Course on E-Business
Professor Mamata Jenamani
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur
Module 05
Lecture Number 27
Data Resources

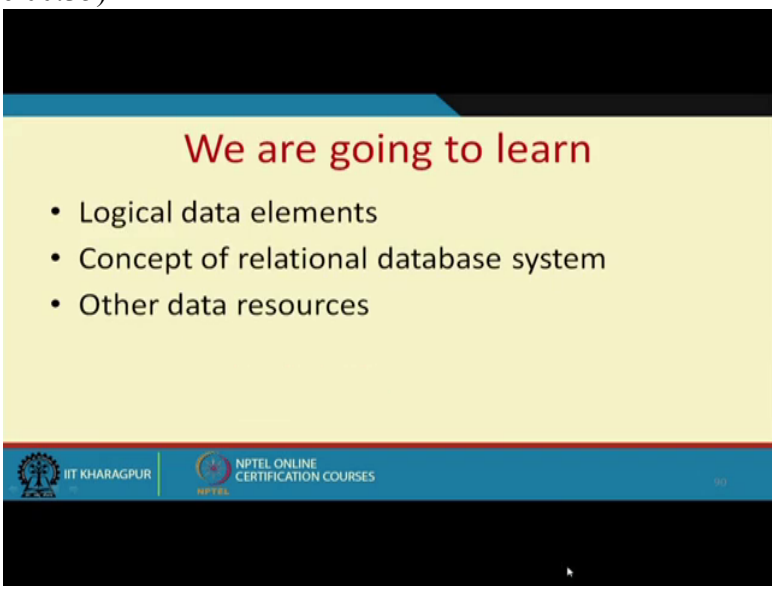
(Refer Slide Time 00:18)

Welcome back, in this series of infrastructure now we will talk about the data resources. So far we have talked about the networking resources, hardware resources and software resources. The next one is our data resource.

(Refer Slide Time 00:34)

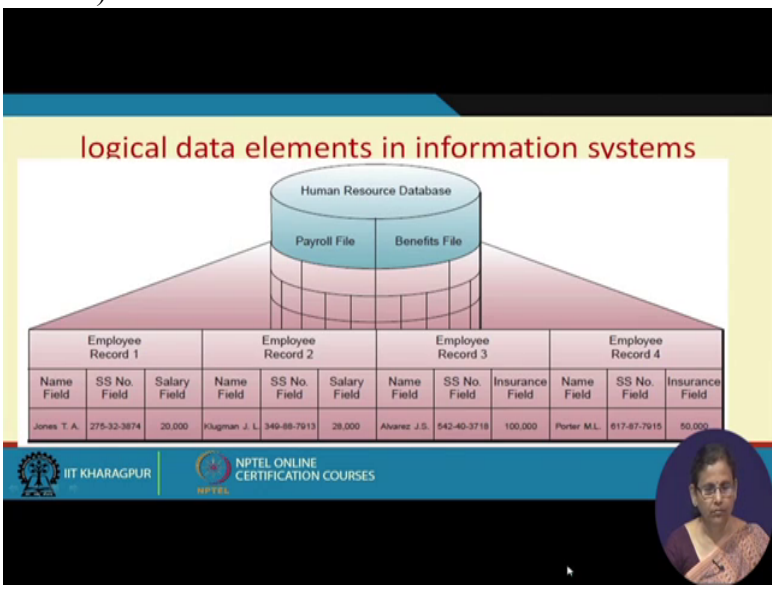
So in this lecture

(Refer Slide Time 00:35)



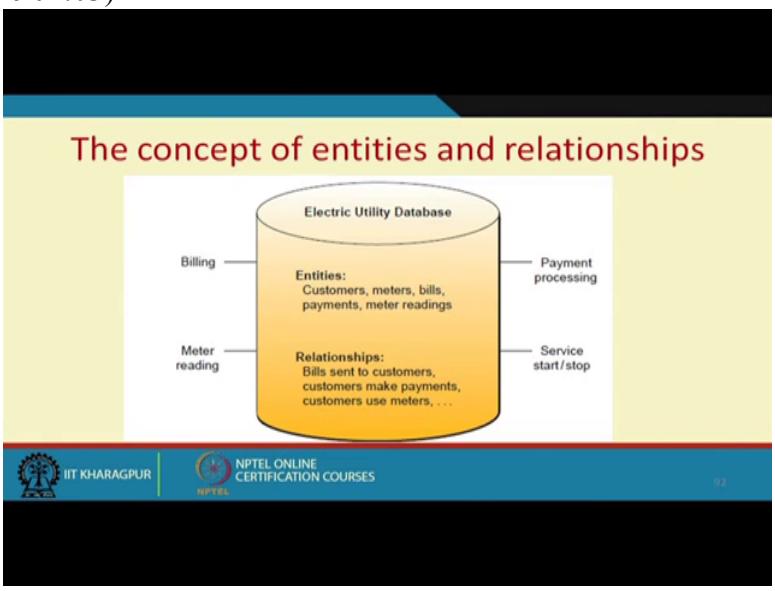
we are going to learn about different logical data elements, then the concept of relational database system which is fundamental to, which the fundamental data, fundamental to maintaining the data resource of the organization, then various other technologies for keeping the data.

