Systems Analysis and Design
Prof. V Rajaraman
Department of Super Computer Education & Research
Indian Institute of Science, Bangalore

Lecture - 22

Last time, we were looking at data input methods. And I actually ended up with how to detect errors in a code. And the modulus 11 in check digit is one which is normally used. And we looked at a theory of, why it is modulus 11. And we also saw that if it is an alphanumeric code that is both letters and digits. Then, it is modulus 37, system which you use.

That is depending upon the total number of symbols like if it is hexadecimal. Next add prime 17 and you use 17 as the modulus number. So, the method was to actually take add digit in such a way that, when you multiply the rest of the digits by distinct weights 2345 and so on. And divide by 11, you will get remainder. And that remainder is the one which is used as the added digit.

And later on if there is any error there is either single transcription error or single transposition error. Then, that will be detected and we also said that 95 percent of the cases. It is either single transcription error or single transposition error. And those which are left out also almost about 90 percent of what is left out are detected. Even though, you cannot guarantee detection. And so, there is a very simple method which is used in many situations particularly.

If you look at a practical example, all books will have what is known as a ISBN printed in the copyright page. ISBN stands for International standard book number. And that has got actually as a code which has got total of 10 digits and it uses the modulus 11digit as the last digit.

You can just check it any book and see if it satisfies modulus 11 check system or not. The ISBN is gained some kind of a block code; expect if it is using characters for the detection of for easily explaining. Now, it is meaningful than a normal code with characters in it.

But, it is in certain types of fields like there is a first 3 digits are used for country code. And then the next 3 digits are used for the publisher within the country. And the last 4 digits for the serial number of a book published by that publisher. So, then one more digit is there as a check digit though it is actually formulated. And you could just check it up and there is an interesting example of a standard code used all over the world.

For book numbers that is giving unique number for every book published in the world. That is what it is supposed to be. And those who publishers, who register for that system automatically get all data except serial number which he adds as in when he comes up with a new book published then adds it goes to the next serial number and last 4 digits were there for the serial number.

Now, apart from the error detection, when we are also pointed out, when there is a large volume of data as input. Special precautions are needed to validate the data. Because, if the number is extremely large like lakhs like in CET exam any other board examination and so on. There is extremely that is extremely difficult to ensure that the data entry

operator will not make any error. There is every possibility of the data entry operator to make an error.

Sometimes, it is also possible for a data entry operator to be mischievous. In other words he may change some marks intentionally. And one should essentially guard against such fraud. And they are also very essential in any kind of a system to guard against fraud for instance if any changes 59 marks into 95 marks. Or then of course; obviously, it is intentionally done.

Of course it could be done accidentally, but whether it is accidental or intentional. It is the duty of the checking program which checks the input data that such errors are detected and corrected. Before, it is stored or further process by the machine. So, there are number of checks normally used to make sure that in any batch of data. We enter normally manually in such large systems there are checks and balances you might say.

So validation checks, one is to put sequence number for each record. So that you would count the total number of records and you do not error input or records together. So, what you normally do is give a sequence or serial number for one starting from one to the end which may go up to something like more than a lakh. But, when you give it to data entry operators, there will be more than one operator.

To whom you would have given that divide up the job and give it to number of operators they all will work in parallel. So, the work will be done faster. So, you make batches or something like 50. Arrange by batches. So, serial number 1 to 50 you give to operator number 1. 51 to 99 or 100 you give to another operator and 151 to 100 and 101 to 150 and 151 to 200. That is the way you divide up into batches of 50, 50 say.

And for each batches batch see the advantage of a serial numbering is that. When a batch is entered the program also counts. The total number of entries made and checks again the total number which you also manually write. That in another words apart from the data itself. There is one more data item called a batch control data which occurs at the batch of 50. So, in the batch control you put the total number of items in that batch. If there are 50 enter 50 and there are some other checks you do.

For instance manually add all the marks in a column. And when you add all the marks in a column and put that as a new number. And when the data entry operator enters, then

the computer will actually add all marks in a particular column. And check it against the total which you have put in the box control. So that, if there is any kind of an error or any mischief it will be brought out by the control total.

So, the batch control totals adds the serial total number of records in that particular batch. And also certain checks like say certain fields are taken. And the total number of items in that field is added and given for instance, in examinations where ABCD are given in subjects. So, what you could do is, at the end you can say total number of As is so much. Total number of Bs is so much, Cs is so much, Ds is so much and Fs are so much.

And then, you can actually count it by a program and count against the manual number which is written by the person who creates the batch control. Because, batch control one should make sure that the batch control is not tampered with anyway. And if there is any problem one should also be able to track down the from the batch control total that which particular batch error occurred.

And then you can pinpoint the particular operator who entered it, so that if there is any suspected fraud you can investigate it. This is called an audit trail that is you got a method of trailing or who did this mistake whether this is intentional or is accidental. (Refer Slide Time: 03:30) And say batch control total is absolutely essential. It is always used in any large data entry to have smaller batches and have batch controls. And batch controls actually control line is manually calculated. And it is put there and sometimes that particular in order to eliminate frauds.

That particular total may not really be known to the operator who enters that data. So that somebody else can enter that and entered data you can check. And so that, the total can match. In other words the point, I am really making is if suppose operator knows that the controlled total is being taken on a particular say field say for instance physics marks. And there is one 95 and one 59 and he may intentionally interchange these two marks in which case the total will remain the same and no error will be detected.

