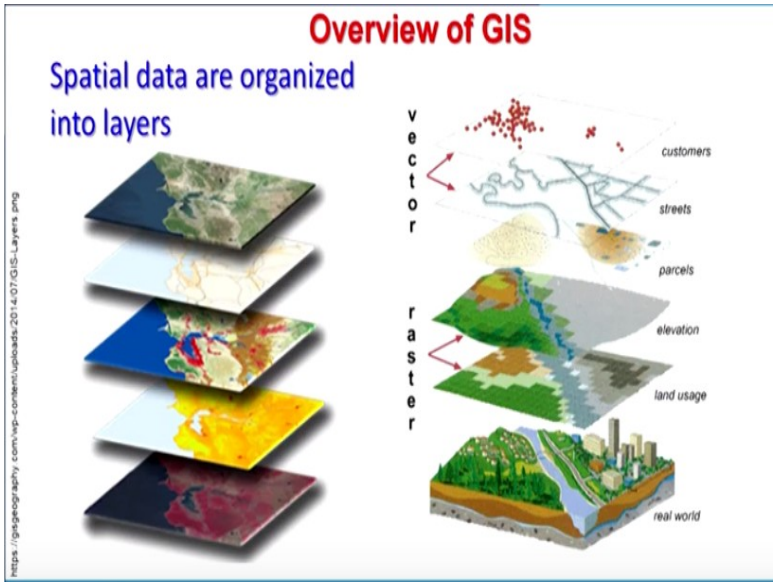
Geographic Information Systems
Prof. Arun K. Saraf
Department of Earth Sciences
Indian Institute of Technology – Roorkee

Lecture - 02
Essential Components of GIS

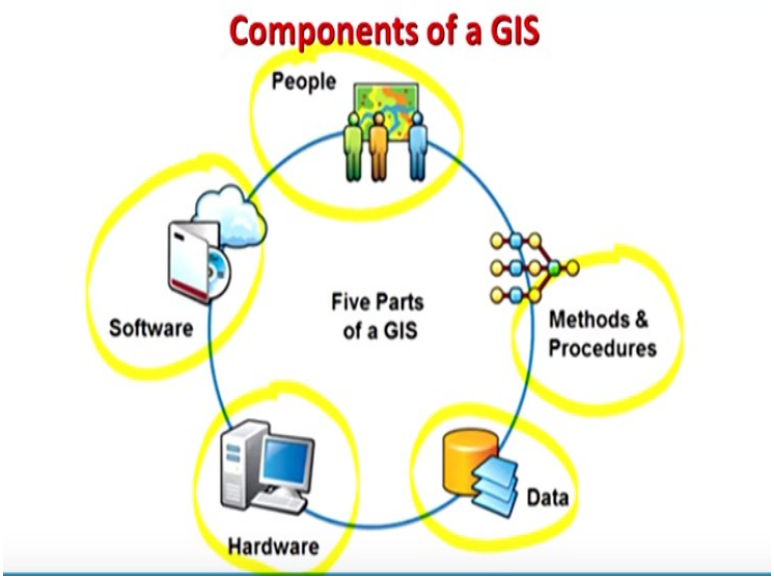
Hello everyone and welcome to new discussion which is on essential components of GIS. Very briefly when we were discussing, what is basically GIS? Indirectly, we touched these components but here now we will have full or detailed discussion about this. The figure on the right side; we have already discussed that the real world can be segmented into different layers.

(Refer Slide Time: 00:45)



And these layers represent the different types of data which we handle in GIS platform like point data, line data, polygon data, raster data, images data and so on so forth. So, generally as we know that in GIS, spatial data is organized into layers. One important point which I want to mention; whenever I mentioned in this discussion related with this course GIS. Whenever I say data; I always mean is I am talking about the spatial data.

(Refer Slide Time: 01:23)



Now there are basically in different literatures, books or even on the webpages, you may find that people are having different number of components against the GIS. But I find that these 5 components are best representing what are basically GIS or components of GIS. Like every system will have its own component. Like if I give an example of a computer; say a desktop computer will have a screen, will have a keyboard and will have a mouse, hard disk and a processor and many other components also.

So, similarly GIS is also having components. More or less now, people are agreeing that these components can be divided into 5 categories and basically as shown along the circle. So, each component is important. We cannot say that this is the first component and this is the second component. So, that is why it is shown the cyclic thing is there.