(Refer Slide Time 01:00)



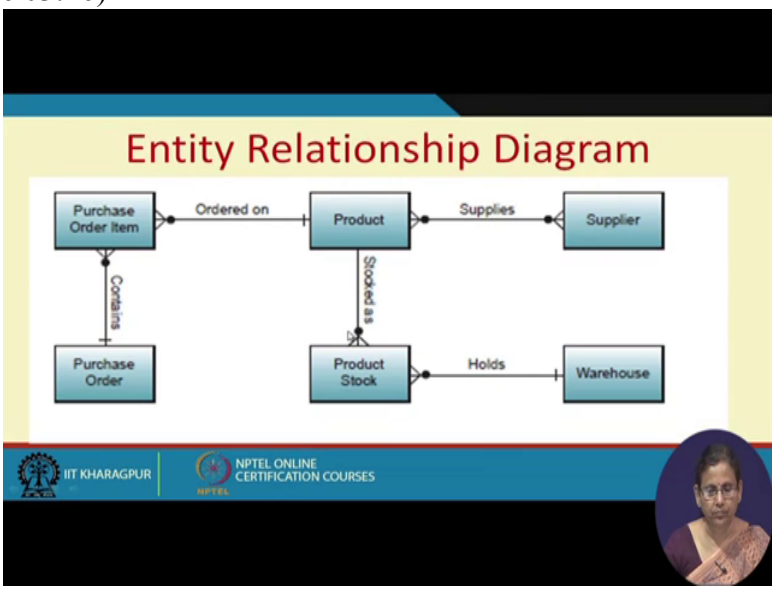
If you look at data, the way it is stored, and the logical elements, the top, at the top we have the database. A database consists of many files, for example a human resource database may consist of a payroll file, a benefit file and so on. Each of these file contain many records. For example here in this case there can be records for the employee Employee 1, Employee 2 etc. so each record. Each record again has many fields. So these fields, records, files and files together many files make one database makes a complete database for your organization.

(Refer Slide Time 02:03)



Now within the database, when you keep the data, you keep the information about both entities and their relationships. By entities we mean the elements such as the customers, let us say this is one electric utility database, the customer's meter reading, billing, payment, bills, payment etc. This data comes from various sources, it may be, and it may be going to various other sources as well. And there is some kind of relationship among these entities which connects two entities. For example a particular meter will belong to a particular customer. The bill will relate to the customer and so on. So these relationships are also kept in a database.

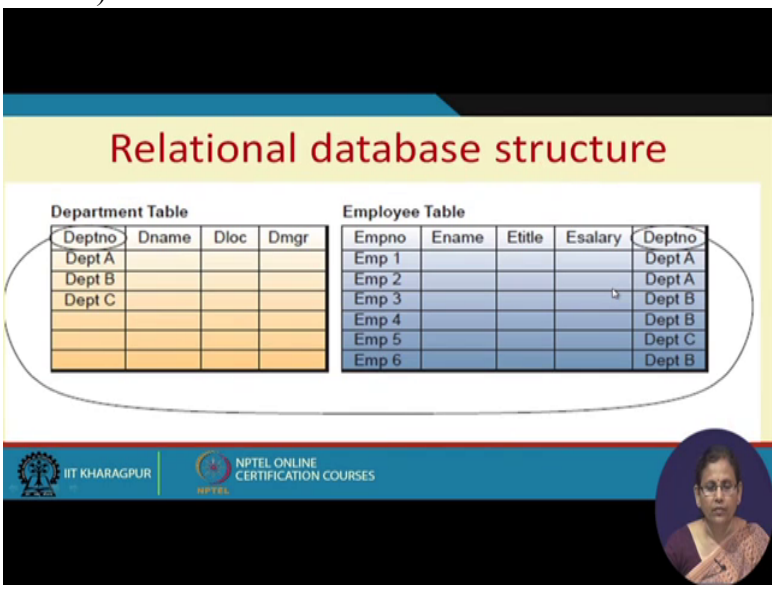
(Refer Slide Time 03:16)



These entities while designing a database, these entities and relationships are first discovered. This particular diagram which is called an entity relationship diagram shows how various

entities can be related with each other. For example here the products and suppliers they are related with each other. Product, your supplier supplies the product and the relationship can be either one-one, many-many or many-one. For example, this supplier and product has many-many relationship which means one supplier can supply many products and one product can be supplied by many suppliers.

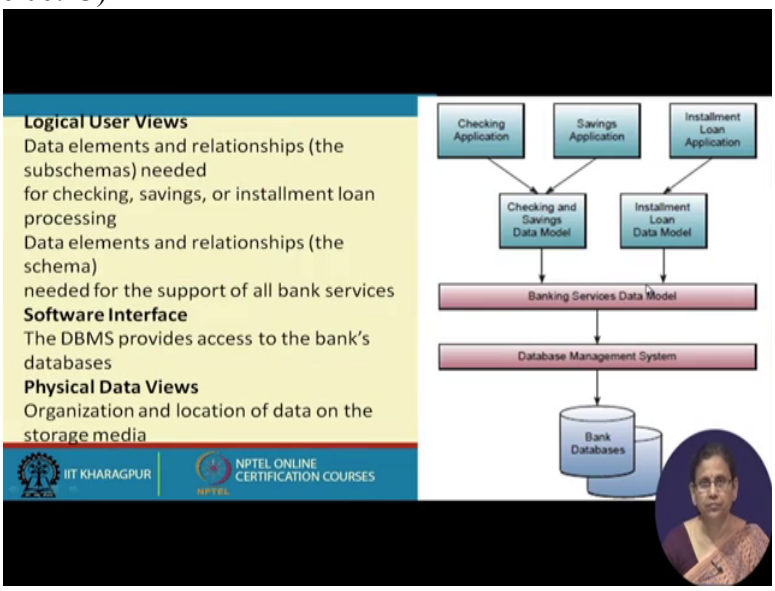
(Refer Slide Time 04:10)



Relational database keeps the information about entities in table format. If you look at it it is a two dimensional format in which the, we learnt the, the, we learnt that the basic elements of the, of keeping the data are the files. So each table is a file which in turn contains the records and each record contains the fields. So these fields make up record and many records make one file. In this case, it is a, this file can either be a text file which typically is not the case because of the kind of search and other functions has to be executed so it has to be stored in a specifically designed file called the, called the database files and these databases, the relationships, entity and relationship in a database can be put in various ways. There can be, and the structure of the database can be accordingly made for example there can be a network kind of database but of the databases actually relational databases which puts the data in the form of puts the data and the entities in the form of tables and as an example here, may be the relationship each department I mean the relationship between department and the employee can be established. So by establishing this relationship you can, establishing this relationship actually avoids many kinds of anomalies that can happen. For example if you have three departments like Department A, B and C and suddenly somebody tries to enter the department of one employee as a department D where as department D does not exist. So if

the relationship is made then proper arrangements can be made so that such situations can be avoided.

