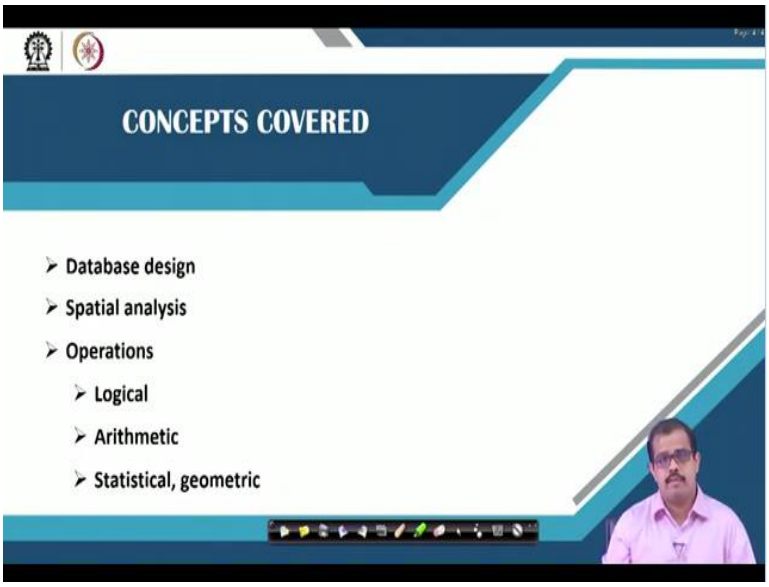
Geographic Information Systems
Prof. Bharath H Aithal
Ranbir and Chitra Gupta School of Infrastructure Design and Management
Indian Institute of Technology-Kharagpur

Module-08
Lecture-38
Spatial Analysis

Hello namaste welcome back to the course on geographic information system. In this module as I did inform you earlier, first thing is we looked at database as a whole than we looked at the spatial queries, how we can do cut spatial queries in this particular lecture we will get into more understanding about how do we do basic spatial analysis. So, next 2 lectures is completely dedicated to how do we do spatial analysis.

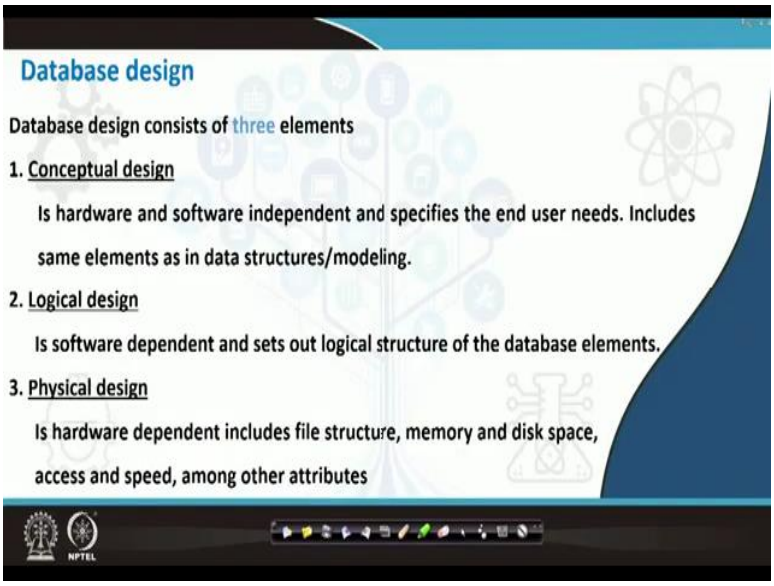
What are the different spatial analysis, the first this particular lecture is very basics of how a spatial analysis works. So let us look at this we look.

(Refer Slide Time: 01:04)



First we will start with the database design, okay, then we look at spatial analysis operations such as logical arithmetic and statistical and geometrical operations, I am repeating database design in because only to terms that everyone has to understand that how the design definitions of a database should be. So, this has to, if you understand this, then the rest will be very easy.

(Refer Slide Time: 01:26)



So, whenever you are looking at a database design, there are 3 elements that has to be looked at, first thing is in the conceptual stage, when I say conceptual stage, when you are actually looking at the entire module as 1 that that is where you start looking at conceptualizing how the database has to be worked out. That is first you have to look at what is a hardware, the software, independent and software that is independent of all the resources that you use, and that specifies the user needs.

So for example, if there is a database where there will be hundreds of users who is using it then if you can, if you create an old outdated hardware system, where you cannot handle even 100 or 200 users, so, that will crash down at every any point of time. So, you should have a very efficient hardware also a software that is customized and designed in such a way that it can handle any resources any point of time okay.