So, I start from this bottom left that hardware; you need to have because by definition as we know that GIS is a computer-based information system. So, you need to have hardware. Of course, you need to have the software and then comes the data. Here I would like to give an analogy which is like a GIS is a carpenter's toolbox. Some people will laugh on this that what kind of analogy is this.

But of course, carpenter toolbox is not a computer-based design or computer-based system. But let me elaborate on this that this carpenter's toolbox will have only tools; like software tools so the GIS. Carpenter will have the hardware tools. In GIS, we are having software tools but tools this is what is. But the carpenter's toolbox will not have the data. So data; there is missing.

On the GIS platform also, if you get proper hardware and install GIS software which are many, still you will not have any data on your machine. So, your toolbox which is including hardware and software do not have any data; the same way as carpenter's toolbox. So, it will have different tools like hammer, saw, drill machine and all those things.

But it will not have wood, mica and some glue and so many other things which are required to prepare furniture. And second thing that carpenter's tool box will also not have a design part; how to design a chair. I can buy a carpenter's toolbox very easily in the market by paying few 1000 rupees. But the design of a chair will not come; design of a table will not come.

That has to be developed from somewhere. Some expert has to be there or some experienced carpenter can design a thing or a user can provide the specification that I want table of this height, having this length and width and made from this material. Now that carpenter will prepare according to user defined specifications. Same thing here we are doing that a GIS expert will prepare maps or data or analyze the data as per user defined specification.

The GIS expert can suggest that in order to have that kind of design or map, you require this kind of data. And this data may be available from these sources. Like a carpenter can suggest you that if you want to buy the raw wood, it is available at this shop. If you want to buy the mica or glue, these are available at this shop. Otherwise, after installation of GIS software on computer hardware, you still do not have the data.

As I recall the definition, data coming from variety of sources. So, it is not residing on your machine initially. It will come from variety of sources and then you organize and make it, you know in such a manner then you can apply the design or user defined specifications for analysis and modeling. And finally, of course comes; the people. People are at every stage; for handling software's for maintenance of hardware, for generation of data or analysis of data and of course, the method and procedures for towards the development.

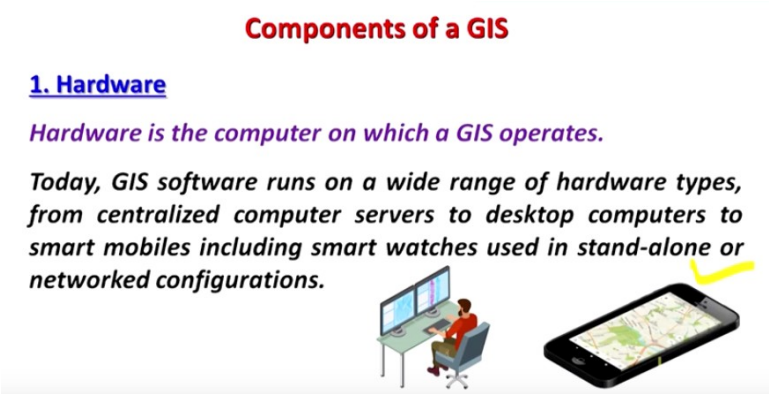
And top of these; the decision makers are most important because they are the real users of GIS and they are basically compelling since beginning of the development of GIS, for further development of GIS because they keep asking questions, they keep asking new products from

GIS. And GIS experts keep looking either the tools or methods available within that GIS platform or in another domain. Lot of things might have developed in mathematical domain or in computer science domain which has not yet been incorporated into GIS. But if we users are looking certain solutions and if we know that people have already developed that kind of thing then it can be incorporated.

I will give you one very common example of this like wavelet; it was wavelet theory. So, first it was in the mathematics then it reached in the computer science; coding started and now wavelet is being used also for digital image processing or in GIS and also for data compression which is again a very common part of GIS. So, something which is developed in mathematical domain may not directly be incorporated into GIS.

But when it goes to the computer scientists, they develop code, create software or they incorporate with the GIS software and further your GIS software or whole GIS technology becomes much more powerful. So, there are many such examples where things have been developed on different domains and then they have been incorporated into the GIS platform.

(Refer Slide Time: 08:05)



Now, we will look these components one by one. Hardware is basically is a computer on which a GIS resides or a GIS operates basically. Now basically hardware you know can be a very high-end workstation or some other machines or a server or hardware can be even nowadays smart mobile which is also shown in this slide. So, from you know big computers or workstations or servers to your mobile, all kinds of hardware's.