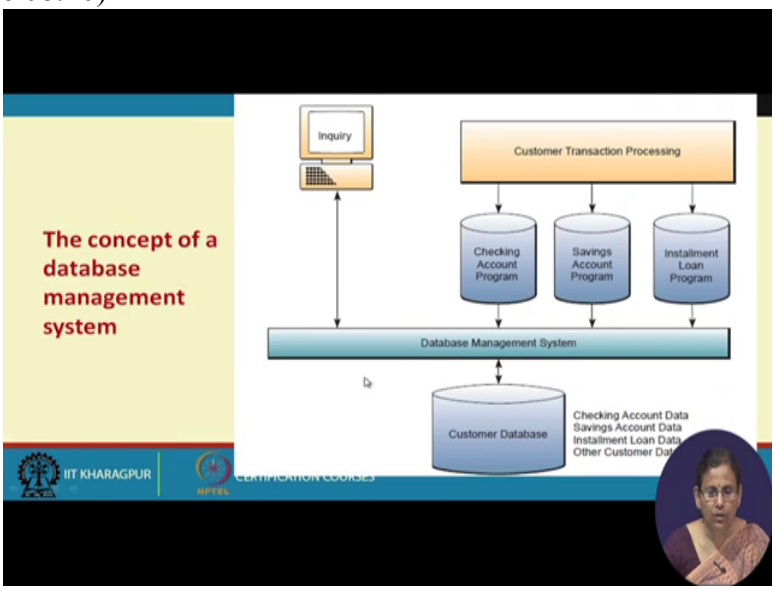
(Refer Slide Time 06:43)



Then organization database when it is kept, it is kept in three levels. First in a very lower physical level the it deals with what kind of storage media the company is going to; we have already talked about various kind of storage media. Your databases can reside in those storage media. So within the storage media in the physical level how the data will be stored relates to the physical view of the data. Then there has to be some software interface to access this data and manage this data. So this software interface is called, is a database management software. Then at each user level depending on their, depending on what kind of permission they get they will be responsible, they can actually access fully or partially the database. So at the logical user level they would be seeing only the entities and relationships which they are entitled to see.

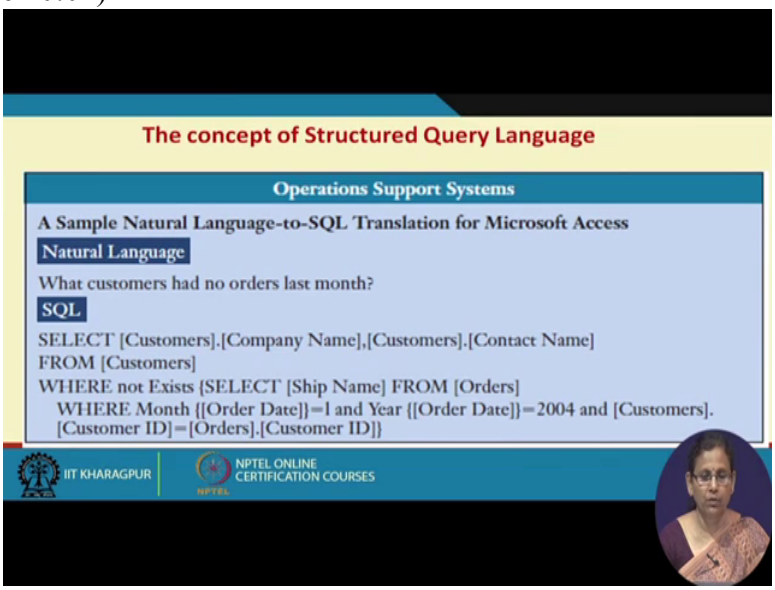
These are few example situations which you can look at,

(Refer Slide Time 08:10)



look. Then this data as I was telling you this database management system is responsible for connecting to these physical databases and it provides many facilities to the, for maintaining and for storing maintaining and visualizing this data. And these databases we have talked of transaction processing system and in fact let us talk little bit about the E R P system because that time we were talking that how E R P actually manages the flow of data among, while conducting while performing automating interdepartmental business processes. So that time we discussed it has to have a centralized database which in turn is attached to with which in turn is managed by a database management system and as the transactions occur your E R P provides some kind of interfaces or that way any kind of it is not only your E R P your database is behind almost all the application whether it is a transportation management system, warehouse management system everywhere database is there. But whatever may be the case whenever any transactions whatever may be the situation whenever any transactions happen this data is being kept in the database system and once it is kept if any inquiry is required then some program can be written which is called S Q L programs