So in other words, the person who is entering and the program which is checking are the same. In other words, the person who is entering knows that this total is being done the total is being checked. So, it does not provide security. So, what you should really do, is that the total should be separately calculated and kept with you with the supervisor. And

after the data entry operator enters all the 50. Then, you check it against the control total by a program. And if there is any kind of a discrepancy, then you can track down.

In fact, either manually or by means of a program, where the error occurred and if you suspect fraud of course, you can have an audit trail that is who entered this. And that is the way, in which it is normally carried out. Because, it is the entry operator does not know which particular marks column is being added. Because, it could be physics chemistry mathematics and he may randomly decide that one batch who add physics marks and other batch you add chemistry marks and so on.

So, in other words if randomly it is done. Then, the operator does not really know which particular field. You have added and he may even not be told at all about the control total being kept there secretly without his knowledge. So that, any such fraud can be eliminated. So that is important. And there is another data entry and verification which is very expensive, but in some very important situations, where the any single error in a data entry could become fatal to the application. You in abundant caution you do this.

In other words, you have two different people enter the same data. And; that means, is duplicated and you actually do a verification of the entire data by means of a computer program. In the computer program, because two independent people have entered the same data and a person does not know, who is the other person who has entered it. There cannot be any collision.

So, any kind of a fraud will easily come out, if verification is done by complete duplication. But of course, it is expensive, because it doubles the work. The same work is being done by two people. The other kind of a check is called record two totals. The batch total is for taking items in a particular field of a record and adding all of them for the particular batch.

(Refer Slide Time: 03:30) Record total may be the total of certain random fields in the record added and put at the end of the record. And that can be separate control, which can be checked after the data is entered. So, total is record tells you identifies which particular record the total does not match. And so, you can actually track back and get to that record where does not match.

And modulus 11 check digit is for key fields. Key fields if any error occurs modulus 11 check detects them guarantees detection 95 percent not so guaranteed in for more percent. Almost 99 percent of the cases, the error come out in the modulus 11 check. There is also other method, which are used earlier. But, they are not as good as modulus 11. One method to just add the digits of a code and take modulo 10 and that is very easy to do; that means, all the weights are equal.

And then, you just do an addition and modulo 10 you put the number and that can never detect single transposition errors. It in some cases, it will detect the single transcription error. So normally, the transcription error normally detects, because the sum will be different, because different number is used. But, if suppose two different digits are the same.

And then, both of them are changed there of course, it will not detect. So, these are problems with modulus 10. But it is a very easy method. But nowadays, it is not very commonly used. Modulus 11 has got popular.

(Refer Slide Time: 17:14)

Now, there are checks also on individual fields in a record which you normally carried out. One is called radix error. For instance, suppose you are entering minutes and seconds hours minutes and seconds. Then, radix cannot exceed 24 if you have an 24 hour clock minutes cannot exceed 60 seconds cannot 60.

So you can actually check, your incorrect entries made. Then immediately, it will throw it out say in that radix error. Similarly, for instance date for instance there is a date field also say kind of a radix error. In other words, the number of months in a year cannot exceed 12. So, it got to be number less than 12. The number of days in a month cannot exceed 31. So that is a error which you can find out.

So, it is one of the possibilities of detecting errors in dates. Dates and times are simple examples of what are known as radix error correction. Range check. Range check fairly straightforward. Normally if, you are entering marks, then the marks will normally range between 0 and 100. If it is a 10 is the total marks in that particular subject.100 is the maximum marks.

Normally marks cannot be negative and marks cannot be greater than 100. And so, if there is a proper range check, then any incorrect entry of a number above 100 would be detected and thrown out and that is called range check. And they are very easy to kind of include any program, while the data is being entered. So, the program as soon as the data is entered. A little program kind of say the data and takes out the obvious errors like radix error or range check.

There is some also called reasonableness check. Reasonableness check for instance is a little more complicated. Suppose a person is being getting telephones which are normally say a 1000 rupees a month. Suddenly one-month, he gets a bill for 10000 rupees. It is; obviously, a possibility of mistake. So, you go back and check that. Do not send the bill of 10000 rupees to the person immediately.

There will be a furor or in other words there are in fact those items in news papers which you would see that somebody got a telephone bill for 2 lakh rupees suddenly. And then, it becomes a news item and it is a kind of reflects very badly on the billing program of the telephone company. Saying that person is normally paying 1000 or 2000 rupees a month ends up with something like a lakh.

You expect there is some obvious error is there or if it is not. Then, you should actually go back and check why it came 1 lakh. So, these are important for the supervisor after data entry to check these things. A program can check it right away. A program can check it by looking at the average bill. And the bill of this month and if it is unreasonably large it can just throw out that record for manual inspection by the supervisor.

In case of data like for instance 31 April, because April has only 30 days, so it cannot be 31 or 30th February can be 28 or 29 and then not 30. Most date checking programs are built in and because it is used so often that you have little function to check the date. And in fact, there is one exercise which is always given in computer programming courses including the leap year checking and so on.

Incorrect data is found out by batch total the batch control total. Missing data also batch control, because the serial total number of items in the batch is counted. And if it is a missing record then the count will show you out of 50, there is only 49 or if it is one more extra 51 instead of 50 then; that means, some incorrect data is sneaked into it. So, then you recheck it. And something called inter field relationship check that is also interesting.

