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 44

Database Management Systems

Hello everyone. Welcome again 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 Professor Deepu Philip and Dr. Amandeep Singh.

Till now we have discussed various kinds of carbon accounting databases that are available in different countries and particularly in India. We have also talked about the roles of these carbon accounting databases and why they are important and why all the organizations are using it.

We will first introduce with the database management systems, what are they, what is the difference between database and a database management system. And then we will start using a particular kind of database management system that is relational database management system and we will start coding.

Outline

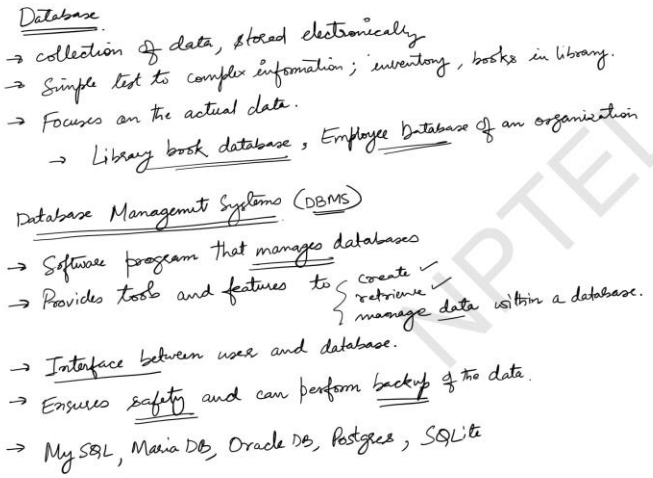
- What is a Database Management System (DBMS)?
- Evolution of DBMS
- File storage vs. DBMS
- Advantages of DBMS

So the outline for today is the first we will talk about what is a database management system. Then we will talk about how these technologies evolved. So we will talk since 1960 and till today.

After that we will also have a brief discussion about why we are not storing all these information in files simple files on our operating systems and then we will dive deeper into the advantages of these database management systems. So let us first talk about what is a database management system.

So, usually we have used two terms in two terms till now that is a database and a database management system. So, database is a passive data whereas database management system is a complete system that is managing all the aspects of this data like how to store this data, how to retrieve this data all of these things.

DBMS?



So, let us first understand or refresh what is a database. We already talked about database which is a collection of data stored electronically. Now anything that you are storing electronically that is storing it as in a database and it can be anything from simple list to complex information. Example could be inventory of a particular product, how it is changing, how many are available or the books in a library. And who have taken which book in the past or who is currently holding a particular copy of the book. So these are all the use cases of our database.

So a database actually focuses on the actual data that we are trying to store for our information retrieval or information modification purposes focuses on the actual data. So example could be library book database or employee database of an organization but then database management system evolved. Now why do we need a DBMS. This is known as DBMS in short. So we will use this term from now.

So why we are using DBMS and what is the difference between a database and a DBMS. So DBMS is nothing but a software program that manages databases. So this is important why because you can have different kinds of databases available with you but a software program that is managing these databases is known as a DBMS database management system. So while managing what they are doing they provide this software program provides tools and features, to create retrieve or manage or modify data within a database

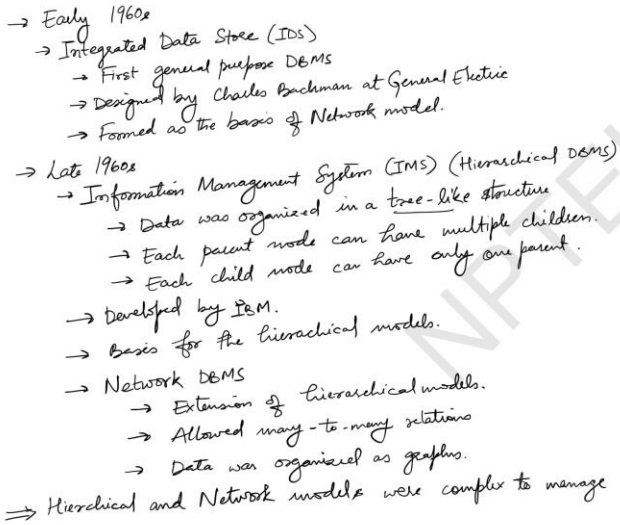
and again it is this software program is just an interface between user and the actual database user and database.

So this to either create something new in your database or retrieve existing data from the database or to manage or modify it. We need a software program and this software program not only provides these capabilities, these three capabilities, it also provides an interface to conduct these operations, right. So, it also ensures safety and can perform backup of the data. So in a state of system crash, these kinds of features are also required by a user so that DBMS can perform these operations and his or her data is safe with it in a state of system crash, right.