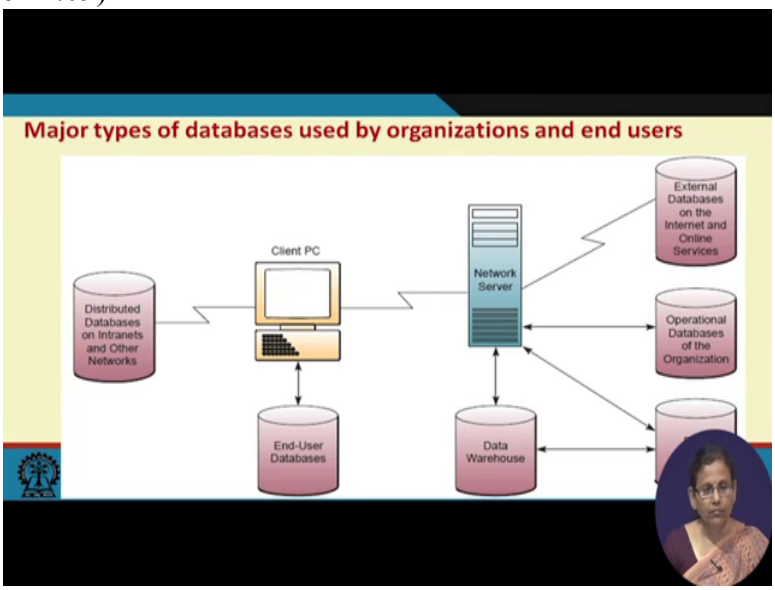
(Refer Slide Time 10:02)



which can be written to access this data. This is when we talk about the think about the programming languages. S Q L is a is a language to access the data we are not we did not have any discussion on the different kinds of programming language and we are not going to have that discussion as well but this is a fourth generation language which is does not require which is very simple to code once you have the appropriate logic of converting a natural language sentence to correspond using corresponding facilities provided by S Q L. Usually the S Q L queries will presented as a select from where form. For example here question is one example is given what customer had what customers had no order in the last month. Let's say your customer relationship management, it's a part of customer relationship management and your marketing guy is trying to figure out who are the customers, let's say, your retail store and regularly the customers buy stuff from you. So as a part of customer relationship management program, your, your sales personnel would like to find out whether the person has made any kind of purchase last month. If he has not made possibly they would be taking some action. But whatever, we are not going to that part of it, this particular query can be written using this specified S Q L format, using this select from where structure, structure to generate output for this query which is given by the user, which means somebody has to write this query. So in case you are using a web based system these queries are to be pre-written. Only thing is that this parameters corresponding to these customer, company name etc has to be provided. So maybe you can have one web interface where simply this natural language system, natural language query is written and once your user clicks on it then corresponding S Q L which is precoded gets the parameters from here and this query is sent along with your H T T P request. We were discussing web-based system which is a three-tier architecture with

a web server followed by one application server and finally a database server. So the query, if you click on this natural language query, it is, the information is sent to the server and in server already some pre-coded application will be there with this query. The parameters for a specific user who has fired the query will be inserted into this pre-coded S Q L query and the query will be made to the database server, database server will produce the result in certain format, and it has to be reformatted, and inserted into a web page for generating a dynamic page which will be sent finally to the user.

(Refer Slide Time 14:09)



Besides these relationship database systems which is behind every application, there are other major databases used in an organization by its end users. So they are, they can be your individual end user databases, they can be distributed databases, they can be your various, just now we were talking about your transactional databases that we have just now talked about, then you can get from some external databases like your stock prices etc. On the internet you can get that and combining all these data sometimes the organizations, in fact most more organizations are now going for data intensive decision making. So it is, it has become essential for the organizations to create data warehouses and data marts out of this data.

Now the difference between these data warehouses and data marts is data warehouses is the organizational repository of data from, of different kind. But data mart is something which you can think of as a data warehouse which contains data corresponding to a particular area. Let's say you have a data mart for all your sales and marketing operations. You have one data

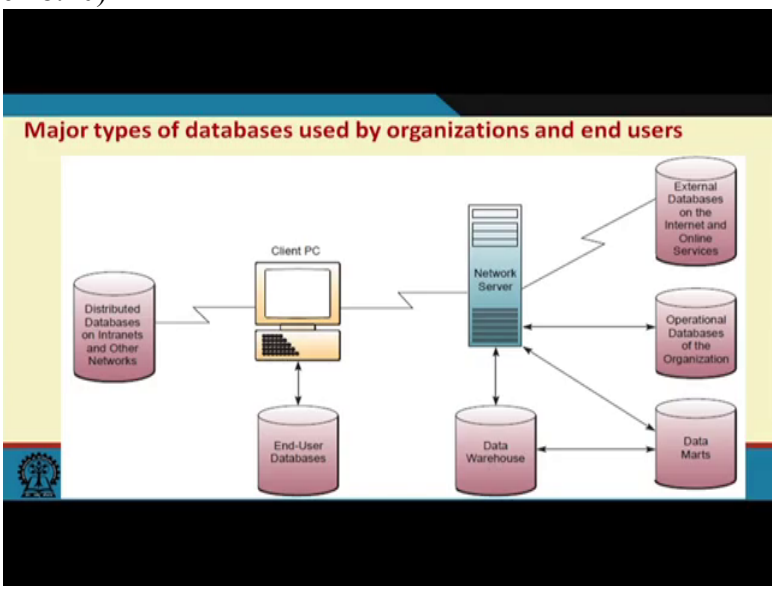
mart for your operations management activities for managing your manufacturing operation and if you combine all this data, every organization has four subsystems. It has finance, it has manufacturing, operations. It has, your sales and marketing, it has human resources, if all these data, and all these data are related together and if they are put, combined together and put, we call it a data warehouse.