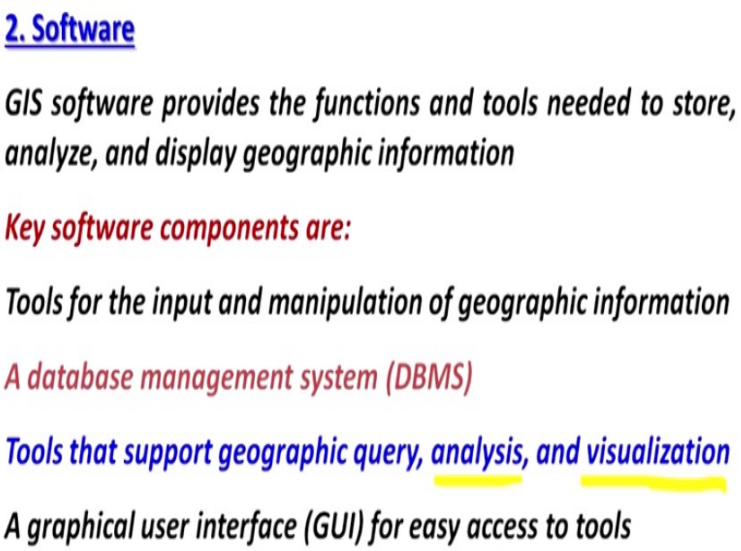
Now GIS software's are available or custom designed GIS software's are available which can work on different hardware's which is available today. So, GIS can run on a wide range of hardware types from centralized computer servers or to desktop computers to smart mobiles

including now even smart watches. Some stripped version of GIS or very small utility of GIS is available, standalone GIS or network-based GIS and then you use the GIS.

So, the hardware now varies having a wide range of hardware types on which GIS can run. Next component, very important component is the software. When I started discussing about the historical development of GIS and in the previous lecture, discuss about how Roger Tomlinson was known as father of GIS, he basically developed initially the software that is you know, the Canadian GIS or CGIS.

So, the software is very much part of GIS. GIS software provides the functions, tools which are required to organize the data, store the data efficiently and of course, analyze and model and display that data or that geographic information as per user defined specification. So, software is equally important as your hardware.

(Refer Slide Time: 10:03)



And key components of the software or you can say subcomponents of GIS are: the tools for input data. Because sometimes data may come in front of you in analog form, a table, a register. Some organizations have noted the data. Now you have to convert into digital format. So, maybe it has to be typed even sometimes, yes. All data even today, it is not available in the digital format or in organized fashion.

So, we have to bring the data. So, our GIS software should be capable of allowing data to input into the GIS database. So, input tools must be there. Sometimes we are having a map in a print form. Now we want to bring on a GIS platform, so that has to be digitized. And for

that also the software tools support. And nowadays these modern GIS software'; most of them supports the input or digitization of maps and other geographic features.

Manipulation does not mean that you know frilling with data or you know, creating errors with the data. No; not at all. Manipulation means changing from one format to another. Maybe changing you know, the way the data was originally collected to the data which is suitable for this. And manipulation might be; the original data may not have the location. But now, I want to make it as a spatial data.

So, I will attach location with each data set. That location might come through fields investigations and using the GNSS receivers. So, these tools must be available in good GIS software that is what the components here are. Database management system; each GIS software either will have its own database management system which can be used.

But for a large data analysis or when you are having you know; huge amount of data or people say big data then maybe sometimes you have to resort to the external DBMS systems like Oracle and other things. So, if I give an example of ArcGIS which is ESRI; Environmental Science Research Institute. It's a private company; commercial product is nowadays called ArcGIS.

And ArcGIS supports to a large extent its own database management and it is having own system. But if data is very big then you would find limits within ArcGIS then you have to resort to some external GIS or database management system like I gave the example of Oracle and its other. Also, the software; a good GIS software should also support the query. Once the database is there, there may be query of one type today.

But tomorrow some other user may raise a different query from the same data. So, we need to efficiently retrieve the data and that can only be done if data has been organized in a proper manner. And of course, once everything is organized then our GIS software should also support analysis as far as possible. All types of analysis, even with the software which I gave the example may not support.