So we know that we can create database, but what are the different kinds of DBMS. So a DBMS, famous DBMS are MySQL, MariaDB, Oracle DB, Postgres, Postgres SQL, so we also call it as Postgres, or a simple SQLite. This is the light version of SQL database. Now we understand what is the difference between a database and a database management system. So now let us talk about how these technologies are evolving.

Because this requirement of using a management system to manage your data is not new. So and with time the computing power and computing technologies have emerged. So, we should understand first how these technologies are emerging and what are the new things that they are adding. So, majorly we will talk beyond 1960s. So, pre-1960s we are majorly using file storage systems and after 1960s the work on DBMS started, actively started to improve these technologies.

DBMS Evolution



In 1960s, the DBMS systems were storing the data in a tree-like structure and these systems are known as hierarchical data models. Also, the next change occurred with the network models.

In 1970s, the relational database model was introduced by a researcher. Further, in 1980s, object-oriented database models were explained and these models were actually complex, but they provided a new technology to manage these databases. In 1990, data warehouses came into picture, which were collecting the data from different kinds of databases under a single database management system.

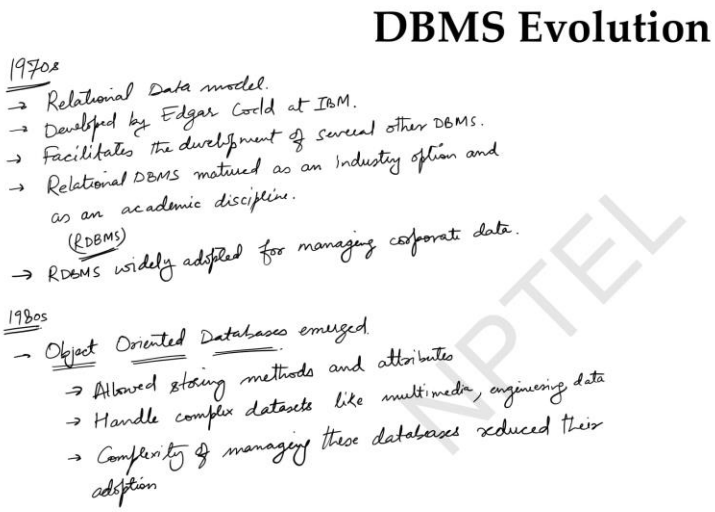
Next comes the internet age, where websites were very prevalent and in the initial stages, what people were doing, they were using file storages for their data on their websites. Then these databases were optimized to use this data in a database using DBMS for their own websites. So websites are now extensively started using these DBMS technology. In current age, we can now manage complex and unstructured data like multimedia, image, everything we can store in a DBMS system. So, this is how they are evolving.

So, let us now understand how these things are evolving since 1960. So, in the early 1960, we had integrated So, this was the first general purpose DBMS. So, we talked about various kinds of features that this software program provides. Now, this integrated data store was the first version of a general purpose database management system that provides most of the features that we are talking about in early 1960s.

This was designed by Charles Beckman. At GE (General Electric) and this data store IDS this IDS was formed as the basis of network model. These DBMS works on different kinds of database models. So, initially we had this hierarchical database management systems and now this network model was also developed based on these systems. So, in late 1960s what happened?

IMS (Information Management Systems) was developed which was a hierarchical now data was organized in a tree like structure, so in a tree like structure there are parent nodes and child nodes so each parent node can have multiple children However, each child node can have only one parent. So, many to many relations were not available at that time and this was developed by IBM.

Developed by IBM and these systems form the basis for the hierarchical models also network based models network dbms was also introduced in this era and they were nothing but an extension of hierarchical models and they allowed many to many relations further data was organized. Since it is a network model data was organized as graphs. But the issue in this era was that these hierarchical and network models were complex to manage.