In inter field relationship check, there are couple of examples I will give. One example for instance would be that. For suppose a student takes about 5 subjects in an examination. And he gets in physics 99 marks, but in mathematics he gets 10 marks. Then, there is a inter field when you compare these two. Then, you do not expect this student who gets 99 marks in physics to get 10 marks in mathematics. So; obviously, there could be an error either in data entry or in marking or whatever.

So, this kind of an error should also be thrown out at the time of data entry. So that detail check can be made. And not send the mark sheets the poor student and then ask him to go from pillar to post and pay money for rechecking and stuff like that. It reflects very badly on the data processing system and the contract who was employed by the university or the school board whatever it is to not check these things and put the students to a lot of inconvenience.
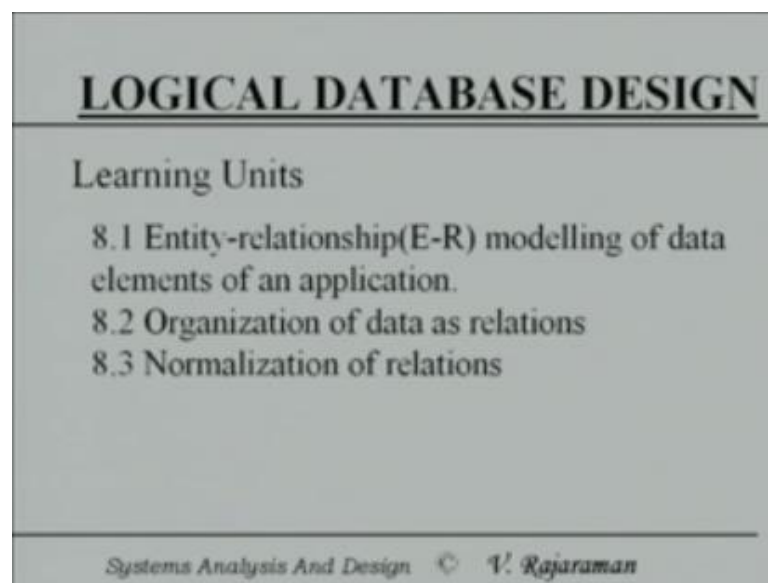
As a responsible system analyst programmer it is your ethical duty to make sure that these stupid silly errors do not occur. And if it occurs it reflects very badly on your professional competence. And it seems to be occurring again and again and again in our country as you always hear about so many students having to go back to do a reevaluation or rechecking and total checking and stuffs like that. That should not really be happening.

But of course, sometimes it will be real the student did not do too well most often. In case like 99 percent in physics and 10 marks in mathematics most often, it is an obvious

mistake either in data entry may be he got 100 in mathematics. And the person entered only 10 and; obviously, this is a very bad thing to happen. And similarly another example, is if you have a neighbor of a student in the class in which he studies. Normally a student in 8th class cannot have an age greater than 25 having exceptional cases may be there.

But, if it turns out then you have to do a manual check. So in this particular module, we looked at primarily the various methods of data input. And we also looked at the importance of the correctness of inputing data. And methods of detecting errors in data entry, because in any large volume data entry errors do happen and it is the duty of the systems analyst to make sure that at the time of system design. He or she puts in controls appropriate controls to eliminate such situations.

(Refer Slide Time: 26:11)

LOGICAL DATABASE DESIGN

Learning Units

8.1 Entity-relationship(E-R) modelling of data elements of an application.
8.2 Organization of data as relations
8.3 Normalization of relations

Systems Analysis And Design © *V. Rajaraman*

Now, we go on to the next module which is one logical data base design. The reason why I am going there now is. So far we have looked at data flow diagrams, we have looked at how to enter data. Once a data is entered the data is got to be stored the reason, why I am going there now is. So far, we have looked at data flow diagrams. We have looked at how to enter data. Once a data is entered, the data has got to be stored in some data base.

See the, what goes into the data base has to be clean. In other words, if there is any error in the data which is entered in the data base later on you will get into difficulty in terms of various processes. So, before anything gets into the master data base. It is essential to

make sure that the data is cleaned up. And cleaned data you might say laundry must do a laundering of the data before the data gets entered into the data base.
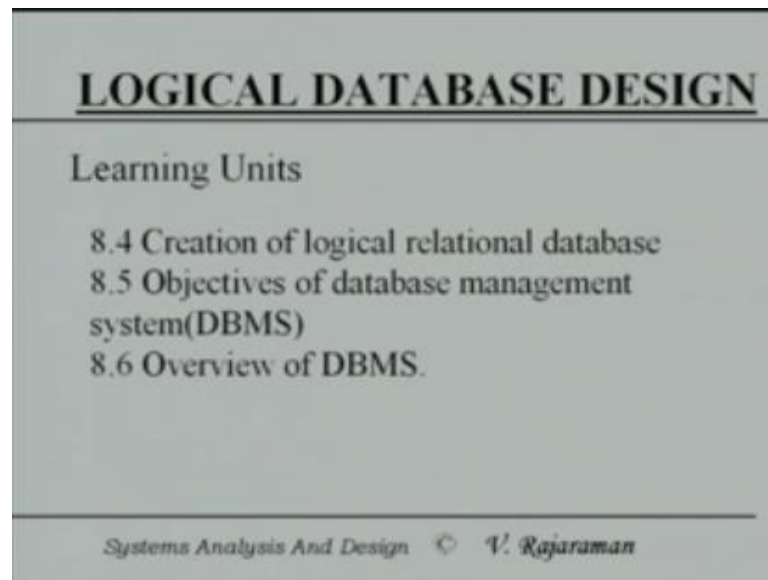
So, the data base organization assumes that the data going to be entered is correct. In terms of the there is no errors in data entry instead of like that. And further processing can proceed. And primarily it is the module; we look at what is commonly known as relational data base system. Relational data base system is now very commonly used. In fact, it is might say ((Refer Time: 27:40)) standard used by every organization.