So, suppose I want to do analysis related to the air pollution. Now, special tools may not be available which mentions that for air pollution analysis, you use this tool but basic tools are

available. So, if I know the inside of GIS or when inside is out of GIS to me then even using basic tools, I still can perform analysis. So, you know the analysis part is very important.

Earlier we used to spend lot of time to create the spatial database on a GIS platform. Because in early stages of the development of GIS, lot of data was not available in digital format and not organized with geographic locations. Now, large amount of data is becoming available which is organized and which is geographic data. Like for example satellite images.

Now, lot of satellite images are available free of cost and they are georeferenced that means the geographic coordinates are now inbuilt in those images after the initial processing by those who have acquired the data. So, that analysis processing, I do not have to do now. I just download the image and it is ready to use on a GIS platform.

Similarly, many other data set like census data or groundwater data or meteorology data. Lot of such data sets are available now in organized fashion, compatible with GIS platform, can be used directly on a GIS platform. So, now instead of spending time of organizing or creating a spatial database on a GIS platform, now we can spend more time on analysis.

Because analysis is really the part where we should spend lot of time and learn different tools and techniques. And analysis which is being done in other domain may not be today in GIS. So, we should try to incorporate how we can perform that such analysis on a GIS platform; bridging the gaps which are in different domains which are available. And of course, visualization is very much required.

So, in form of either map which people prefer many times. Because you know that very old saying, a picture tells 1000 words and I have added into this phrase that a picture tells 1000 words and a map tells 10,000 words. So, people prefer because we can visualize things, we can see the pattern; we can see the distribution and along with some other data sets. The example of that is corona epidemic.

So, as soon as the data was plotted, people knew that which well water is causing problem to the resident of that area. So, visualization in form of maps is always preferred by many people. Otherwise, charts can also be created or you know other graphs or tables and other

things can also be created on a GIS platform. So, the second component is equally important as first component.

Once you have setup or got a hardware, rest of the things are fixed. But the software keeps getting updated and becoming more richer and richer and new tools can also be developed. Another important point before I go to the next component regarding the software; if you are thinking to have your own GIS system or one would like to install or buy a GIS software then you should look at GIS software, which is having a possibility of appending further tools in it.

That means it should support some outside programs to which I can integrate with existing installation because that will allow me to do lot many other analyses which a standard installation may not support. So, that is one thing which one always should look. Another thing which also one should look if you are choosing a GIS software which are different; many are available in 1000s then you should also see whether that software is in public domain or a commercial one or a open source or commercial one. Because for commercial one; you may have the support from the company.

But open source sometimes you may not get support. I am not against the open source but these are the difficulties sometimes you may encounter. So, these things are there. I will give you an example of open-source GIS which was there since ages. It used to call as GRASS. And it was developed by US Corps of Engineers like NCC in India.

And those people developed this software. It was available free of cost. But people face including myself when I started using GRASS, we face the problem of getting support. But nowadays things have changed. And QGIS is one of the best open-source GIS software available. Anybody can install and can use. It is very good software. It has got lot of support, discussion groups; all kinds of things are available with QGIS also.

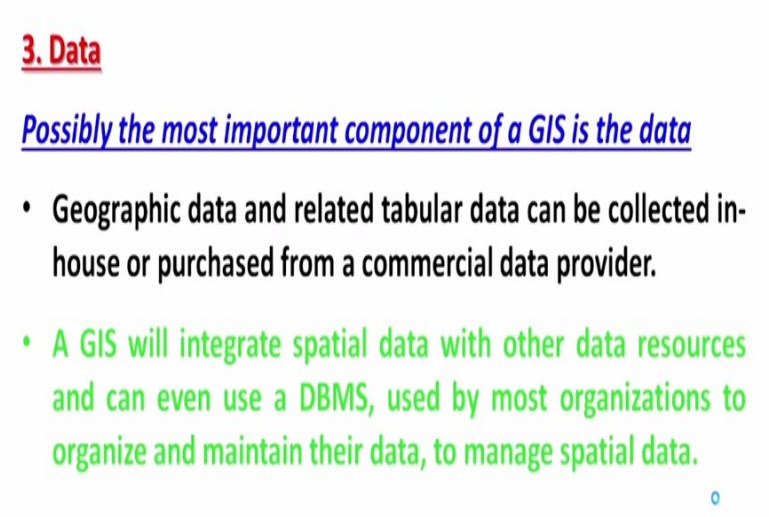
And QGIS will whatever the software tool which I have discussed is having. I have seen; it is having. I have used that one and it is all having as commercial GIS. So, it is not that commercial GIS should always be used. No, if your work can be done using open-source GIS that is wonderful. Now, when a GIS was developed in 1964 and onward by Roger Tomlinson. It did not have the graphic user interface that means everything was in command mode.