And specifies that and basically catering to every query of the end user. This includes same elements in the data structure and both in data modeling. So, you have to look at both data structure and data modeling. So, when you are actually conceptualizing any database designed, the first thing is how your data structure will be, what are the different data that fit in and what are the different models that use whether it is vector data model, raster data model, or you have a combination of both of these models.

Then you will look at that the conceptual design includes the software, so you need a software that is capable to handle both or you need a software that is capable to handle a single model, okay then what kind of I mean what kind of software that you need as a back end there are various other users. So you need a Python console to start programming in it, you need a different other consoles to look at different aspects of it.

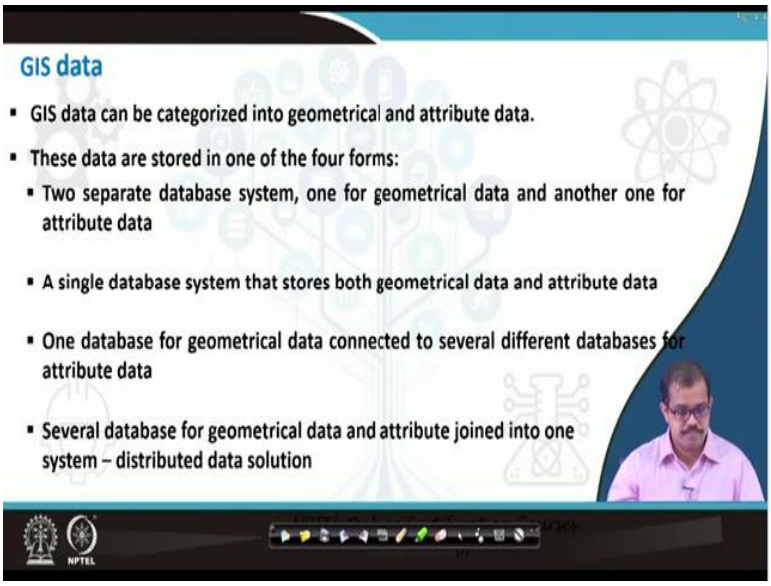
So, what basically you need. So, all of these things are taken into consideration that a software is decided, once the software is decided the same reciprocating hardware should be looked at to handle that software efficiently also to handle the data queries from the user very efficiently. The next thing is logical design. Is your software dependent. Is your design software dependent and sets out logical structure of the data database element.

So that is what we look at what kind of database element logically is this the way that your thing is that, it is more, more dependent on how software handles your data. So is more of most of the original design is based on a dependency on a software than a physical design, okay is hardware dependent include the file structure, okay, it is more of if you look at the hardware, what kind of file structure has to be maintained.

How much memory is required to maintain that file structure, what is a disk space that you may need, what is the access and the speed that you may need okay if you have a smaller database you may not need the such huge amount of both RAM and virtual memory but you need a high RAM, high virtual memory in order to run tasks which are quite intensive. So you need such qualifying hardware.

So, look at all of these aspects when you are designing it physically. So, 3 things that you have to look at one is conceptual design, one is the logical design other one is a physical design.

(Refer Slide Time: 04:56)



So once you have looked at it, then come to the GIS data. So, when we look at GIS data basically both GIS data has a geometrical data and the attribute data. Attribute data is a very compulsory part of GIS, because it gives you information about the geometries that is associated. So, now, these data are stored in 4 forms, whenever you are looking at GIS data, they are in 4 forms, 2 separate databases, 1 for a geometrical data other one for attribute data, okay.

So, when you start looking at the QGIS probably in the 10th and 11th module, we would give you an extensive hands on QGIS. So, that entire module will be handled by my Ts who will actually give you start with how do you create a data, how do you store a data, how attribute data manipulation happens. So, you during that module, when you save particular data, if it is a vector model, you are trying to save it and safe file.

So, you will find out that there are classes of files that has been stored, now it is not a single file. And mind you the most of the students make a mistake in this particular aspect. If I asked them you have to submit this particular record data file, what they do is they just pick out the share file dot shp extension and send it, they do not send all other files. So what happens if you just send an shp extension it means that it is an corrupted file.

It does not have it associated data files that has to be there, which has stored geometric data which has stored database file. So it will not be able to open. So if you are actually sending out

data, sharing our data, you have to save the data, you have to send out the entire set of files. It has 4 or 5 files that are together depending on the extension. For example, if normally if we take 2 separate database systems, one geometrical database, other one is for that database are created by default, okay.