I will explain, what is a relation and what is called a relational data base. But before, we start or create a relational data base we add a kind of model the data. Something called a entity relationship modeling data which will load about and how to organize data in relations and normalization of relations. Normalization, I will explain in great length as we go along creation of logical relational database.

Logical data base is something which is very very distinct from a physical database and logical database. Logical database is something which you create to understand how the data will be organized. A data has to be ultimately stored in a physical device like a disc. And there is something called a database management system which transforms logical database and stores it as a physical database.

And while storing as a physical database it should make sure that the physical database is uses the best possible economy. In other words make sure that optimization is done in the storage. So that retrieval will be fast and things like that.

(Refer Slide Time: 29:45)

LOGICAL DATABASE DESIGN

Learning Units

8.4 Creation of logical relational database
8.5 Objectives of database management system(DBMS)
8.6 Overview of DBMS.

Systems Analysis And Design © *V. Rajaraman*

And we will not get into great details of DBMS database management system. The reason, we need not go into great detail in this particular module about database. Relational database management systems are because in any computer science. Curriculum is a full course on database management systems which will be compulsory in almost all universities.

Because, it is a pillar or a corner stone in the learning of computer science engineering for database management system is a very very important course and because, the whole semester is devoted to database management system in this series of lectures. I cannot really afford to it is not a purpose to be able to cover the entire subject. Purpose here is to give you a flavor of what database management what is relational database, why it is important, how it kind of gain idea of what DBMS is.

So, it form a proper groundwork for you to study further. For this you might say the foundation on which you can build later on your other course.
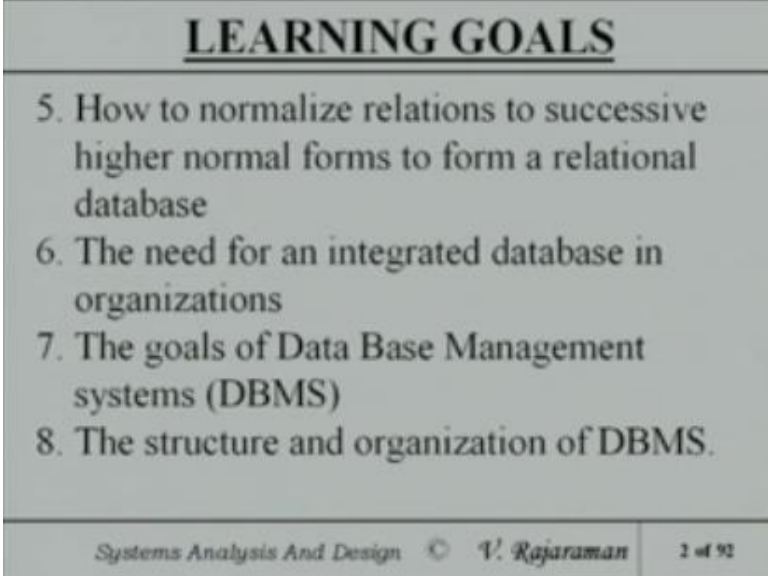
(Refer Slide Time: 30:46)

LEARNING GOALS

In this module we will learn:

1. The Entity-Relationship(ER) modelling to develop a conceptual model of data.
2. How to organize data required in an application as relations
3. The need for normalizing relations
4. The various normal forms and their relevance

Systems Analysis And Design © *V. Rajaraman* | 1 of 92

So this, module we will learn entity-relationship modeling to develop a conceptual model of data, how to organize data required in an application into relations and need for normalizing relations. I will tell because more often need not students do not understand, what is there are number of issues one is what is normalization, why do you normalize a database, how do you do it.

Most often students know mechanically, how to do it, but they do not understand why normalization is required. So, one of my major objectives is to essentially talk about first what is normalization, why is it required and how it is done. There are a number of normal forms and each one is relevant in certain situations. We will say which form is relevant in which situations.

(Refer Slide Time: 3:45)

LEARNING GOALS

5. How to normalize relations to successive higher normal forms to form a relational database
6. The need for an integrated database in organizations
7. The goals of Data Base Management systems (DBMS)
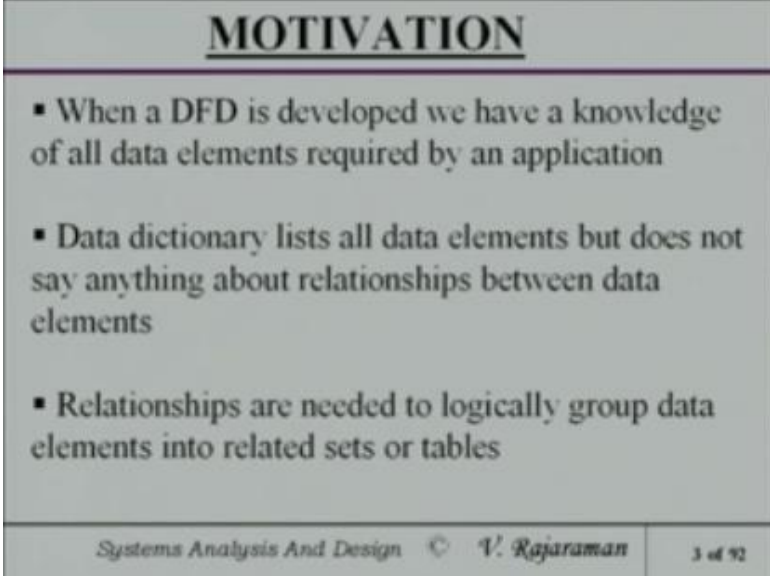8. The structure and organization of DBMS.