Right now, you might be confused that in a centralized, in a E R P system, in a centralized database we also put the data about various, we put various type of data together, we also put your finance accounts data, we also put your human resource data, we put operations data, we put sales and marketing data, so on. Unless otherwise you put them together, automating inter-organizational business processes is not possible. Then what is the difference between storing the data in E R P and again taking this data and putting into

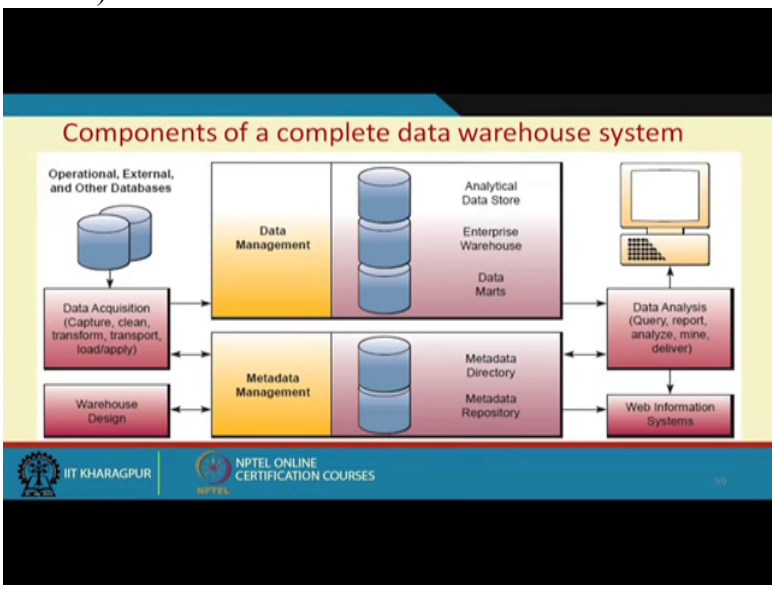
(Refer Slide Time 17:15)

the warehouse. The idea, while showing, talking about this relational database systems which is a typical way of storing the organizational data we saw that it is actually two dimensional it gives a two dimensional view of the data. You have, in the rows, you have the entities, the records and in columns you have specific fields. Now if you would like to connect two tables to generate a multi-dimensional view of the data, the S Q L queries generally become very complex. So therefore if you have to have a specialized query system, first as case you have to create a multi-dimensional data. So most of your time, your warehouses helps you creating this multi-dimensional databases.

(Refer Slide Time 18:20)



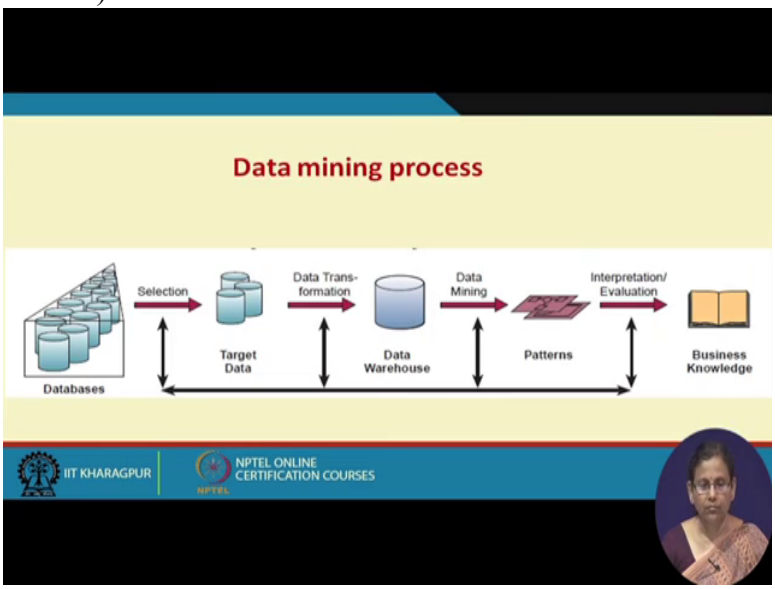
(Refer Slide Time 18:21)



So this, you can say is components of complete data warehousing system. You have your data from various sources which are put together and you have another system for managing your meta-data. Now what is meta-data? Meta-data is basically data about the data. For example, consider a customer record. A customer record consists of many fields, its employee code, its name, its department and so on. So this individual column names, and how the data is organized, how the data, your relationship is set among this data entities, all this data has to be kept in certain repository as well. So these such repositories are called meta-data repository or data, meta-data directory. Now your warehouse has both the data as well as this metadata about the data. And data comes from various sources like your operational databases and so on.

And after all, after this appropriate, your warehouse design, you can have a proper query system which can be your S Q L query, which can be your o l a p query, which can be your specialized business intelligence applications which will be mining this data and send this to the end user. And many times this data will be after analysis, will be sent to the end user through web based information system. In fact we were talking about enterprise data portals where all the employees through, employees have a common portal through which they can access various kinds of data for different applications for which they are entitled to. Depending, based on their role, they will be given access to different category of data. And those portals were also responsible not only for getting data from data warehouses and all; through some web-based system they were also getting the data from external sources.

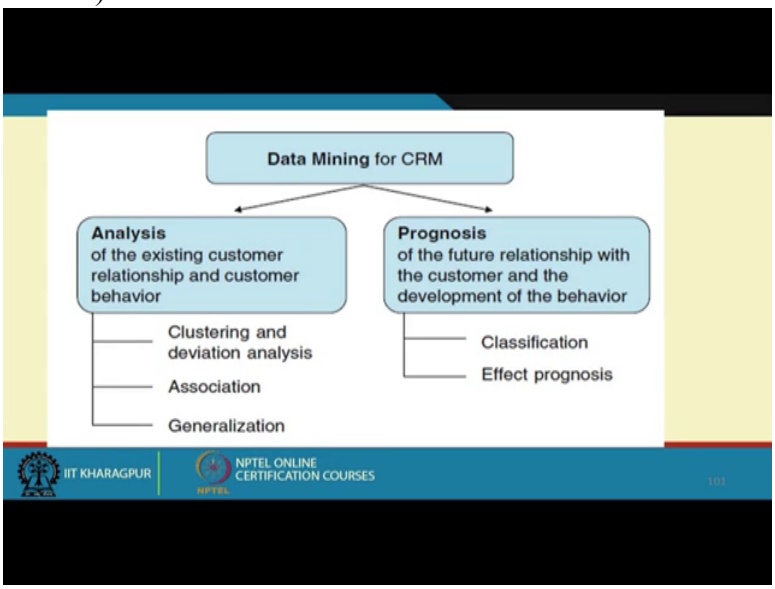
(Refer Slide Time 21:09)