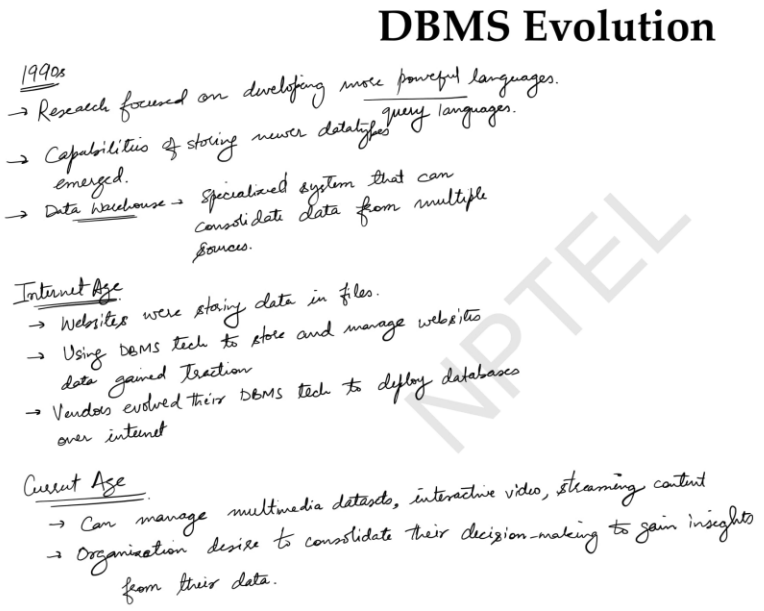
So there were different efforts that were being done to mitigate this problem of these database management systems and in the 1970s relational data model was introduced. And this is the most prevalent data model until now. And we will also learn when we talk about MariaDB, it is also an RDBMS, which is based on this relational data model only. So this model was developed by Edgar Codd. And it was developed at IBM.

It facilitates the development of several other DBMS also in coming decades. Also this relational DBMS. Started maturing, matured as an industry option and as an academic discipline. So, what happened was different kinds of organizations were actively started using this RDBMS and they also started developing it further, this technology further. Also, academic research also started.

So, this DBMS, RDBMS was widely used. For managing corporate data so next in 1980s these RDBMS was established as a dominant. DBMS paradigm but another thing that

happened was object oriented databases emerged. So we usually talk about object oriented programming language and other things. So here we are talking about object oriented databases that means the data and the related attributes are stored in a database as an object.

So these database allowed storing methods and attributes the feature of this kind of databases, they can handle complex datasets like multimedia engineering data but the issue with these kind of databases was, that the complexity of managing these databases reduced their adoption by industry.



So coming to the next decade in 1990s research focused on developing more powerful. Languages for what languages for querying the data or editing the data in the database. So these kinds of languages are known as query languages and new capabilities emerged capabilities of storing newer data types emerged further a new thing known as data warehouse. This was introduced in this decade and what it is it is a specialized systems. That can consolidate data from multiple sources. So in the last lecture of the carbon accounting databases, we already talked that there are different kinds of data sources in a particular organization while doing the carbon accounting process.

So these kinds of technologies like data warehousing where the data sets are coming from different sources can be consolidated is a very effective feature for our carbon accounting

process as well. So, these kinds of technologies are currently available in the different kinds of DBMS. So, in the next decade, also you can say as internet age. Now, during this time, the websites were very prevalent. So, websites were storing data in simple files.

So using these DBMS technology to store and manage websites data gained traction So what happened was initially there were only static websites where whatever you write, only you can see. And you cannot dynamically change that data on the website. But if you use a database and you have different kinds of data inside the database, so this DBMS technology can facilitate the dynamic content on the website.

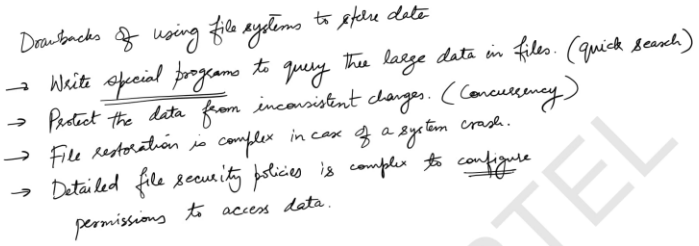
So these technologies developed using this DBMS technology. So, DBMS vendors evolved their DBMS tech to deploy databases over internet. Current age, right now we can manage as we have already said can manage multimedia data sets interactive video streaming content. All of these in a database further organizations desire to consolidate their decision making as well. To gain insights from their data.

So, this last point actually we have talked during the carbon accounting process that some of the prevalent databases provide analytic capabilities. So, in the current age these databases are capable of performing simple to complex analytics operation on the data sets available inside their database. So this software program, we are terming it as DBMS, is having different sub-programs that can perform these analytics operations. So this capability is currently available in these systems. Now, let us talk about the fundamental question, why we are not using FISE instead of using this complex system known as DBMS.

So consider an e-commerce website. Any e-commerce website like amazon.com or flipkart.com and they have they provide different kinds of product informations. They have different product categories and all these informations are stored in a database. Let's say we start storing these informations in operating system files. So what happen is.

There are different kinds of problems that may occur. Like at a particular instant, there are multiple transactions that are happening. So, and if we are storing all these transactions in a particular file, then the particular file is opened for editing by different. So, this kind of multiple operations at a particular time may not be suitable for the simple operating system files that we are storing. So, these kinds of features are available in DBMS.