Systems Analysis And Design   © V. Rajaraman   2 of 92

So, how to normalize relations into successive higher normal forms, there are a number of normal forms called first normal form, second normal form, third normal form, fourth normal form fifth normal form and so on. So in fact, every few years somebody comes up with a N plus 1 normal form.

But very often in practice, people are quite satisfied with going at third normal form. Because, for all practical purposes three normal forms are sufficient beyond that in pathological cases you go. So, I will emphasize more up to the third normal form in this lecture. The need for entity database in organizations that is an important question which will be raised, because relation relations we are components of a database system.

A database will consist of many relations and all these are will be managed by DBMS structure and organization of DBMS.

So, when a DFD is developed. We have knowledge of all data elements required by an application and data dictionary lists all data elements, but does not say anything about relationships between data elements. In other words only the data DFD at the end of it you come up with a data dictionary. Data dictionary is more to really has a documentation aid for a programmer who is going to maintain the program to understand the meaning of a data, where it occurs and so on.

So, the maintenance can be easy and it lists everything. Data dictionary is actually a Meta data. Data about data and logically organizing groups of data elements relationships between data elements are have to be known. These relationships are not really available in the data dictionary. It is only a listing of all data.

(Refer Slide Time: 34:50)

MOTIVATION

- Such an organization

  - Reduces data duplication
  - Simplifies adding, deleting and updating data
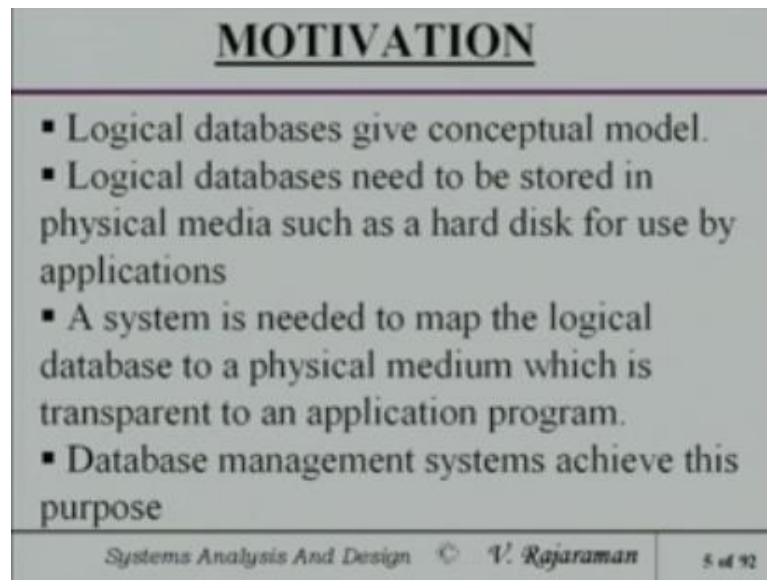 - Simplifies retrieval of desired data

So, actually the last point namely relationships are needed, what is a relationship? And that is what is important that is relationship. And relational method of organizing reduces data duplication. Simplifies, adding deleting and updating data. Simplifies, retrieval of desired data, in other words, why should we do a relational operation? That is a motivation.

So, if you do a relational database organization it reduces data duplication. Because, if you duplicate data it is unnecessarily storage is wasted. And also there are duplicates there is possibility of so called duplicates can be inconsistent. So, one should not have inconsistency and you should be able to add delete update data without any errors. In other words if you delete something you may unknowingly delete some important information about some other item that is what is eliminated.

Similarly, when you add something it might be inconsistent, what it already exists. And when you update you may make a mistake in update. And all these things are really what are detected or ensures that a relationship data base a system ensures that these things would not really happen it simplifies and also simplifies retrieval of desired data.

So, logical database give conceptual model. Logical databases need to be stored in physical media as hard disk, I told it already. System is needed to map the logical database in a physical medium which is transparent to our application programmer. In other words application programmer should not be worried about how the relations are actually mapped into hard disk. Because, hard disk as you know are organized into number of cylinders, number of sectors and so on.

So in sector organization, the cylinder organization, so on should not be of any concern to the programmer, how it is mapped. In case of the job of DBM as far as the programmer is concerned. He looks at the data as a logical data for all the programs he writes. So, this is what is important. That is what as far as the program is concerned. The actual physical organization should be invisible unknown to him. DBMS database management systems achieve this purpose.

(Refer Slide Time: 37:24)

So, the purpose is to develop a conceptual model of data. This model specifies relationships among data items. Using this raw data is organized into tables of related data. So, a table of related data is really what is called a relation.