So, we used to type even when ArcInfo came on a desktop machine like in name of PC ArcInfo, everything was in command mode. So, whenever we want to display, we had to type and give the instruction or commands to the software. Now after this invention of mouse and graphic user interface, the accessing tools that means software tools and data has become really very user friendly.

And that is the biggest you know user friendly development which has taken place in the GIS domain is this graphic user interface. So, it has become much easier, much user friendly as 20 years back. The heart of GIS is the data. I am talking about spatial data. Without data, it is something like carpenter's toolbox having tools but does not have the material to prepare or manufacture furniture and it does not have the design.

So, if we talk about this material then material in GIS is data. As rich your GIS database can be, for an enterprise or for an organization then you can do lot many analysis on it. But if it is not rich in data and good quality data, clean data then such analysis cannot be performed. So, possibly the most important component of a GIS is the data.

(Refer Slide Time: 22:21)



I would definitely repeat again that if I have to put in hierarchy then it becomes very difficult to say that hardware is least important or data is most important. But still I can buy a hardware; I can arrange a software or by a software or open source software. But if I do not have the data, everything becomes non usable. So, that is why here it is mentioned that the possibly the most important component of GIS is the data.

Now, geographic data or spatial data and related tabular data; data in GIS be divided into different categories. But 2 main categories initially: one is the geographic data or spatial data. Another one is also called non spatial data or tabular data. Tabular data because when we display it is displayed in form of tables like Excel sheets. So, that is why it is called a tabular data or also called non spatial data or also called attribute data.

So, each spatial data can have as many as or n number of attributes data depending on type of data and requirements. For example if I am having earthquake epicenters so, location of earthquake epicenter becomes my spatial data. But what is the magnitude of that earthquake, at what depth it has occurred and what are other details about like fault plane solutions or geology or other things; that will go as an attribute data.

And there is a direct linkage or dynamic link between spatial data and non spatial data. What I mean here if I do a query on tabular data, the same time, the same selection is done on the spatial data. Vice versa is also true; if I perform certain analysis or query over my spatial data, the same time it is also shown with my tabular data or reflected in my tabular data. So, you know, these 2 data sets, one is spatial data and non spatial data.

They are integrated in a manner that they goes hand in hand and whenever I want to use it, one data or query one data, other is automatically is also queried. Now data can be created by ourselves; going in the field or collecting elsewhere. Or we can purchase from a commercial data provider. So, lot of data now also available through companies which collects the data, organize the data and they make the data GIS compatible.

Or, nowadays lot of web portals is available which also provides data; many times free of cost. I gave the example like remote sensing data; satellite data of various satellites of various countries on daily basis, it is available. As soon as the data is acquired by the satellite, let me give you an example like Landsat 8 series; as soon as the data is acquired, in 10-15 minutes time automatically that is available for download.

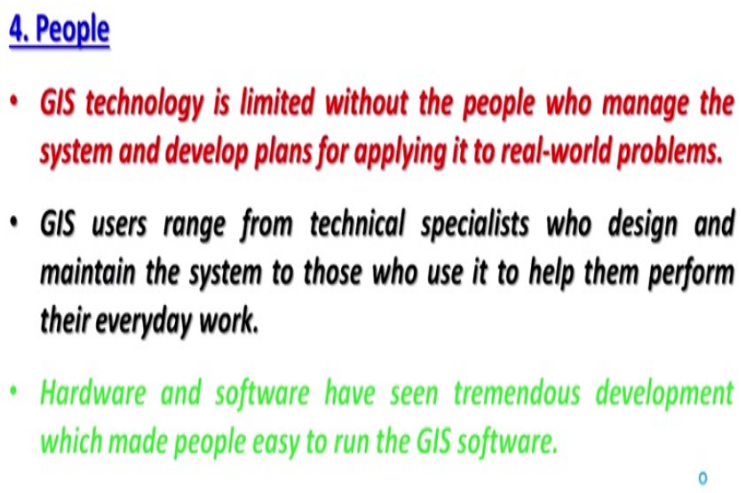
And that too in an organized manner; organized means it is already geo-located or georeferenced. And directly you can bring it onto a GIS platform. And there are satellites which may provide data on daily basis or hourly basis and that too in organized fashion,

many times you can download free of cost. Or otherwise if it is not available then there are commercial providers.

