Carbon Accounting and Sustainable Designs in Product Lifecycle Management

Prof. Deepu Philip
Department of Management Sciences

Dr. Amandeep Singh Oberoi
Imagineering Laboratory

Dr. Prabal Pratap Singh
Department of Management Sciences

Indian Institute of Technology, Kanpur

Week 10

Lecture 45

Database Design

Hello everyone, welcome to the course on Carbon Accounting and Sustainable Designs in Product Lifecycle Management. I am Dr. Prabal Pratap Singh and we are co-teaching this course with Prof. Deepu Philip and Dr. Amandeep Singh. Till now, we have covered an introduction to database management systems in our last lecture. And now we are trying to dive deeper into the designing aspects of these databases.

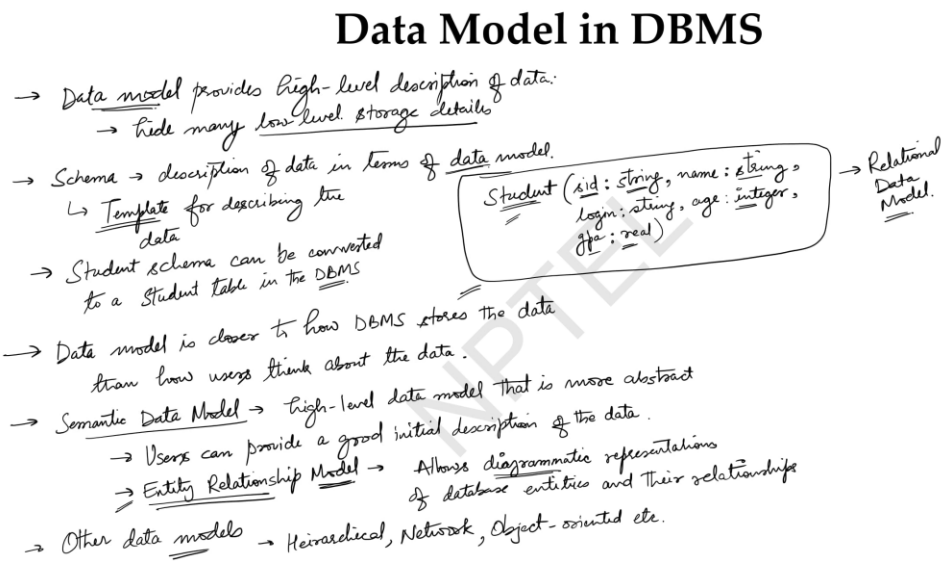
Outline

- Data Model in DBMS → what is it?
- Database Design Process → six steps
- Fundamental Terminologies →
- Diagrammatic Representation → ER diagram
- Relationship sets →
- Features of ER Model →
- Conceptual Design Decision

So, today we will start by learning what is a data model. So, this would be, what is this data model? What is it? Then we will try to understand the complete database design process. So it is divided into six steps and we will try to grasp a brief introduction to each step and then we will also see how these steps are implemented by a designer. Further we will understand the fundamental terminologies that are involved in the database design process next we will also see a diagrammatic representation that is the entity relationship (ER) diagram.

So we will take a real world scenario and try to convert it into an ER diagram. In the next section, we will try to understand other concepts of the database design like what is the relationship between two entity or entity sets like these things. The next thing is the features of ER model. So, what are the different things like whether a particular entity set is a weak entity or not, we will try to understand how to identify these things. Finally we will look further into the conceptual design decisions like if you are designing your own database.

Then what are the major distinguishing decisions that needs to be made so that your final design of your ER diagram can become efficient or better.



So let us start by understanding what is a data model. So it is nothing but a high level description of the data. Data model provides high level description of data. Now we

already talked about how things are stored in a database and how they are different by storing it in a simple file of operating system.

So these data these DBMS does not store file in a similar way as that usual files are stored. So therefore what a DBMS system does is they divide the data into different aspects and then try to create a different kinds of schema so that these become efficient while storing and they can provide all the features of the DBMS technology. These data models helps them to hide many low level storage details.

So these low level storage details mean that the actual storage of your data is not provided to the user or the database designer. However this data model will only provide how you are representing the data not how you are storing the data right so schema is another important term which is useful in these data models and it is a description of data in terms of data model.

So, let us first see what an example of a schema so that we can easily understand how we are going to design our database and then how we are going to create a schema out of it. So let us say we are talking about a student in a school and how you are going and we are creating a database for the school management system. So a student is an entity, the basic entity of this complete system. So an example of a schema for student is student ID which can be represented using SID. And this SID will be a string so these are all the design decisions that have already been made and we are just showing you an example of this schema.