Now once the data is stored in these warehouses or in the databases, while talking about this differentiating between this data and information and knowledge, we have already defined what knowledge is. But I would like to repeat here is whenever you send any query to a database it is not knowledge discovery. It is about generating the information from the data. Information will be, simply relates the data elements and brings it to a some meaningful form. But while firing a database query you should be knowing what information you would be like to get. So accordingly you would be writing the query whereas there might be certain hidden facts which you are not aware of. So therefore this concept of query etc doesn't really come into picture; because if you know, then only you write the query. But if do not know what to get, it is difficult. So therefore, there are softwares which are called data mining softwares which will be using this data from the data warehouses and find out the interesting

patterns out of this data. And they are appropriately; if they are appropriately interpreted they generate what is known as business knowledge. So this is how your business intelligence is, intelligence comes into picture.

(Refer Slide Time 23:11)

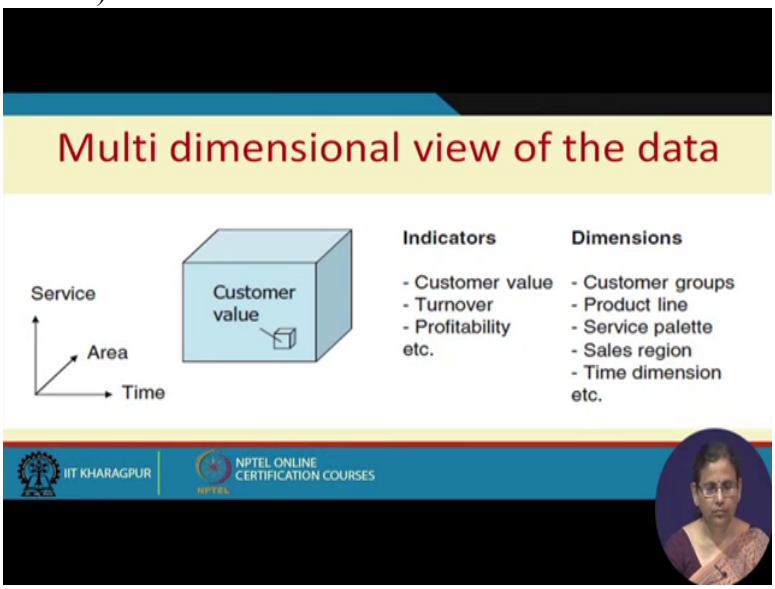


For example, consider while talking about this customer relationship management we were discussing about analytical C R M. And at that time, we told that in analytical C R M, you simply do not have the transaction details or you write some queries for which you already know that such data exist but you also try to discover some interesting knowledge out of it. So here, during this data mining process, let's say in C R M, you can analyze the existing customer relationship or customer behavior and try to find out various techniques such as clustering and deviation analysis. For example if you apply clustering on the customer's data and you find out a small cluster which is different from other regular customers. Then what do you say? This special group of customers must be doing then you will be analyzing the data that why this has been clustered separately from others. Once you try interpreting this cluster, probably for example you may find out that this is a group, this is a group which is, which has been regularly getting items from your site and they are most frequent purchasers. So they should be included in your loyalty program. Similarly it can also have, you can also apply some approach to find out the association among data elements.

For example in one of the earlier classes, we have, we were discussing about finding the frequently purchased items together and once you find the frequently purchased items together, you can take many strategic level decisions for improving your sales. So in the,

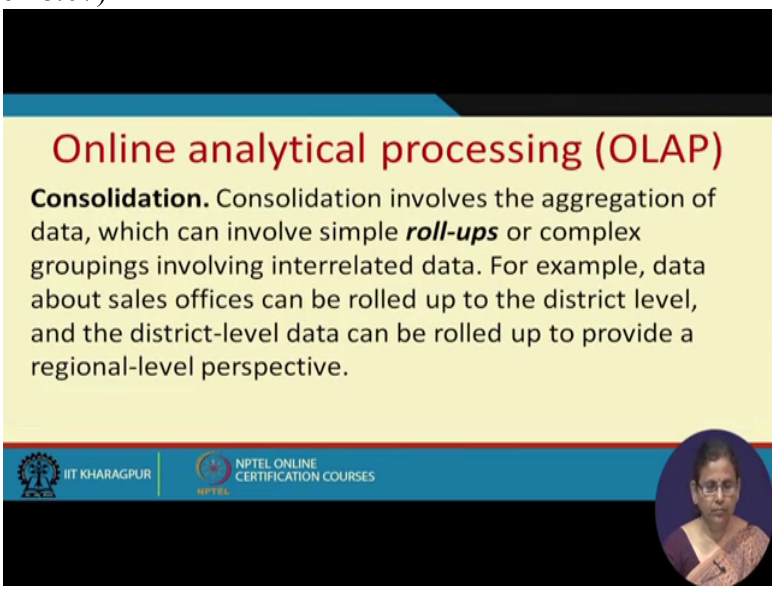
when the customer puts the certain items in his shopping cart, probably you are going to show the advertisement for the related set of products which often other users buy together. So such kind of association can be also be discovered. Then it also does some kind of prognostic stuffs which helps deciding certain your future set of activities based on your customers' behavior. For example, you find that if the customer's character is, customer has characteristics x, y, z etc he is likely to buy the product. If such customers, buying pattern of many customers are put together and you can probably train some kind of classifier to see that whenever a new customer comes to your store, in fact we are talking about your recommended systems in one of the later lecture we will be talking more about these kinds of activities.