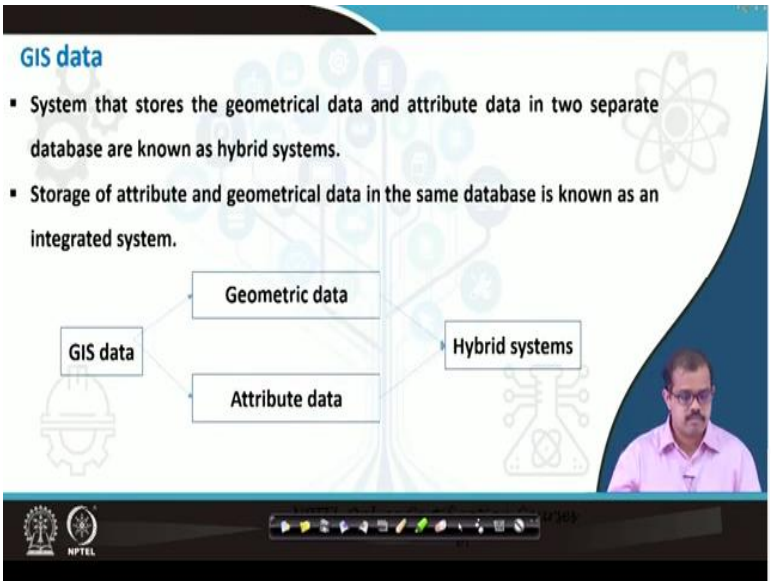
For example, in share file you have as a checks and you may have dot mdv or dot gdb So, that file is created. So, a single database system that stores both geometrical data and attribute data. So, that is the other one and one more database for geometrical data that is connected to several different databases and the attribute data for attribute data. So, that is also stored. So, you have 4 different files, okay.

So, these 4 different files for example, let us say someone sends only a single database system. So, it cannot sustain without other things, if someone sends only geometric data, so it cannot sustain with other database because you cannot query it basically cannot open. So you will have one file which has relationships only so it cannot open because it does not have geometric data or different databases.

So all these files have to be sent together, in order to handle that data if you want to share for example, you want to send out the data to someone. Maybe this has to be zipin, user WinZip to zip all these things and you can send it, okay. Are the best thing is converted to kmz file. So once you have converted to kmz file, you can probably send it. So though you have a lossy format, it is more of a lossy format, but you can always save in a kmz file, it is more of a zip file and it can be sent okay.

So several databases are in a vector data, okay, I am speaking about vector data basically, several databases for geometrical data and attribute joined into one system. These are distributed data solutions which have already spoken out in the previous session. So that is also another way of looking at it.

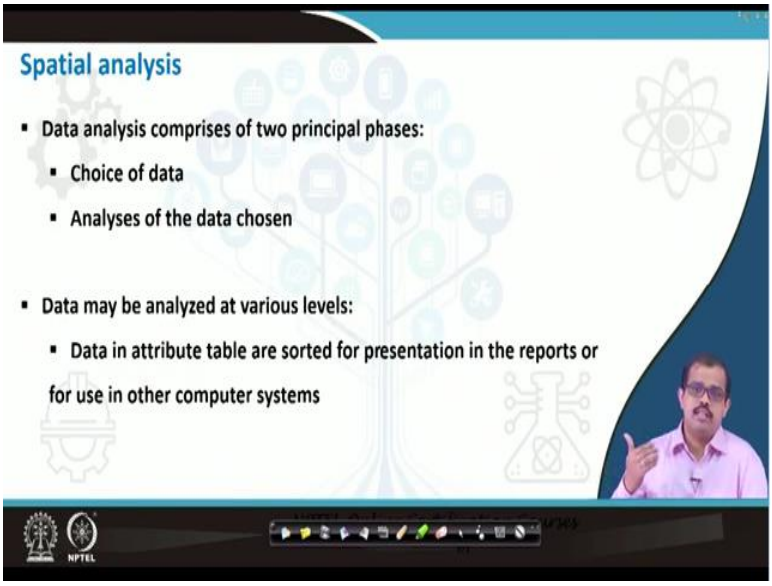
(Refer Slide Time: 08:50)



So you can send out the way you need. And when you look at GIS data, basically you have a geometrical data, you have an attribute data So, in 2 separate databases, so, these are called hybrid systems, whenever geometric data and attribute data has stored separately then it is called hybrid system and most of your file extensions today are the format file formats today have hybrid systems.