The next thing the student can have is a name, so name will also have a data type of string similarly he or she can have a login username. So it will also be a string age of the student it will be an integer and finally the GPA the grade. So this will be in decimal, so it will be stored as a real number in the database, now this is an overview of an example of a student entity in a database schema. So this is describing the data in terms of a data model now which kind of data model we are using here this is a relational data model. Because relational data models are widely used and we are studying the relational data model only.

So, schema is a template for describing the data. Now why we are calling it as a template because we are not mentioning a particular student here while showing you this schema. We are not saying that this is ABC student whose SID is a string or name is this, login is something else. We are just telling that any student of the school will have these basic attributes in it. So this is a template.

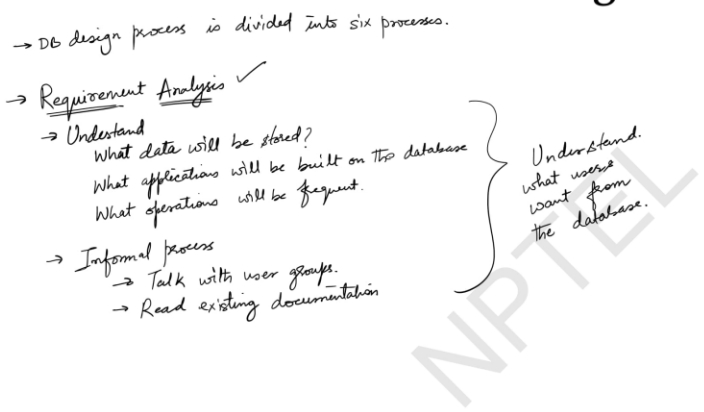
Now based on this template, we can fill our data. And the data of student will not go beyond this definition, right. So, after designing this schema, the data model will provide that student schema. This student schema can be converted to a student table in the DBMS. So, the first stage is to create a schema and then create this table actual table in the DBMS and then fill the data there, right, for all the students of a particular class or all the students in the school.

So, now the thing is that this data model is a representation of the data. What we are doing here in the schema phase is we are trying to represent the data how it will be saved. But we are not telling that how this data is actually saved on the physical disk. So, therefore, data model is closer to how DBMS stores the data. Then how users think about the data.

So in the database design process, we will learn how we try to capture the real world in the process and then using by talking with the users of the database. And then we will try to convert this input from the users into data model. And that is how data model will try to make your data more towards the DBMS systems than from the user's input, right. So, now these data models are also of different kind. So we will be trying to use a semantic data model.

So what is a semantic data model? It is a high level data model that is more abstract. So with these kinds of data models, the users can specify their requirements and then we will try to convert it into the tables of a DBMS. So users can provide a good initial description of the data. Now, in these semantic data models one example of these kinds of model is an entity relationship model.

So this is the most widely used model for semantic data models that database designers use. So what it does is it allows diagrammatic representations of database entities and their relationships. Today we will try to see how these diagrams will be made and how to convert these diagrams into schema will be our next task. So, this is not the only model the other data models could be hierarchical or network data model, object oriented etc. So, we are using the relational data model and in those relational data models, we will try to use entity relationship model to create an overview of the diagrammatic representation of these data models.



Let us now start the database design process. it is just these database design process is just one part of the complete system software system design process and while designing a database. The actual data intensive applications will require a better diagrammatic representation and clear overview of the complete real-world system. So this complete process this DB design process is divided into six processes. So, let us start with the requirement analysis.

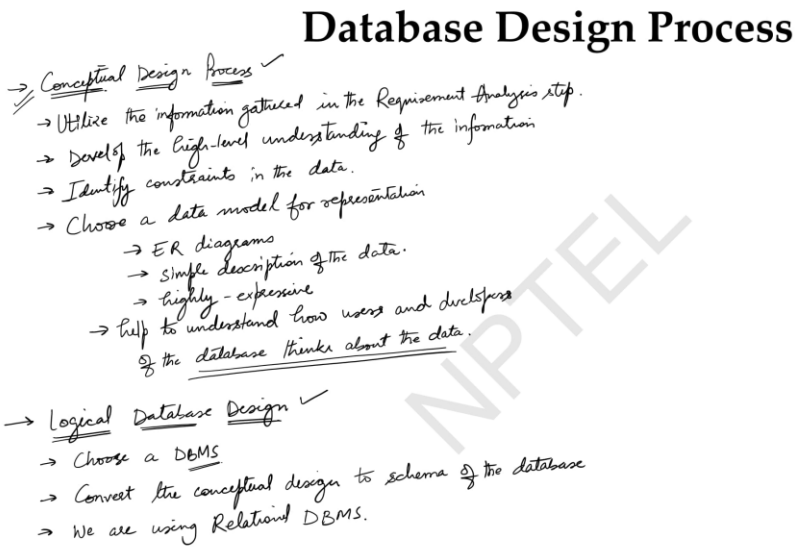
The first step of this database design process is requirement analysis. So let us say a school management has asked you to create a new database for their school that will encompass their complete student systems, their faculty information and everything else that will help them to manage their daily activities better.