(Refer Slide Time 27:01)



Now I was telling you that the data that is stored in a regular relational database system is actually two dimensional in nature. So therefore if you would like to integrate the data from multiple dimensions like representing, for example here we are talking of sales database where we are trying to look at the sales data from three angles, time, area of sales and service provided to the customer. Now they will be in separate databases, in two dimensional databases then combining them together to build a multidimensional view of the data is often carried out by your data warehouses. So the indicators like turnover, profitability, customer value etc which depend on many dimensions can be seen together which helps in decision making.

(Refer Slide Time 28:07)

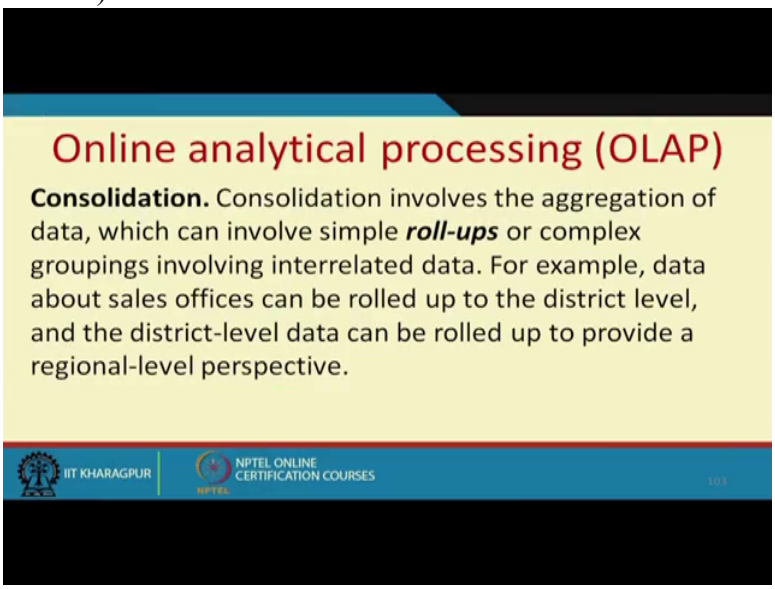


Now to generate this multi-dimensional view of the data, a specific class of processing is used which is called as online analytical processing. This online analytical processing has, processing has queries which are called OLAP queries which are different from your S Q L queries, structure because in your two-dimensional database, we have these S Q L queries, now these OLAP queries are actually, in case you have a multi-dimensional database you can actually run these OLAP queries. These OLAP queries, there can be various category of OLAP queries, first one will be consolidation. So this consolidation sometimes has more specific name for a special work called roll-up. For example here if you have the data, sales data and you would like to view the dist, you have the entire sales database but you would like to view the data in the district level. Then you have to roll up the data or

(Refer Slide Time 29:21)

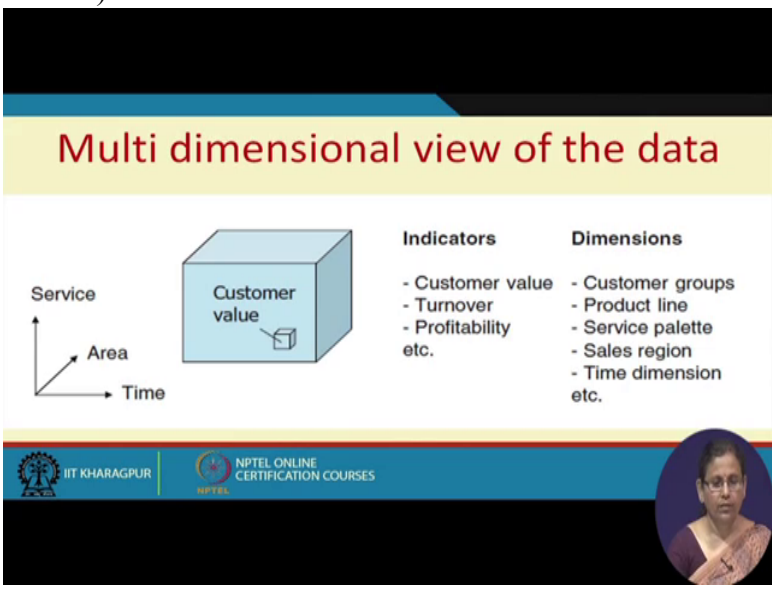
group the data so that you can view them together. Similarly from the district level, if you would like to see that, in the state level how you are performing. Then again you will be rolling up this district level data and put them together to see that how you are performing at the state level. Possibly you can even roll it up to see that what is your region-wise sales performance, Ok.