Sometimes when these resources get exhausted, you do not find the desired data either through your free or commercial providers then you have to go in the field and collect the data. So, data can come from variety of sources; remember that thing. Now basically what GIS is doing?

As we have been discussing that GIS will integrate spatial data and other data; your tabular data, non spatial data, data from other sources and in a DBS or in a database management system and then organize, maintain their data and manage spatial data. Now fourth and again very important component because without users, without people; GIS is nothing, so, it is equally important. All 5 components are equally important; so the people, users.

(Refer Slide Time: 27:21)



GIS technology is limited without the people who manage the system, who run the system, develop the system, develop the tools, and develop the bridges between different domains; mathematical domain to GIS, mathematical domain, computer science to GIS and many others discipline. Somebody might have developed a hydrological model to predict discharge on a river stream.

Now, can I use or can I import, can I decode or recode in a way so that I can also use on a GIS platform. So, you need expertise; either external or yourself to integrate that model into a GIS. So, such understanding or models may be available in other domains, one has to only

bring it in. And of course, people develop plans for applying it to real world problems. I am looking for groundwater or some mineral resources or petroleum.

Then I plan my exploration using GIS and using different data sets. The data set might be satellite images, data set might be the digital elevation model and data sets might be the borehole data or log data. And then I tried to analyze and try to see and other knowledge; geological, geophysical knowledge. And then I tried to estimate that where I can find the water and how much water I can find, how much I can bring on the surface.

So, many such things can be done by the people within the GIS. So, real world problems can be solved. As in the previous lecture, in the last slide we had the project development of GIS. So, first to understand the problem; what are the key components of the problem? What are the key parameters? And then try to organize those things on a GIS platform for analysis.

GIS users; there are a variety of users range from technical specialists who design and maintain the system to those who use it; just users to help them perform their everyday work. Let me give you again the same example of like Uber or Ola apps. We are simple users. We are not a technical specialist or we are not as a developer. But our experiences, our feedbacks give feedback to these specialist or developers to improve their system further.

So, we are not designers but we may suggest sometimes that if you do this thing, if you add this also in your app then it may become further useful. So, the feedback or the questions with user raised always compel a better development of particular or custom design GIS software. So, users can come from variety of places or variety of range and of technical specialist.

So, the decision makers are also users. Users those who are maintaining the GIS system, they are also users. Those who are developing GIS software, they are the technical specialist and so on so forth. So, they are all people. They all are contributing towards; even users are contributing towards the development of GIS. Without users; GIS is nothing. Now hardware and software; tremendous development have taken place in these 2 domains which made people easy to run the GIS software.

As I have been mentioning that initially when GIS was developed in Canada, it was on Dec 20 machines and the programming language was FORTRAN. It does not have the graphic user interface and a very limited hardware capabilities and of course, software capabilities. It used to only handle initially vector data. Because Roger Tomlinson was given task to maintain these networks which were available within the municipal boundary of Ottawa.

So, he developed only that but later on, along with vector data, raster data handling also became and now lot of variety of data can be handled on GIS but these 2 major things; vector and raster and of course, non spatial data. So, vector and raster comes under spatial data category and attribute data comes in non spatial category. All these types of data can be used on a GIS.

So, as today if more development takes place in computer hardware which is happening almost every day, after sometimes that is percolated into the GIS. Similarly more software tools, better language, better programming, more efficient programming is done then that also percolates on the GIS platform. Let me give you an example. I have already mentioned that initially the GIS were developed in FORTRAN programming language.

