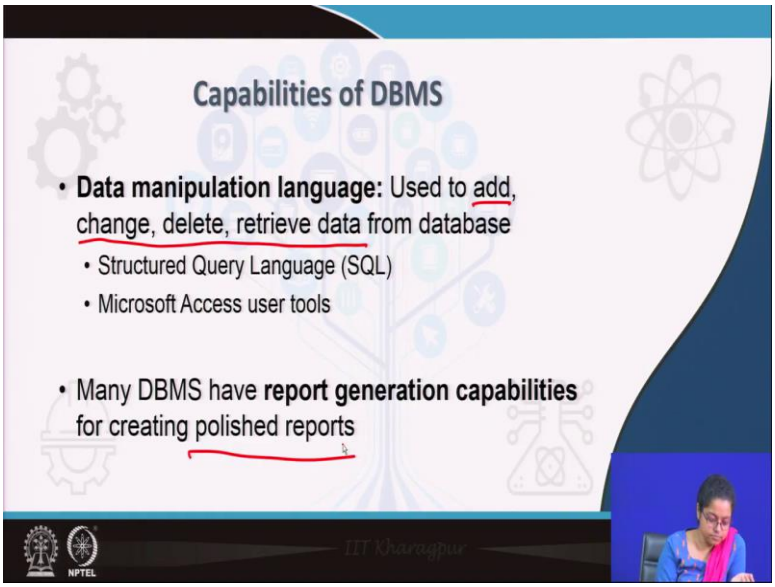
Now, coming to the second operation that is join; join operation is used to map the two tables, the supplier table and the part table. And what it will do is it is going to match with the help of the lookup field that we saw before thus that is the supplier number. With the help of the supplier number, it is going to match the two tables and it is going to create a huge table which has only two records; that is 137 and 150 part number and has multiple attributes; huge number of attributes.

So, all of these attributes of both the tables that is the part number, part name, unit price, supplier number, supplier name, street, city, state and zip; all of this information will be there in the join operation.

Now, post the join operation comes the third operation that is the project operation. So, now what we want to do is; we only want to find suppliers for parts 137 and or 150. So, we want to find the suppliers only and not their other details such as; you know city, zip etcetera right.

So, the project operation is only going to give you certain selected attributes from that huge joint table. So, here we have only part number, part name, supplier number and supplier name; fine, I hope the three operations are clear to you.

(Refer Slide Time: 21:12)



So, let us move ahead; now capabilities of database management systems. So, database management systems are; have a lot of capabilities as we have seen before we; there is something called data manipulation language.

So, data manipulation languages are used to add, change, delete, retrieve data from databases right. And very popular data manipulation languages are structured query language, as well as Microsoft Access user tools.

So, structured query language is used to you know perform addition operations, change you know modification operations, deletion, retrieval of data from databases, from different DBMSs such as MySQL such as MS Sequel server, such as Oracle; all of them use such structured query language.

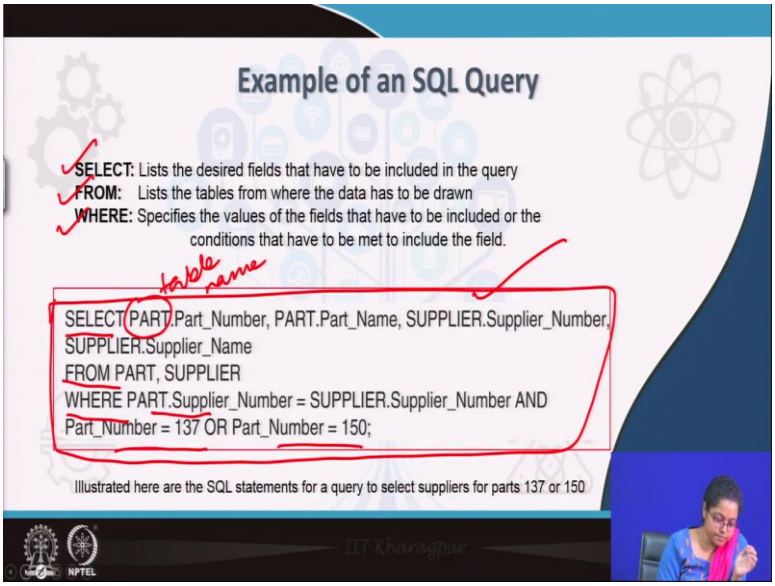
And Microsoft Access is an inbuilt; you know package that you see along with the entire office package. So, in the office package; you have Excel, you have Word; along with that you also have Microsoft Access right. So, Microsoft Access is an in is inbuilt with your office package and there you have tools for all of these; changing, deleting, retrieving and adding data from the database.