And storage of attribute geometry the data in same database is known as integrated systems. These are very less number of systems which are integrated, there are most of the systems that use are more often hybrid system okay.

(Refer Slide Time: 09:33)



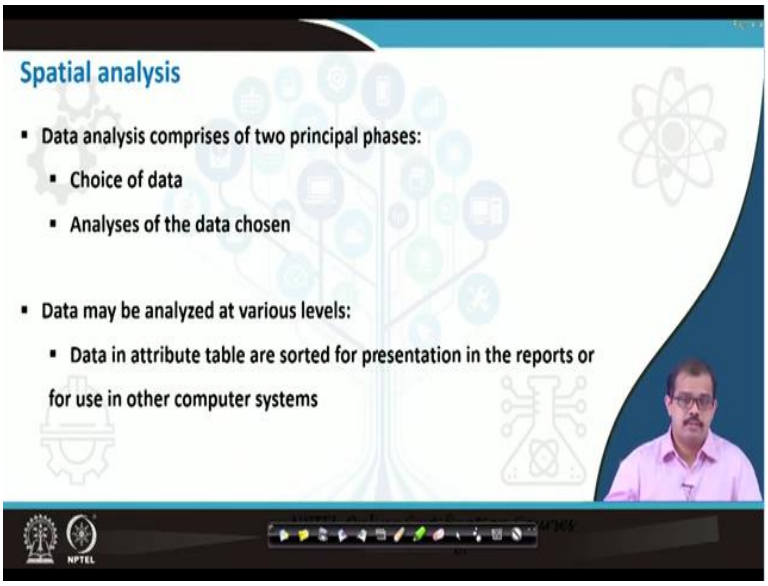
So, next coming once now we have understood what is database, then we have looked at GIS data. So, we have looked at how do you store a data okay a vector model are basically how do you store a vector model. So, similarly, you have lot of raster formats you which you use in normal daily routines. So, all of those formats also can be saved. Now, once we have done, now we have data stored in the database.

So, the next thing is if a user has to do any spatial analysis, that is what we are trying to see in this particular lecture. So, any kind of analysis that is like a normal data analysis can be done on all of these data okay, permitted that data can be used for such analysis, you cannot use a yes or no data to say how many number of I mean you cannot give a quantitative information with yes or no okay or having the color information you cannot say the quantitative information.

So, look at what kind of data and similar analysis can be done. So, when you look at data and said analysis consists of 2 different phases, one is the choice of data, what kind of data then analysis of the data chosen. So, both of this has to be understood, choice of data will be telling what kind of data it is basically, then how do you analyze the data that you have already choose it is another part of the spatial analysis that you will take up.

Now data may be analyzed at various levels, data in the attribute table are assorted for presentation in reports or use it in the computer systems okay. So that depends on the user.

(Refer Slide Time: 11:12)



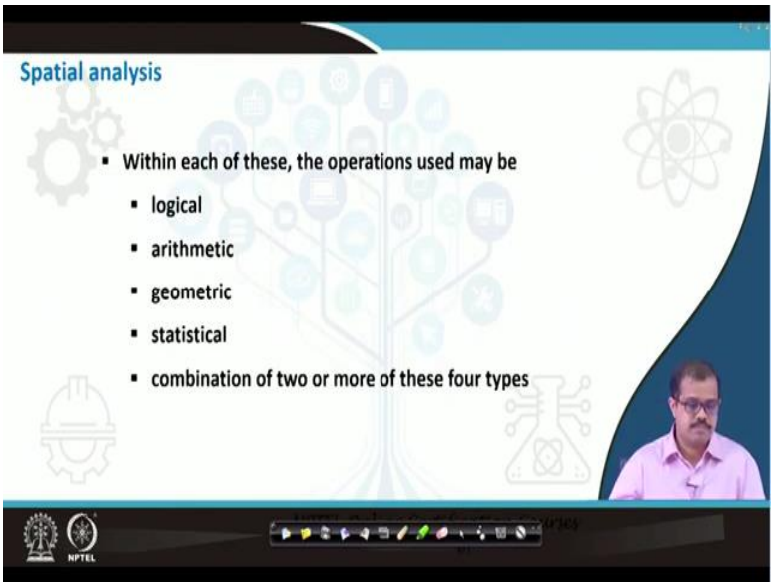
Then operations are performed on geometric data either in a search node or in the computational purposes. So, that depend again, probably when you start using the data, you will understand what is a search mode and or of computational mode on how the operations are performed. Arithmetic, Boolean, statistical are performed in the attribute table okay, I will show you at the end of the probably in the 5th lecture of this particular model, how an attribute table looks.