Now, as a database designer, your first step is to first identify the actual persons that are handling these kinds of information while managing the school. So, you will try to talk with them and find out what are the different data and their data types that are needs to be stored. So, all of these kinds of activities where you try to have discussions or have focused talk groups with the management of the school will come under the requirement analysis step.

So in this step we majorly understand the problem now what we need to understand what data will be stored in the database. Or what applications will be built on the database. Or

what operations will be frequent while using this database. So, overall this is an informal process. Where you will talk with user groups or read existing documentation if it is available.

Overall this is understand requirement analysis you will understand what database users want from the database, right.



The next step is the actual conceptual design process. Now under this second step of the database design process, what we will do is we will try to utilize all the informations that we gathered in the requirement analysis step in the previous step. And we will try to create different kinds of database schemas out of it, so and also we will try to identify the different types of constraints that the identified data has. So we will utilize the information gathered in the requirement analysis step.

The next thing is as we talked that we need to develop the high level understanding of the information. Which we need to store in the database, right. So this information is going to be stored in the database the other thing is we will try to identify different types of constraints identify constraints in the data. So, let us say in a school management system, if a faculty cannot have a limited number of hours in a particular week, right. So, we need to identify these kinds of constraints that whether while assigning the number of hours a faculty can perform in a particular week.

Whether it is unconstrained or constrained or whether the age of a particular student is having some kinds of constraints for studying in a particular class or not. So, these kinds of constraints will be identified using the data we have collected during the requirement analysis process. The next thing is we need to choose a data model for representation, which we have already chosen. Choose a data model for representation. So, this is we are going to do this with our ER diagrams.

And it is one of the several high level model, so what it will do it will provide a simple description of the data. And there are different kinds of data models that can be used but er diagram is widely used because they are highly expressive. So when we are going to make these ER diagrams you will understand that why they are highly expressive. And that is why they are widely used. And these kinds of diagrams will help to understand how users of the database and the developers of the database thinks about the data.

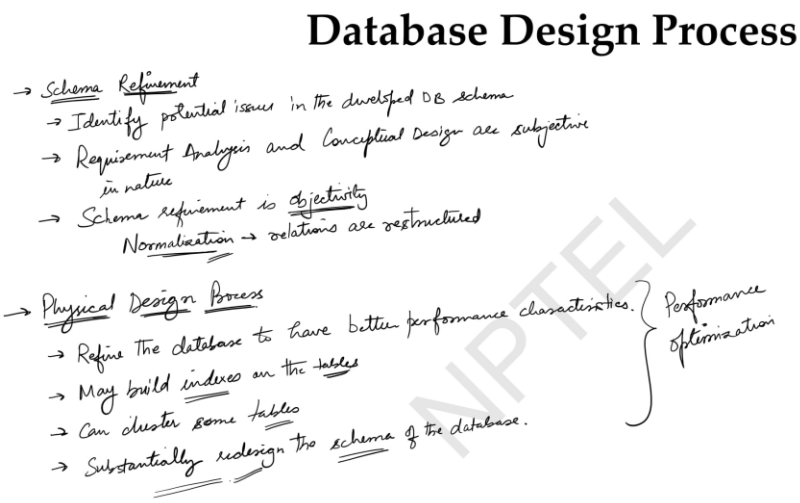
So, this is this is an important step while creating your conceptual design process because the user may have a different idea about the reality. Whereas the developers who have just done the requirement analysis may not have identified or captured this reality in its true sense. An example could be if you are designing a database for the employee management system in a particular organization, then the actual manager of a department.

And a simple employee can be a manager of more than one department but maybe it is restricted by the organization that a single employee can only be in a manager of the single department. So these kinds of constraints needs to be identified by the developers and needs to verify with the users or the employee management system of the organization.

So that they can create a efficient database with less number of faults in it. So this conceptual design process is important because we need to identify this reality and try to convert it into our database schemas. So the next process in the database design is the logical database design. Now until the conceptual design process, we have a diagrammatic representation of the complete system. And now we need to convert our conceptual design into a schema so that we can use our database management system.