(Refer Slide Time: 37:47)

In fact, I am being very loose in my definition here, but I think it is more stricter definition will be given as you proceed. The tables are organized in such a way that duplication is reduced. Operations of adding deleting changing data is simplified and

systematized. Of course, systemization reduces accidental errors. Retrieval of data is facilitated just repeating what I said.

(Refer Slide Time: 38:24)

LOGICAL DATABASE DESIGN-INTRODUCTION

- Collection of these tables are called the database for the application

- Loosely one may call organization of related data put in a table as a "RELATION"

8.1.2   Systems Analysis And Design   ©   V. Rajaraman   8 of 92

So, collection of these tables is called a database for the application. Loosely, one may call organization of related data I just want to put into a table as a relation. The reason I am calling it loosely is because there are lots of other. There are huge theories or relations, which is there in the literature.

There is a person called as Codd who first invented this idea of relational database. And if you add what is known as a relational calculus which could be applied to this. And in some sense it is somewhat like the relations in mathematics relations functions and so on. A very specific meaning in mathematics, similarly he used very specific mathematical definitions and which are not as loose as I am making out to be.

They are lot more stricter and they are the ones theoretical foundation on which entire relational database idea is based. So, the reason why it is so popular is because it is sound theory of relations. Underlying the concept and which ensures all these ideas of simplification. Non duplication elimination of inconsistency all those things are part of that mathematical structure which has been made up by the person who proposed relation database idea.

As usual some of the ideas when they are first proposed are not immediately accepted. Because, it turned out when it was first proposed it was a little profligate or in other words did not economically use storage. Because, it is actually proposed in 60's at that time disk storage was a premium. The largest disks available in 60's were something like 20 MB.

Today of course, no sophisticated person will have pc less than 80 GB. So, storage has become very large and at very low cost. So, the so called uneconomical method of storing is no more relevant. So that is why when computers improve slowly people realized the importance or ease of using relational database. And so, other methods which are prevalent earlier were all kind of abandoned. And relational database what is currently being used as the standard almost everybody.

(Refer Slide Time: 41:33)

LOGICAL DATABASE DESIGN-INTRODUCTION

- Systematization by which related data are put in a table is called "NORMALIZATION"

- A method called entity-relationship analysis facilitates creation of relations

8.1.2   Systems Analysis And Design   ©   V. Rajaraman   9 of 92

Systemization by which related data are put in a table is called normalization. A method called entity relationship analysis facilitates creation of relations. So, how do you create relations, we use a modeling tool called entity relationship method. And again this is proposed by a person called Chem. again some years ago. It is very simple idea and the idea of entity. This again and again occurs in computer science. Later on the same idea of entity will occur in object oriented design which will come later in this course.

ENTITY-RELATIONSHIP MODEL

ENTITY: Specifies distinct real world items in an application

For example:
vendor,item,student,course,teachers

8.1.3 Systems Analysis And Design © V. Rajaraman 10 of 92

So, the question is we have to define, what is an entity? We have to define, what is a relationship? Entity specifies a real world item in application. In other words it is something which exists as either animate or unanimated. Normally, some kind of a noun phrase, examples are like vendor is a entity, an item is an entity, a student is an entity, a course is an entity, a teacher is teachers are teacher is an entity.

So, there are entities normally the entities are the ones. There are many many entities in particular relation and those entities are related.

ENTITY-RELATIONSHIP MODEL

Relationship: meaningful dependencies between entities

For example: vendor supplies items
teacher teaches courses

Relationships are underlined above

8.1.3 Systems Analysis And Design © V. Rajaraman 11 of 92

And we are talking about relationships. Relationship is a meaningful dependency between entities. For example, a vendor supplies items. So, vendor and items are entities, whereas supplies is a relationship. In other it connects the two entities namely vendors and items. Through the relations applies.

One way, we were looking at it is vendor and items are nouns and supply is a verb. And normally relationships are verb classes and entities are noun classes. So, teacher teaches courses. So, teaches is a relationship. And the two entities which are related namely teacher and courses is by the process of teaching. So here, you are underlying the relations.

(Refer Slide Time: 45:20)

Entity set is a collection of similar entities. There is suppose you have a collection of records or for instance when you take a beautiful diagram vendor you take the vendor as a key item, vendor codes say. But, there are many vendors with different codes. All vendors who are supplying items to the organization vendor set, where the elements are entities. Individual vendors are you might say members of the entity set. And there are in a set the member will all have a common property.

A set is consists of a number of similar items. So, a set of chairs a set of tables similarly set of vendors. Set of all items in a store is an item set. Normally entities are members of a set. Like a teacher is a member of all teachers in the university or college. Every

member of an entity set is described by certain attributes. There is every, you take a vendor the vendor will have certain attributes.

(Refer Slide Time: 45:58)

ATTRIBUTES

Attributes specify properties of members of entity set. Attributes also specify properties of relationships

Examples:

Entity : Vendor

Attributes : **vendor code**, vendor name, address

Relationship : supplies

Attributes : **vendor code, item code**, order no., qty. supplied, date of supply, price/unit

Like for instance vendor will have a vendor code, because I have said you have a unique identification of any particular entity. And I have put in bold phase like the key field words say or the unique identification of that particular entity. Because, the entity has got the attributes namely the vendor code, vendor name and address. They are the three important attributes of a vendor.