So, another very important capability of a database management system is report generation. So, DBMSs not only help you manipulate the data; they also help you to create very beautiful, polished reports out of the databases right.

So, and these reports that are generated are actually you know they represent certain calculations on the data that is present in the databases. At the same time, these reports can have some; you know they can have some graphs, they can have some charts which are actually used to give you a very beautiful graphical and pictorial representation of data. So, as we know you know a picture speaks a thousand words.

So, when you represent data in form of a; in form of say a graph or a chart, it represents the data much better then storing data in databases right. So, reports would help you; you know make the data much more vibrant and make them much easier for the users to understand. So, these are the different capabilities of DBMSS.

(Refer Slide Time: 23:57)



Moving ahead, example of an SQL query. So, an SQL query as we saw prior to this that SQL queries are part of data manipulation language. So, here we will see; what we will just get the flavor of an SQL query right. So, an SQL query here will be used to perform the exact three operations that we have seen some time back; such as the selection operation, the join operation and finally the projection operation.

So, and that will happen through an SQL query here. So, SQL query has three basic parts; the ‘selection’, the ‘from’, and the ‘where’; these are very important; so, let us quickly understand what these mean.

So, selection operation lists the desired fields that have to be included in the query right; the desired fields. The ‘from’ clause here lists the tables from where the data has to be drawn. So, if there are two tables that are involved say, supplier and part, the ‘from’ operation or the ‘from’ clause will talk about those two tables and the third clause here is the ‘where’ clause.

So, the where clause is very important because it specifies the values of the fields that have to be included or the conditions that have to be met to include the field. Let us see with an example; here we have a very simple illustrative example of an SQL query.

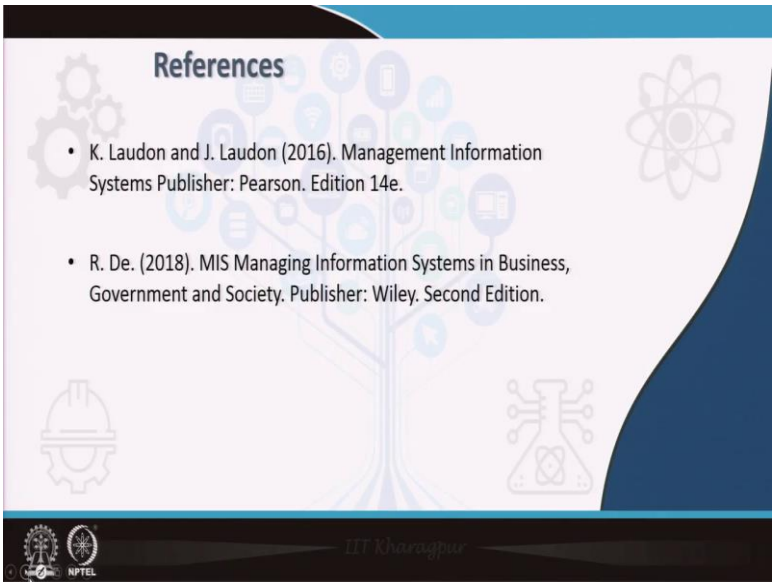
So, this exactly does what we had seen few slides back the three operations right. The selection operation where we had selected say part number 137 and 150 and then we had joined it with the supplier database and eventually we had projected only four fields; that is the supplier number, supplier name, the part number and the part name right.

So, we see this; select, this is the first clause; part dot part number. So, part represents the table name and again part dot part name; so these two are from part table; then supplier dot supplier name number and supplier dot supplier name from the supplier table. So, all of these attributes come from; this is the ‘from’ clause; the second clause, the part table and the supplier table. And thirdly the ‘where’ clause where we actually do the matching or we do the; we call it the lookup or the mapping; right.

So, where part not supplier number would match with supplier dot supplier number right only where this mapping would happen or this matching would happen and where part number is 137 or 150. So, this will give you the exact you know all this entire SQL query will perform the exact operations that the three operations we have seen in the previous slide performed.

So, selection then joining and finally, the projection right. So, this is an example of a simple SQL query.

(Refer Slide Time: 27:17)



So, today we have spoken about databases and how they are used in information management. I have primarily followed two books and taken some references from there; Laudon; MIS by Laudon and Laudon, and Managing Information Systems in Business, Government and Society by Prof. Rahul De.

So, in the next session, we will be taking up; we will be discussing data warehouses and how they can; how an organization can derive business intelligence from data warehouses.

So, thank you for now!