How do you do and using just a mathematical operator there, how do you perform analysis. I would not really do it but I will show you how it is done. In the practical class, we will show you how the operations can be also be done. How do you compute area, how compute land, how do you compute different aspects if you want to add minus, subtract, add or Boolean operations etc. you can do it and any of these are with the spatial analytics text.

So, geometry and attribute tables are used jointly. So though they are separately stored in many of the distributed databases, so, they can always be used in order to compile new set of data based on original and derived attributes okay. So, just by using the same attribute information that you have already stored or the information that you have already stored, you can create a new set of outputs, reports or you can use those attributes and derive more new information.

And then use the same data for your analysis or whatever that list of the attribute of concern, then compile a new set of data based on geographical relationship. What is the relationship between the geographic data that also we can do with this spatial analysis.

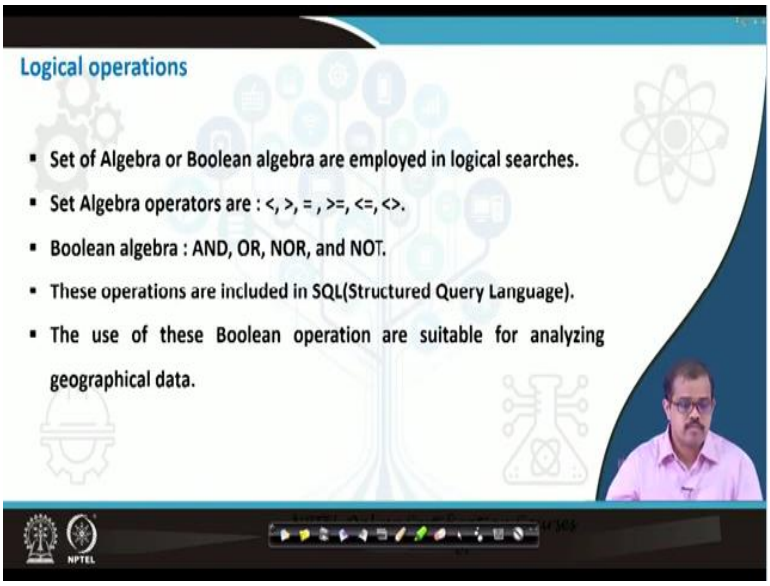
(Refer Slide Time: 13:02)



So, when you are looking at it with each of these operations, the operations use maybe one is logical arithmetic, geometrical, or statistical and combination of 2 or more of these 4 types. So, basically 4 types of data large operations can be done one is logical other one is arithmetic, other one is geometry, other one is statistical. So, I think most of you have done this analysis in excel one or the other way, okay.

So, combining this the same thing can be done using a database excel, you will have to look at where the data is how the data is you cannot query basically and try to do it. But with the database, you can easily query use that query to understand and to the spatial analytics.

(Refer Slide Time: 13:48)

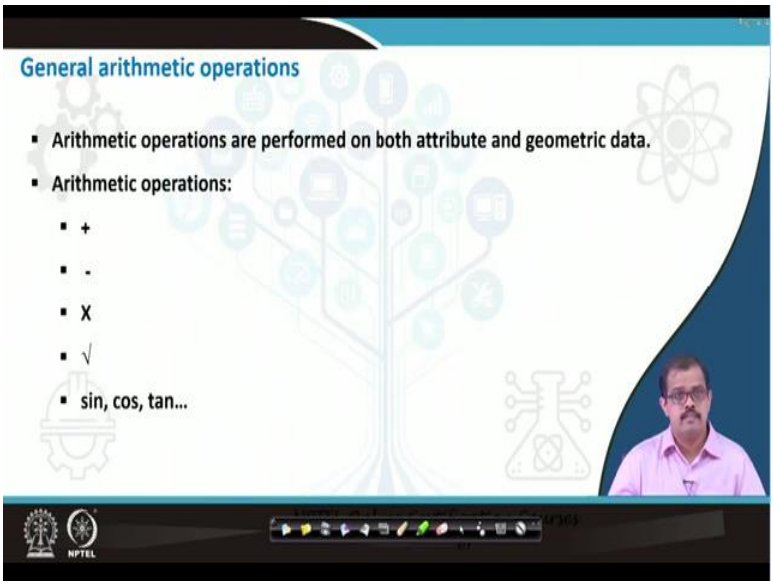


So let us look at each of these operations. So the first operation is logical operations. So when we are looking at logical operations, the set of algebra or Boolean algebra are employed in this logical searches, set operations such as greater than equal to, greater than equal to, less than less than or equal to such operations are normally used. For example, I gave you an example of like operator in my previous lecture, which is actually mentioning that are multiple statements that it is actually taking if the state is equal to West Bengal or the state is equal to Kerala okay.