(Refer Slide Time 29:51)



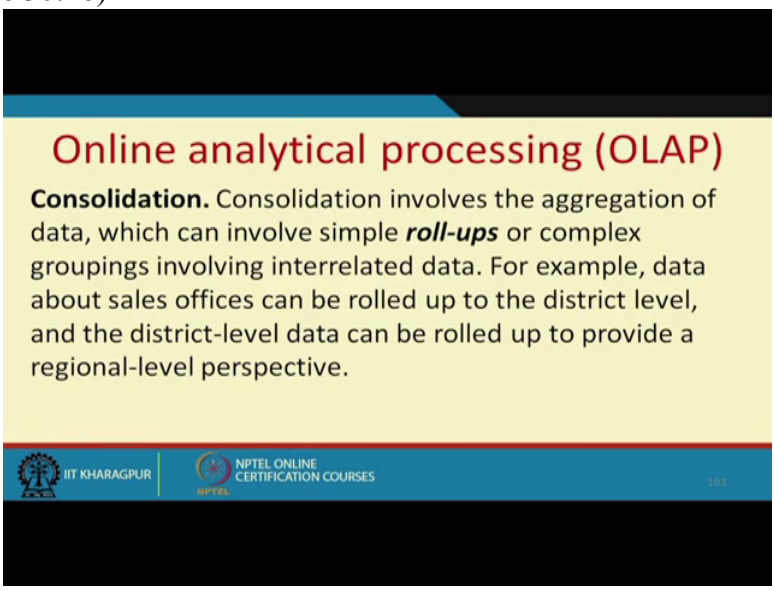
So this is, rolling up is just one example but this is one example of grouping the data. There can be other way of grouping the data as well. And together all these grouping activities are called data consolidation. So there can be many queries for data consolidation.

(Refer Slide Time 30:10)



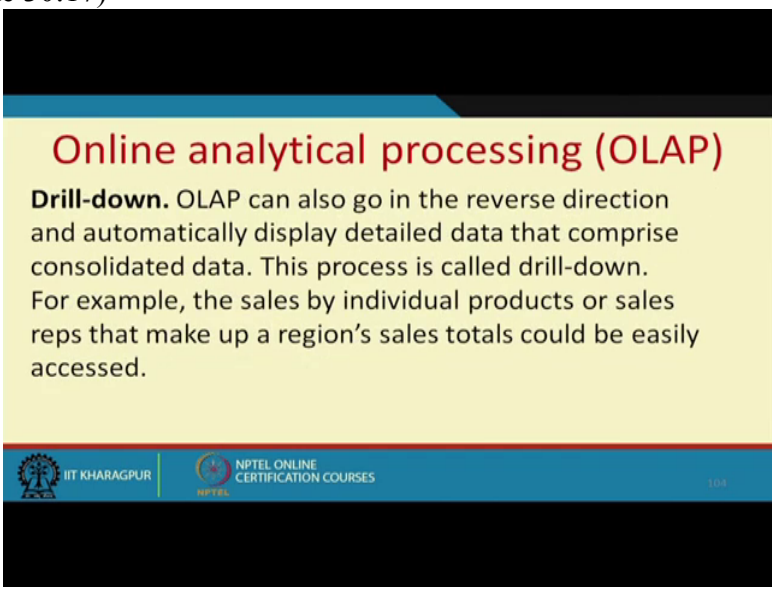
So

(Refer Slide Time 30:16)



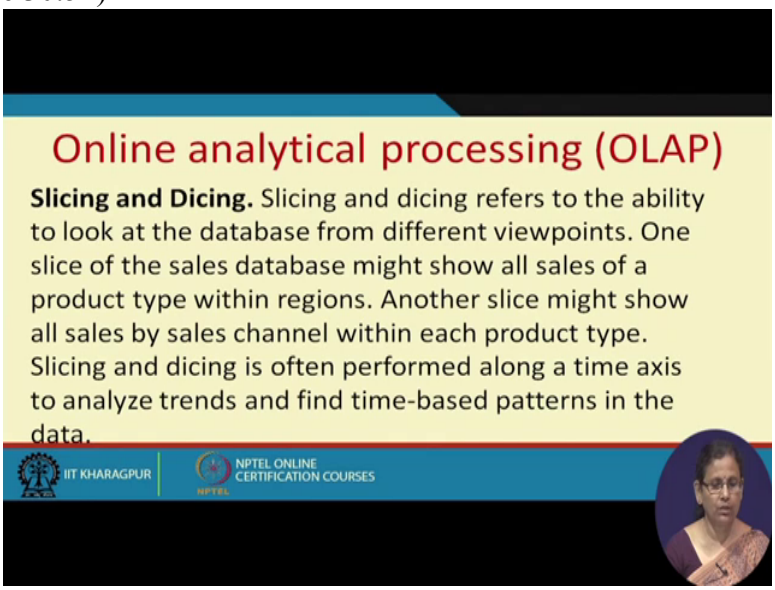
the second category of query

(Refer Slide Time 30:17)



is actually drilling down the data. It is just the reverse of rolling up which means suppose you have the country level data, now you would like to see that what is your state level performance or region level performance. From region level you come down to state level. From state level you come down to district level. So it is about, the original group that you are having, that you will be slowly breaking down and going deeper into it. So this process is called drilling down. This, then the next category is

(Refer Slide Time 30:54)



actually your slicing and dicing. In case of a multi-dimensional view of the data, if you would like to remove some of the dimension, remove one of the dimensions it is called slicing. If you remove multiple dimensions it is called dicing. From data cube, imagine a data cube, let's us you are getting data in three-dimensions and you have a data cube. We were talking about let's say time is one direction, cells is one direction and area is in another direction. So in this three-level database, if one of the, one of the part or a dimension is removed it is called slicing. If more than one such things are removed and you get a smaller data cube out of a bigger cube, it is called dicing.

So with this,

(Refer Slide Time 31:56)

we finish our discussion on the data resources of an organization. In fact, in this series we started with the description of the infrastructure, basic infrastructure in a E- Business system and we saw that E-Business system has four basic infrastructure requirements. First one is hardware, second one is software, third one is networking, network and fourth one is data. So in this series we have now finished with all these four basic resources. Form next class, next class onwards we will be talking on the security infrastructure which is fundamental to the existence of E-Business system, thank you very much.