Why not Files?



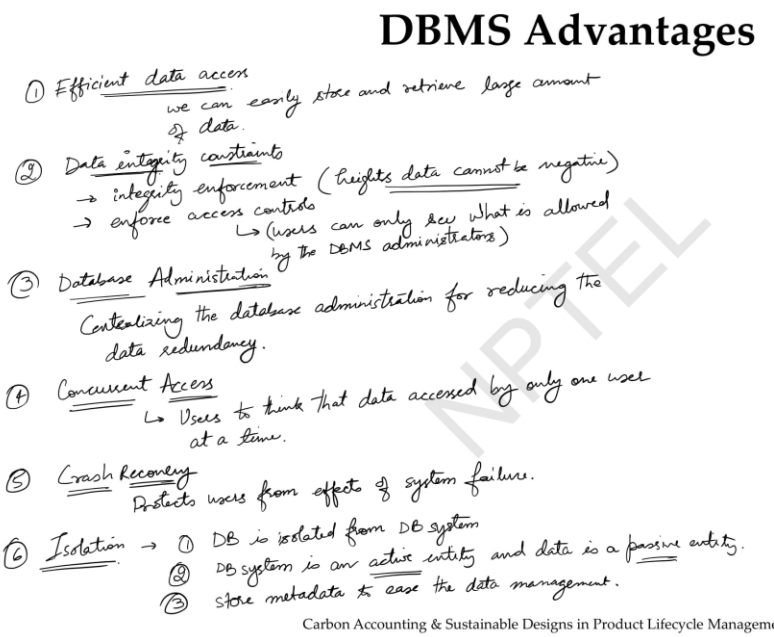
So, let us now understand the drawbacks of using or storing the data into file systems of using file systems to store data first of all we need to write special programs to query, the large data in simple files. So these spatial programs will easily get complex while writing to query the large data in simple files, right. So however if you use the e-commerce websites that we were discussing.

They provide very quick search. So the quick search capabilities on these websites is a feature of the databases that they are using databases or the database management systems that they are using. Also they protect the data from inconsistent changes. So with simple files, the possibility of having an inconsistent change is far more than having it in a database. So these databases provide a feature known as concurrency.

This feature allows every single user that is actively using a database to modify the database concurrently. However, you cannot perform this operation with single files. So the feature of concurrency is missing when if we use simple files for storing this large amount of datasets. Further file restoration is complex in case of a system crash. So the database management systems are having specialized programs already written as sub-programs which provide this capability of easy restoration of datasets in case of a system crash.

So these kinds of capabilities are not available in operating system files. Last, we can also say that the detailed file security policies is very complex to configure. Now what we need to configure we need to configure permissions to SS data. So, as I discussed while discussing about the evolution of these DBMS, that the corporate system actively taken up this DBMS to store their data. Now, in a corporate organization, there are different kinds of users.

Some are sales managers, some are having other kind of roles. So, each role has their own permissions of files that they can use or edit. So, if we just allow storing all the data in operating system files, then what happens is we have to develop the security policies that will be very cumbersome and complex with simple files. However, the DBMS technology provides these security policies in a very efficient manner.



So let us talk about the advantages of DBMS. The first one is efficient data access. So in this what it says is we can easily store and retrieve large amount of data second is data integrity constraints So this feature allows integrity enforcement. What is this?

That let us say you are storing the heights in a data set. So the heights data cannot be negative. So, the DBMS technology will enforce this rule that in a particular attribute the data will not have negative values. So, this is called data set integrity enforcement and it

can also enforce SS controls. It means that user's can only see what is allowed by the DBMS administrators.

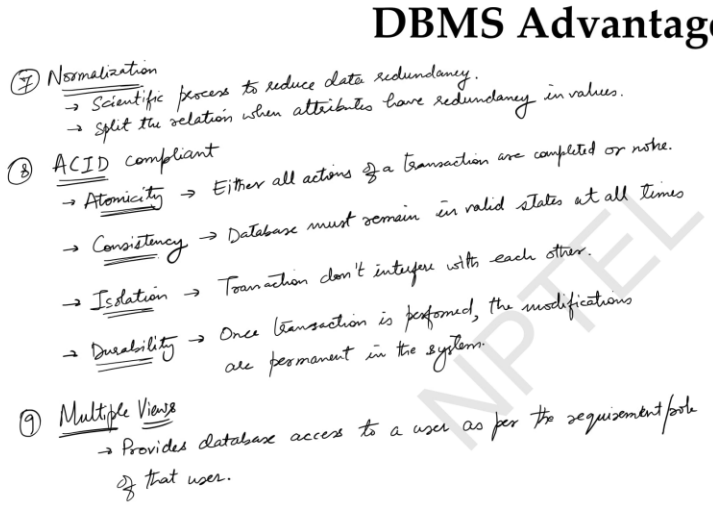
The third feature or the advantage of these DBMS technology is database administration. So instead of administering in different stages the we are centralizing the database administration, for reducing the data redundancy the next advantage is concurrent. So this feature allows users to think that data is being assessed by only one user at a time. However, multiple users are assessing the data, but still the DBMS technology makes one user feel that he or she is only using the database at a particular time.