And his name and address and supplies will have two it relates two different entities. So, the attributes will have both the entities which take part in the relationship between vendor and item. So, there is a vendor code or item code both given to indicate the relationship supplies actually connects these two. And the attributes are you have ordered certain number of items from a vendor.

And normally, it is an order number which again order number is an identification of the order which will be unique identification. So, order number quantity supplied date of supply price per unit these are attributes of a relationship.

ENTITES AND ATTRIBUTES

Example

Entity : Teacher

Attributes : **Teacher code**,teacher name,department,building,room no,phone no.

Relationship : Teaches

Attributes : **Teacher code,Course no**,course name,semester offered, credits, prerequisites

8.1.6   *Systems Analysis And Design*   ©   *V. Rajaraman*   14 of 92

Some more examples teacher, teacher is an entity teacher code teacher name, department, building room, number phone number may be home address whatever. See you may really put all important information about a teacher in that attributes of a teacher. And you will put that information such as address, if it is important in the application. If the address is irrelevant like in the case of one number is given we assume it as a office phone number.

Also could be home phone number. If it is a home phone number then the residence number and office number both may be given. So, it depends the attributes which are important are decided by the application to which you are going to put this entity. If you are going to be sending say some letters to teacher at his residential address. Then, the residential address becomes a important attribute of teacher.

Similarly, if you want to be able to phone the teacher at home outside hours then the phone number becomes important. The point I am trying to make is there is no hard and fast rule about what attributes report. You will put the number of attribute depending upon the situation. On the, what application or what future application, you are going to use this for this entity for.

Because, as I have been pointing out continuously in this course. Software systems, how do we design for expandability and flexibility? Expandability is an important part because no software system is ever static. So, even today field like phone number may

not be required. Tomorrow some policy of the institute or university may change, where home phone number may be required.

You do not want to go back to the drawing board and start redesigning your database at that time. So, it is a kind of better to look into ahead and see what are all possible uses may be put in the near future at least. And in this, you can get some help from the user the person who is going to who is actually asked to design. But, as I said a good analyst or a good company will teach the analyst to kind of put some of these things. It may be considered redundant today, but it may be a great help tomorrow.

So, one never knows better to kind of look at the possibility of expansion. So, relationship is a teacher teaches. See teacher is an entity. Teaches is a relationship. So, if you want to have the attributes of the relationship that is the teacher code and course number. Course numbers are also unique in any university. The course name is there, but course number is the one which is very commonly used in most all universities at least in abroad and definitely in all IITs and NITs and so on in India.

I do not really know enough about whether all colleges use. It but it is a good idea to use a course, course number or a code for a course. And course name and semester in which you are offered. If you were offered in both semesters you say both 1 and 2 or if other term I say one 2 3. Number of credits for that course or in the case of colleges may be maximum marks for the course prerequisites.

In order to take the course, what other course must have been cleared. Like it could be other course numbers which are prerequisites before you take these other courses you cannot take this course. So, the prerequisites are more important part.

So, some entities depend on one another. The in fact, we saw this. Here teacher and course number are two entities which depend on one another. So, they would depend on one another. Entity vendor and entity items and related as vendors supplies items. These relationships are described by entity relationship diagrams. So, it is a graphical method as we have been always finding out a picture is worth 100 words.
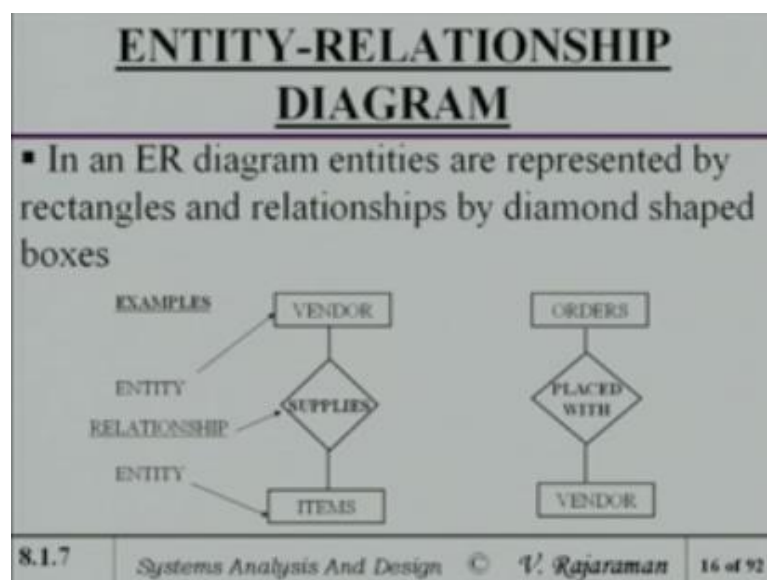
So, a picture pictorial representation is a very easy way of documentation. Good way of documentation and also gives you an idea about what attributes are being used at

vendors, what are the entities which are being used, what are the relationships which are being used. And also later on we will see that what are the attributes of an entity.

ER diagram the normal what I would say the notation which is used commonly used to have a rectangle to show an entity. And a diamond shape box to show a relationship. So, vendor and within that particular box you write the name of the entity and name of the relationship. So, vendor supplies items. So, entity is vendor item is an entity and the. So, that is what the entity is a vendor, the entity another entity is items and the relationship is suppliers.

Similarly, orders is an entity vendor is another entity and placed with orders are placed with vendor. This is another example of a entity relationship model.