And if the population is greater than 5000 and population is less than 10,000 between 5000 and 10,000 you want to query or population is greater than or equal to 10,000. So, give me the outputs of these districts. So, that is what here the logical operations mean the logical and or the logical usage of those values. Then you have Boolean algebra like and, or, nor, not. So they are extremely efficient in terms of usages the is Boolean algebra where you use and, and gives you the combination of both the set of data and the combination of the union part of the data is and or as either of these okay.

So, then you have nor, not. So, any of those operations of Boolean algebra can be used, these operations are included basically in structured query language. So, most of these basic operations as our dictionary is already there, the use of these Boolean operations are suitable for analyzing any geographical data, okay.

(Refer Slide Time: 15:37)



So, when you are looking at arithmetic operations, there are lot of arithmetic operations are to be performed. For example, both attribute and geometrical data arithmetic operations can be adding, subtracting, multiplying, it may be your square roots, it may be sine, cos, tan. So, even this technological operations can also be perform using general arithmetic operations. So, other than this you have not statistical whatever statistical operations that you see in your daily routine, all of these operations can be performed in terms of whatever data set you have.

So, it only thing that you have to maintain is that why that particular operations is being done, okay. So, in case you understand that these are the operations that has to be for example, let us say that there are certain set of data, okay, normally when people develop a database the main thing especially with GIS data, the main thing they forget is that connecting GIS data geometric data with all these data in terms of what is the access level that a data user has, and how this data is being maintained and how the data has been changing.

For example, if you give enter access to a new user, he or she may even or whatever the data is already present, it may be lost, or it may be even corrupted in terms of doing that unless he or she does not have proficiency in doing it. So even when in the geometrical data and your database, maintain the access level to each of the users, so that is very important. If you look at most of the websites that are connecting the geometry data with your databases, there are huge databases that are being employed.

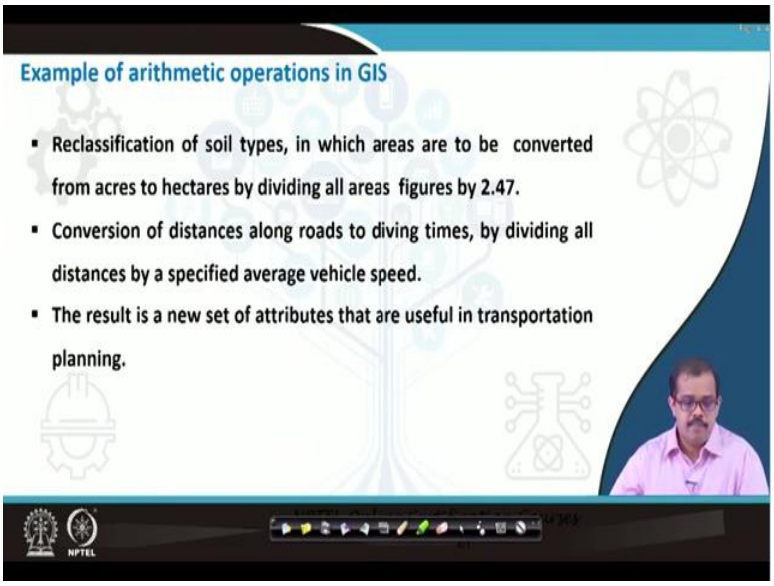
But only thing that what I see is that though they have a front end that is extremely beautiful, but the back end is not so strong in terms of database being extremely secure. So if your database is secure, then probably your most of your whatever the data connect, is happening in a proper way, or do not allow many of users to actually edit your whatever the basic data is. Only thing is that they can add more information subject to that it is supervised by someone who already has a proficiency in that subject.

And approves it, or he or she can add or delete information only in the separate table, which can be used by the organization or can be deleted by yourself when you are maintaining the database. So keep all these things in mind when you are actually allow any access to the user, any type of access to the user, so, be very careful. And most importantly, have security access, encrypt your database have a security access.