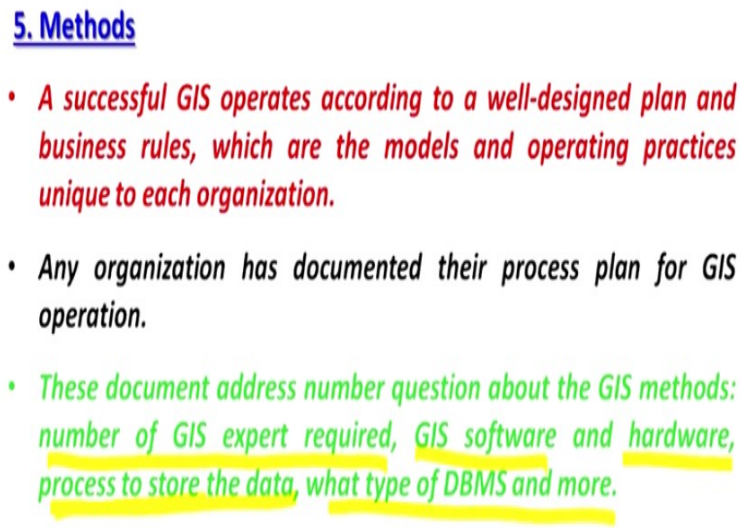
Nowadays not many people use that language. Later on, like if I take the example of this ArcGIS; ArcInfo was developed on UNIX platform and you know, that some different languages. Then they started using this; initially it was on C++ then they started using Visual Basic. And nowadays there is another language; Python. So, lot of development which one can do with Python also is integrated onto a GIS platform.

So, particular software has been developed on a particular language but it can also accommodate the development which is being done or has taken place in other different languages. So, if somebody has developed a software tools for GIS analysis in C++ or visual or Python can be integrated into current GIS software like ArcGIS and many such as software.

So that the development which takes place in hardware and software domain is coming directly into GIS development and GIS tools or GIS technology is enriching from these developments on daily basis. And last point here is; the people are main component for the successful GIS. Successful GIS means having installation, having spatial data and now I have

started producing through my analysis; various scenarios, various results, maps and other things. Then only it would be called a successful GIS.

(Refer Slide Time: 34:50)



And last component in GIS are the methods because these are the algorithms. There might be some algorithms, they might be some equations or some models already developed in different domain. All those can be incorporated; may not be today but after developing better understanding between GIS developer and a domain expert, these methods can also be incorporated into GIS.

So, a successful GIS basically operates according to a well designed plan and business rules which are the models and operating practices unique to each organization. So, depending on for what purpose GIS is being used, for which the models and operating practices will be developed and accordingly then successful operations of GIS will take place.

Suppose if somebody is working for a company which is related with you knows, development of highways and roads and other railway tracks so, they will be using GIS only for that purpose. But If somebody is working for an organization which is exploring the natural oil or gas, they might be using the same GIS software but their data set might be completely different.

And their methods, approaches, operation practices within its rules may be completely different. So, the software, hardware might be same but users because the applications are different. Users will develop their own methods and analytical tools to fulfill their own

organization requirements. So, any organization has documented their processes plans for GIS operations.

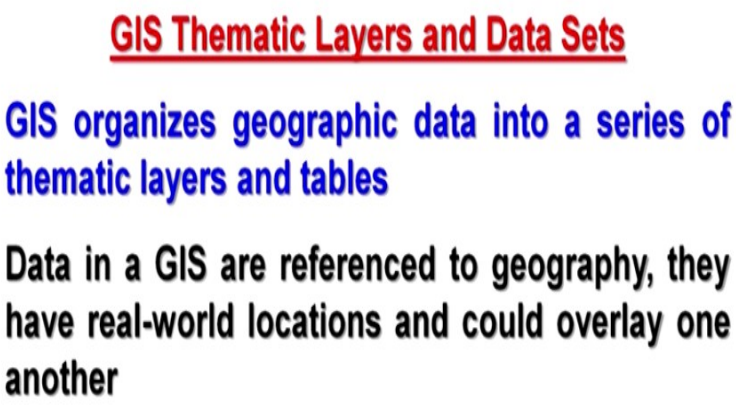
If I install GIS software like I gave the example of say ArcGIS or QGIS, everything is installed. But whatever the tools which are available but end of the day if I am working say for groundwater exploration, I might be using only 10-20% of those tools. But the same software installed if it is with the civil engineer, he might be using different tools maybe again 20-30%.

So, that documentation can be done by different organizations, different departments for their own purposes. So that it can be used by the future generation people that this software and these tools are specific to our organization. And these documents address number of questions about the GIS methods, number of GIS expert required, GIS software, hardware, process to store data and what types of DBMS and more.

Because a big enterprise or organization should follow certain businesses rules and among these rules are documentation. So that an employee today is working for your organization; tomorrow he may leave, some new person can come. So, it should not happen that that person is gone so, everything is gone.