(Refer Slide Time:54:35)
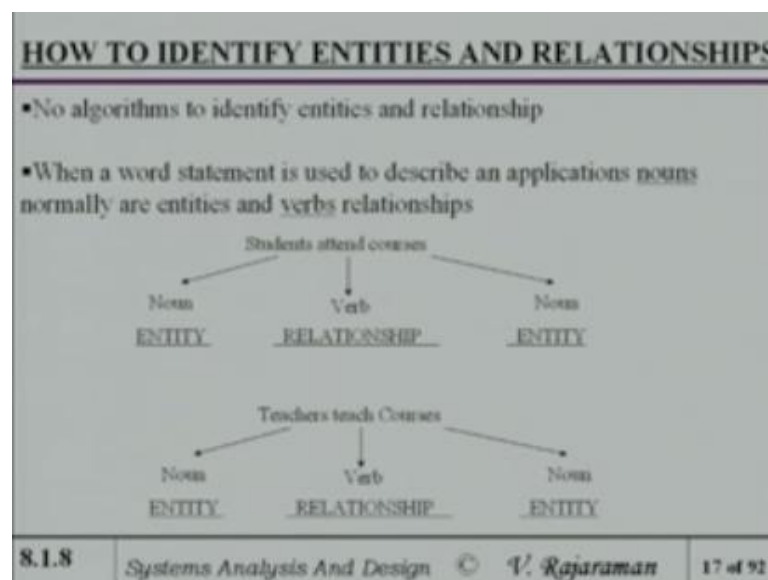
There is no algorithm to identity entities and relationships. In other words program, we can write to detect automatically what are the entities and what are the relationships; obviously because, after checked they are obtained from you might say requirement specification which are normally stated in plain English or whatever language.

And so, you should be able to interpret that language and design the particular identify what are all the entities and what are all the relationships. So, when a word statement is used to describe an application. Nouns are normally the entities and verbs are

relationships. I have told that earlier today I am repeating it. The nouns are normally entities and verbs are relationships.

Like for instance a teacher teaches courses, students attend courses. Student is an entity. Attend is a verb in that. It is a relationship. And courses is the noun which is a another entity. Teachers teaches courses, students attend courses. So, these are two simple examples of how to detect the noun class.

Noun class, there are two noun classes in this case you might say. And then, there is a verb which connects these things. One can of course, pick points in my grammar. But, the important point is that the connection is through a verb. And we are normally known as noun phrases.

(Refer Slide Time: 56:41)
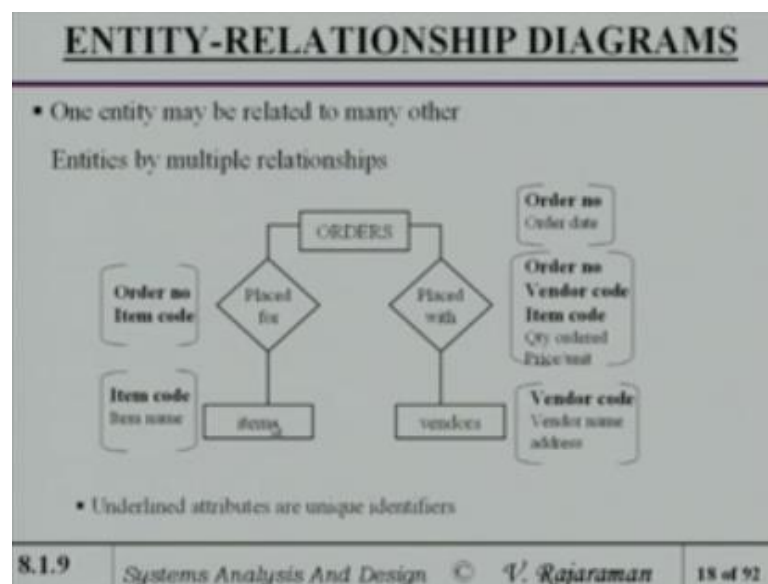
One entity may be related to many other entities by multiple relationships like orders are placed for items. So, order can be order number, order date and vendor. Vendor is vendor code name and vendor address. See, normally in the entity relationship diagram, you put the entity down. And on the side you write down the attributes of the entity. In this case I have just given two parts.

Just two kind of to economize space, an order entity may not just have just two parts because there is a key. Key is essential. The other things may be there many things may be there. I have showed only two here. And placed for item code item name is placed

under items and order number and item code these two are the primary keys of this and this.

So, I am putting it, but later on we will kind of little bit of you might say refining of this. In the sense that only in certain cases, you require both of them as what is known as composite key? Two keys put together as so called composite key. Similarly in this case, I used a composite key of order number vendor code and item code. And in fact, it looks a little odd because the item does not occur either here either unless order number also has got item code in it.

So, item code appearing here is not actually the reason it is appearing is that. In this entity attribute item code must have been there. I have illustrated in the point that just do not blindly put something in the relationship which does not occur in the attributes of the two entities. Normally, the relationship will have these two keys and also mostly something which has like for instance quantity added.

See order number also will have quantity ordered in it. And price per unit also normally will be there. So, normally in the relationship also would really have this, what is there in? In the entities, so because if you illustrate that do not put blindly. You can look at it somewhat carefully and do the proper job of writing it out. Of course, put it as something like what I have done here is not all that it occupies space. Better to put it in a row this is what I will do later on in this in this module.

So, I think we will continue from this point onwards and look at in detail in the next lecture.