So that any user does not tamper your database in any form. So because why I am saying this is all of these operations when it is done, there are certain tampering of the databases may happen and most of the database are prone to have issues so have some backups of if you can store it as mirrors every day a mirror can be created and can be stored or something like a file system that can be stored daily.

So any changes may be previously reverted to the way just a day with the loss of only certain information. So that kind of thing has to be there. If you are trying to implement any of these operations and allow the user to do it, even with yourself, it is better to have all those things ready because you do not know what I mean how the database can get corrupted, how your data the hard work data can be really going in way. So please look at all those issues in encrypting and securing the database.

(Refer Slide Time: 19:31)



And when you look at the example of arithmetic operations and GIS, reclassification, for example, if you have soil types, okay, there are soil types that are done if it is converted, so I have given it as figures and figure 2.47. So when you are looking acre to hectors most of them do this mistake. What they do is when you are trying to do reclassification of soil, where areas to be converted, they do not know what is actually the index terms, whether the representation is in areas or in hectares.

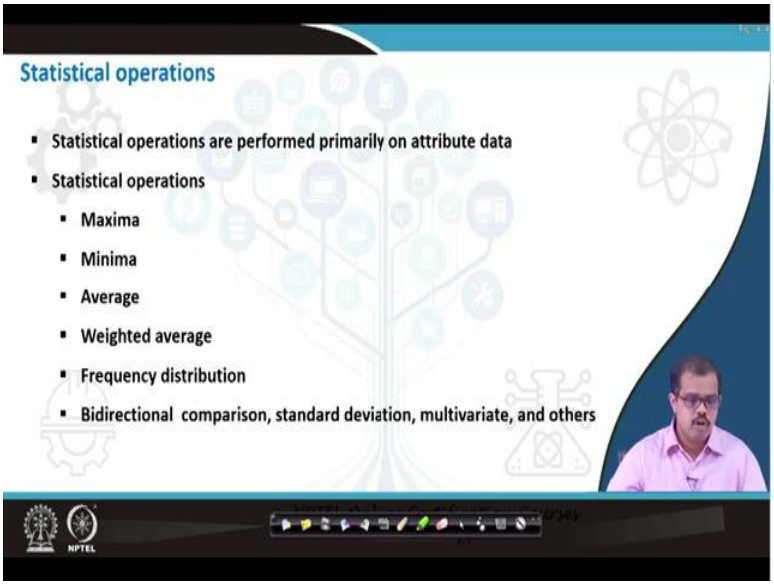
So, they normally just put it as units, it may not be always correct to suggest every time the report to be units okay, you have to be very specific in terms of area. So, look at the metadata how this particular I mean data is being represented normally when you have certain data types that are associated is being shown to you, I normally they may have thematic classifications and each thematic may be given us for a certain value.

And that name okay value in terms of a string and also associated units. So, with the unit try to use it most of them make a mistake here. Either they give it a like for example 65 units or sometimes if it was in acres they represented 65 hectares. So that really gives a mistake, okay. But in case you want to do it, reclassify anything, look at it extremely careful what kind of units are being used.

For example, I am giving an example here with acres 2 hectares by dividing all fingers by 2.47. That is how we convert to acres to hectors, then conversion of this distance along the roads to dividing times by dividing all distances by specified average vehicle speed. So you can do that. So only thing is you using athletic operations, it is much easier. So you do not need to repeat this operations like you do manually in most in the excel sheet.

The result is a set of attributes that are useful in transportation planning in terms of what examples that I have given, you have soil types, you have the road distances. So based on that you look at transportation planning okay.

(Refer Slide Time: 21:56)



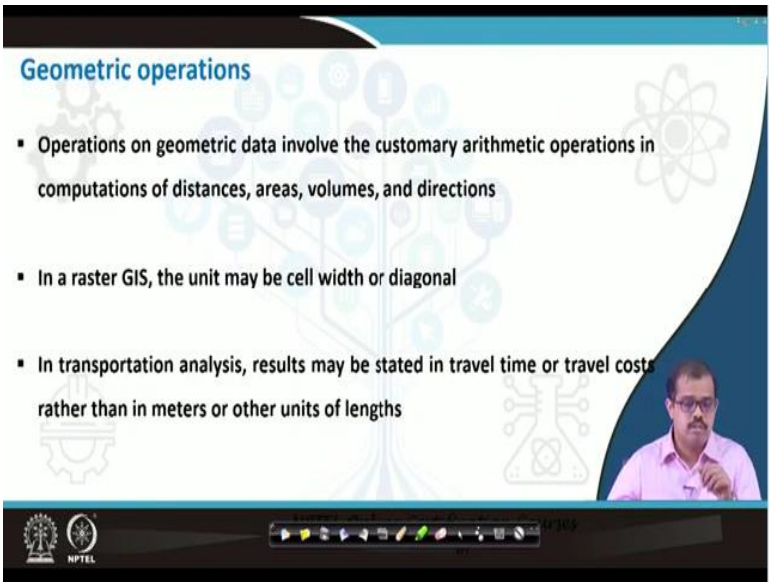
And the last set is that the statistical operations. And when you look at statistical operations, you can perform the maxima, you can perform minima, look at the average, weighted average, frequency distributions. You can look at interpolations, you can look at any of those things that you are trying to do multivariate statistics, you can look at a very basic statistics including standard deviations, plotting curves, plotting different graphs.