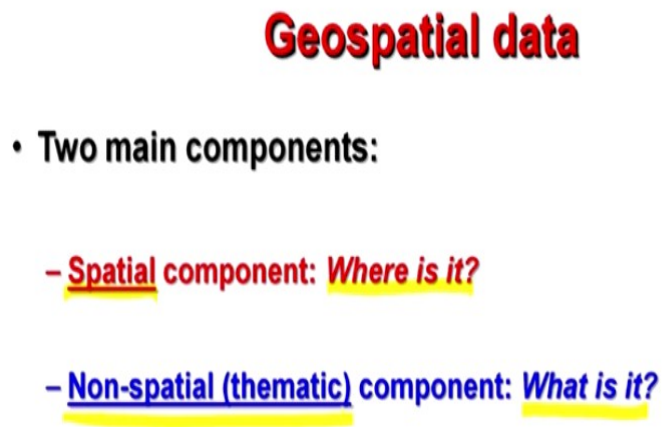
No, it is very well documented like along these lines; the number of expert required for a certain project, what are the software, what are the hardware and processes to store the data, what type of DBMS and methods. If everything is documented, a new person can adopt very quickly and can work on this. And a well designed plan will address all these questions. So, everything if it can be planned properly then it becomes really a successful GIS.

(Refer Slide Time: 39:09)



Now we have been using word GIS thematic layers and data set. So, I will spend few more minutes on this that GIS organizes geographic data into series of thematic layers and tables. And the data in a GIS are reference to geography that is their spatial location. They have the real world locations that are again in geographic coordinates and could overlay one over another.

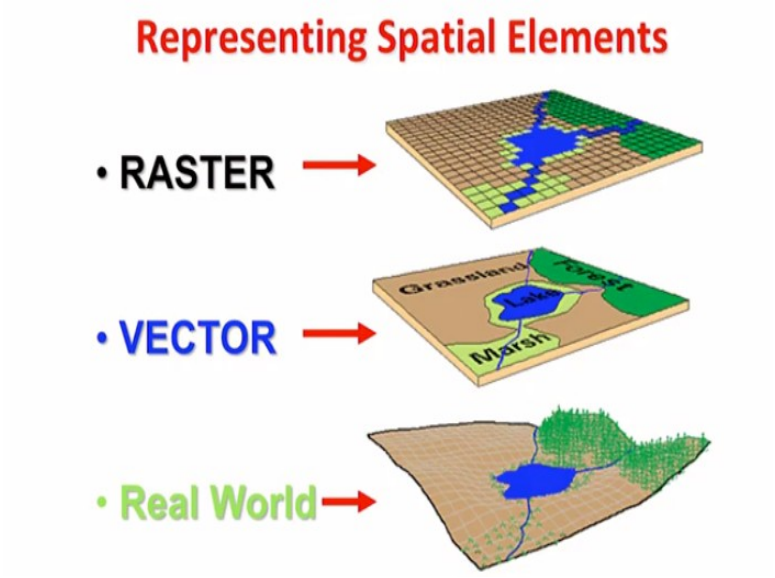
(Refer Slide Time: 39:22)



Sometimes the term geospatial data is also being used. So, geospatial data means spatial data. Spatial data means having geographic location. I earlier mentioned so, very quickly, I will cover these 2 points that 2 main components of the data or 2 types of data which we handled in GIS. 2 main types of data; one is the spatial data which is having geographic location and we can answer a question; where is it?

So, whenever there is a question related with location that will come through spatial data. And if we have a question like, what is it? Then that will come from the non spatial data. So, GIS data can be divided into 2 main categories; spatial data and non spatial data. Spatial data can answer questions related with location and non spatial data can answer questions related with what it is and so on so forth.

(Refer Slide Time: 40:27)



Now representing this part that real world can be segmented into different components; different layers, different themes, thematic layers so real world can be also segmented depending on the type of the theme that you are handling like vector data or maybe the raster data. In raster data you can very quickly understand as in a form of grid but grid and image are 2 different types of raster data.

And in case of vector, there are 3 different types of vector data. Vector entities; we have studied in our high school or 10+2. The same concepts of vectors are there that there is an origin; there is a destination and direction. So, this distance becomes our magnitude like in typical vectors and the origin and destination is known. So, we know the direction also.

So, the same concept but only thing is that here we are having 3 types of vector data; point data, line data or polyline data, typical vector and polygon. So, we will be discussing little later about those. So, this brings to the end of this discussion about the different components of GIS; mainly 5 components are there and different components are important. And we cannot put in any hierarchy, all are important and thank you very much.