No one else is using it. So they can freely edit, retrieve or manage the data sets. Another feature is crash recovery. So most of these advantages are also why we are not using the simple files. So in crash recovery, DBMS protects users from effects of system failure.

Another important feature is isolation. So, here the database or DB is isolated from the actual DB system that is DBMS. So what this feature is saying is that the actual data that we are storing is again isolated from the applications that this DBMS system provides. So DB system is an active entity. And data is a passive entity further this DBMS also stores metadata to ease the data management

So metadata is nothing but the data about data. Like if you are clicking a picture, a photo, then the actual photo is your data. But when that photo is clicked or created in your mobile or in your camera and what are the specific dimensions of that photo or other related things, these are all known as metadata. Now this metadata is very useful in a DBMS system. So, even if you are managing your data sets in your database, then when you have deleted or extracted or retrieved anything from your database, that can also be stored in your DBMS, which is not available, may not be available to you.

But the actual DBMS, that is the management system behind all this, will use this information to manage the actual data in your database. So, this feature is provided by isolation.



The next major feature or the advantage is the normalization. So, normalization is nothing but a scientific process to reduce data redundancy. We will talk in detail about this normalization process when we will design our own database.

So this feature also split the relation when attributes have redundancy values. Now, this will be explained better when we will design our own RDBMS and there are different kinds of normalizations that can be performed 1, 2, 3. So, these will come in the further lectures. The next important property of these DBMS system is they are ACID compliant. So ACID here is an acronym where A stands for Atomicity, C stands for Consistency, I for Isolation which we talked and D stands for Durability.

So what is atomicity? So let's say you are having processing system of a bank and they are performing transaction processing system. So if you are conducting a transaction and during the midway the transaction fails. So if the amount is deducted from your bank account then the whole process of deduction will be rolled back. So that means that if the process occurs completely then only the transaction will be.

Otherwise the transaction will roll back to its initial state. So this feature is provided by databases and termed as atomicity. So it says either all actions of a transaction are completed or none. Now, what is consistency? We talked about this earlier also that let us

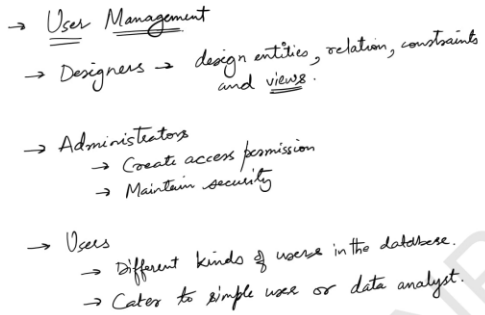
say if you are storing the heights of a particular community in a region, then the height should not be negative.

That means a user should not store a negative value in this attribute. So, that all the data set will remain consistent throughout. So, the database should. And if this constraint fails, then the database will move to an invalid state. So, database should always remain in a valid state.

Therefore, database must remain in valid state at all times. Isolation is nothing but transactions don't interfere with each other and durability is once a transaction is done transaction is performed. The modifications are permanent in the system. Another advantage is multiple views. This feature provides database access to a user as per the requirement or role of that user.

And this can be configured in the database management system. And it is usually performed by the database administrators.

DBMS Advantages



The last advantage is user management. So there are different kinds of users, mainly designers, designers of the database. So designers or administrators or the end users so designers design entities

So if we are talking about the database on libraries, so a book is an entity, student or the subscriber of that library is also an entity. So these kinds of entities are designed by the designers of the database and they design entities, relation between these entities, constraints and views. Views is the same thing that we just talked that what a particular user will see as per his role or requirement, right, administrator performs two major tasks they create ss permissions and they maintain security of the database At last users finally for what the end users there are different kinds of users in the database. So DBMS provides this capability that there could be different kinds of users in the database and these DBMS systems cater to.

Simple users are those that they are only let's say retrieving the data and performing simple tasks or those users that are performing specialized tasks like data analyst. So we have talked about different features of database management systems and their advantages and why we are not using files and how these technologies evolve with time. So now we will try to start designing these databases so that we can finally create our own carbon accounting database in this course.

Thank you.