And plotting also advanced statistics extremely useful in terms of handling or advanced statistics whatever operations that you can make in most of that statistical you can invoke for example, most of the GIS software's have either a connect to any of the modeling framework, okay, most of the GIS software can easily handle using Python programming. So, once you have a Python

programming any package that at a very advanced package can be used with the tools that is actually embedded in there most of the software to analyze and to export any information that may be necessary that is used for various purposes.

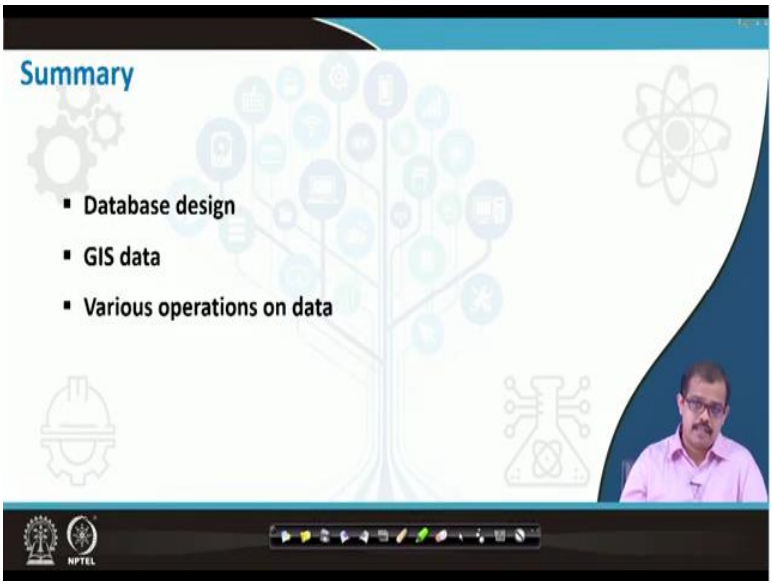
So it can perform any kind of statistical operations. Just I would also suggest people who want to be advanced GIS users look at Python, very effective, very useful in terms of programming, and also in terms of using already available tools in the way you need okay, so you can customize it much easily. Whereas most of the proprietary software will never allow you to customize it whereas in Python, you can customize the way that particular tool is even if it is a proprietary software, okay.

(Refer Slide Time: 23:39)



So, the next thing is geometric operations. Operations on geometric data involves the customary arithmetic operations in computation of distances, areas, volumes and directions. So in a raster GIS, the units may be a cell width or a diagonal. Normally when you look at the rational model, it is based on cells, it is based on raster's that are there okay. For example, in our transportation analysis results may be stated in travel time or travel cost okay. Rather than in meters or units of length. So, look at what application it is.

(Refer Slide Time: 24:16)



Do not use unnecessary conventions of how do you represent a data. So, each of the analysis, each of the output needs the way the data is to be exported. So, look at for example, as I said transportation analysis it is better to give travel time and travel costs, rather than giving that it is now how many kilometers or how many meters or units of land that you may use at time to time. So, look at different units the reports has to be very specifically designed in terms of the queries okay look at every stakeholder that is very important.

In summary, we looked at database design, various operations on databases, we looked at GIS data, what GIS data can handle. So, we looked at we have 2 GIS data, so geometric data and attribute data, how do we store both of these different forms of storages and how do we extract information from it whether you are using logical operations arithmetic operations, Boolean operations or a statistical operations.

And we also understood that how we use different sources of data and how different databases are actually stored. So, this is what we learned in this particular class. In the next class, we will go into more of spatial analysis, we will understand more of spatial understanding of the special tools of understanding how and why it has to be done. So, till then have a nice time and thank you.