So first we need to choose DBMS. So most probably this could be anything like MariaDB, MySQL, Postgres. So we will be using MariaDB in this course. So after choosing our DBMS, we need to convert the conceptual design to schema of the database. And here we have already mentioned that we are using relational DBMS.

Now, we have done three major parts of our database design process that is requirement analysis, then the conceptual design process and the logical database design. Now, developer will have a database that can be used on the system, but we need to further refine the complete system.



So, the next stage is schema refinement. So we have developed our schema. Now we need to refine. How?

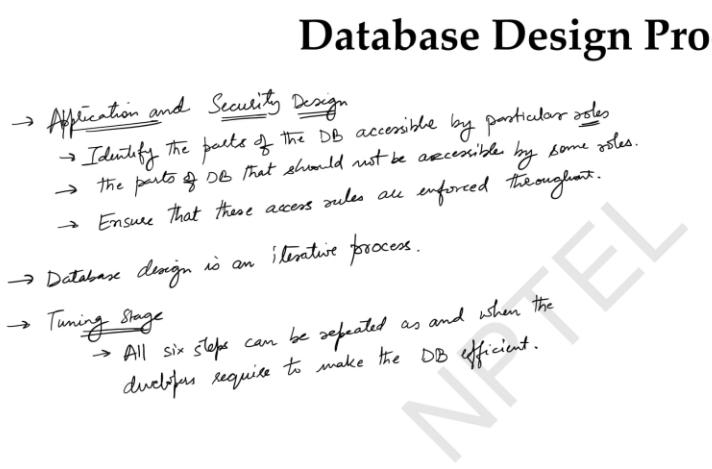
We will use different kinds of theories that already exist in the literature so that we can identify potential problems in the database schema and try to mathematically resolve those schema issues. So until now, requirement analysis and conceptual design process are subjective in nature. But this schema refinement process is actually objective. In this design process we will first identify potential issues in the developed DV schema and as mentioned this requirement, analysis and conceptual design are subjective in nature but schema refinement is objective.

How it is objective we have completely defined mathematical process that is called normalization. So in this process we will try to restructure the relations relations are restructured. And this complete process will try to reduce the redundancy of the data that is stored so the next stage is the physical design process. So we have our refined schema

with us. Now the next task is to how these database will be stored on the physical disk, so we will try to refine the database to have better performance characteristics.

This stage is mostly optimization aspects of the database where the how these actual data is stored on the disk are improved so developers may build indexes on the tables. So these are relational tables that are developed by using the database schema. And now developers will build these indexes so that they can be efficiently utilized and the retrieval or modification process of the data can be improved. They can also cluster some tables. So that the number of tables will get reduced and the performance may get improved by clustering these tables.

And sometimes what happens is at this stage the developers may find a need to substantially redesign the schema of the database. So this is the actual implementation phase. And if the developers think that few aspects, if by changing few aspects of the schema, if the performance can be improved during the heavier workloads of the database. Then they try to redesign the schema by going back into the previous steps of the design process, right. So this complete process is majorly talking about the performance optimization.



Now, the last sixth stage is the application and security design. At this stage, the developers need to understand that any data of an organization, whether you are

developing for an employee or the organization for the employee database or you are doing it for a school management system, The data has different kinds of aspects. That is, maybe some data is crucial for the privacy reasons, some data is only useful for a particular role in an organization. So, these kinds of SS possibilities by the users of the database needs to be determined by the developers while designing the database.

After creating the complete database. Some users will have access to some parts of the database and some users will not have the access. So both of these aspects are important. We need to show what we need to show and we need to hide what we need to hide. So these kinds of roles needs to be defined clearly and the security of the data in the database management system should be clearly worked upon.

So at this stage we identify the parts of the DB database accessible by particular roles. So roles are nothing but different kinds of users of the database. We also need to identify the parts of the parts of DB that should not be accessible by some rule. And finally the developers need to ensure that these SS rules are enforced throughout. So, this six step process is now complete and although it is complete, it is actually an iterative process.

So, like we saw in the physical design process that developers are trying to substantially redesign the scheme of the database and the schema was actually designed during the conceptual design process. From the fifth step, we are going again toward the second step. So this way, this complete design process is an iterative process and the developers need to iterate over different stages multiple times during the development. So database design is an iterative process. So there is another stage that we can say as a tuning stage and at this stage it is nothing but when we are trying to repeat these steps so all six steps can be repeated as.

And when the developers require to make the DB efficient. So, this is the complete design process.

Thank you.