Spatial Informatics
Prof. Soumya K. Ghosh
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture- 01
Introduction

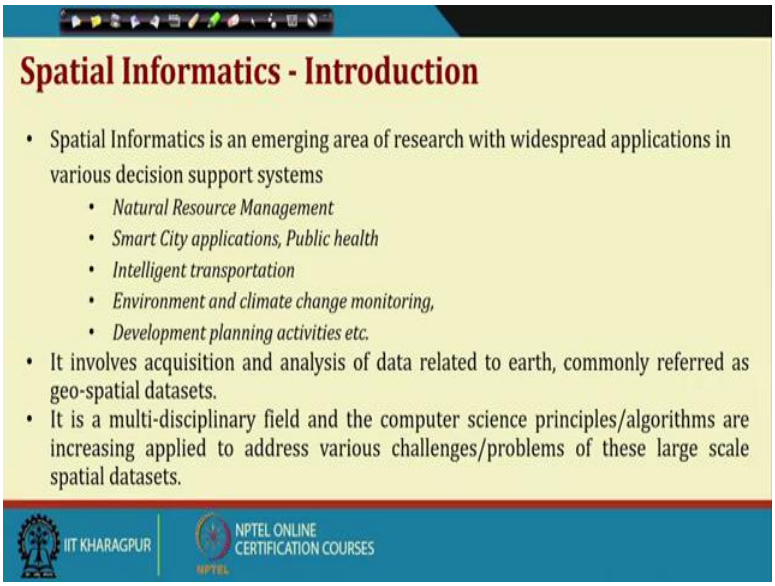
Hello. So, welcome to this course on Spatial Informatics. Today we will be having a broad view of that what we mean by spatial informatics. Rather, for many of you it may be a new course or new term, terminology though we are accustomed with different aspects of spatial datasets these days.

So, when we talk about spatial informatics more specifically we are we will be dealing with geospatial data; that means, data related to earth surface. This has a different implication. If you look at our all type of datasets what we deal with for our day to day activities at commercial level, at different aspects of decision-making you say there is a little bit of reference to the earth surface. Like think about say a student database, it says that which hall he or she is there which is the department he or she is studying at; that means, it is relating it with all other different types of attributes to somewhere put into the coordinate, even if we look at the home address and all other things, right.

So that means, it makes lot of sense to look at this how these datasets related to earth, but what does it mean and type of things, right. So, this course primarily who are around these aspects of the things. Though, truly speaking spatial in a sense can and can be anything on the space, right, it is not only our mother earth, it can be on a universal level, even if I talk about a body as a space then also there are different location, right, like eyes are in some places or location of a heart it is in some places and etcetera etcetera. But we will be mostly concentrating this particular course on the datasets related to earth surface.

Why it is so important? Why we require a separate way of looking at these spatial datasets or spatial databases? That we will see over the over few lectures and I believe you will appreciate that there is a need for have a separate dealing for those type of things.

(Refer Slide Time: 02:56)



If we look this different aspects, like if we look at that what are the different application areas which will be looked at. So, it is a first of all it is a emerging field it is for say one decade or so, it has came up in a big way. And, maybe due to use of maps, use of maps for finding routes, from finding a one location to another locations those are more popular becoming more popular and became more popular for our day to day use also. Not only that with recent or over this advancement of computational resources; that means, availability of huge computational resources as a much lower price these all those this data intensive, compute intensive this sort of domains has become more popular or becoming are more popular.

So, if you see there are these has wide scale application in different natural different domains, right. One is definitely natural resource management, right. Like I want to find out that how this my forestry coverage or how much coverage of the forest is that, whether it is how it is changing over time or say in separate in aspects like any type of disasters like a cyclone or flood or even very man-made disasters or like maybe a fire broke out and type of things we do not need to know that what is my overall resources position and type of things, right.

So, it or and has a wide scale application in natural resource management, not only that agricultural, production estimation, so it is a huge data to be Chand and type of things, where which has a relations with space or a earth coordinate. And, in some of the cases

we will see later on it is not only space it has a relations with a space and time, how things changes; like climate change, how climate change is going on, how things are affecting and type of things, right; we will see.

So, there are application in smart city applications like finding as we are discussing that we do a day to day finding a particular path or between point a to point b or shortest path minimal time or sometimes with different other parameters and type of things So, it has a smart city. There are several aspects of smart city applications which helps us which needs space time information. One is transportation challenges, there are challenges of maintaining public healths or even how overall say pollution map, how things are evolving over time and type of things all those things are there, whether I can make smart city inclusion cities where this sort of informations become a major aspect or maintaining this type of information is a challenge.

And of course, we kept this inclusion transportation as a separate head because the transportation has a huge applications. We will see some use cases how things can be there. We are used to several type of aspects like with car sharing, different type of finding a minimum path and type of things. And not only that whether I can do some of the intelligent type of, whether there is the traffic lighting can be making intelligent or making made more amicable with the cloud. And, you see the one of the major problem with large cities are there is a volatile population in huge amount of people or population come push into the city during the morning hours which goes out of the city or a big portion goes out of the city in the evening hours.

Now, for this the whole dynamics go for a hay way, right or we need to learn rather there are work which tries to learn a typical dynamics of the cities, right, which may help in overall traffic management in a particular city. So that means, intelligent transportation is one aspects, environment and climate change monitoring is another aspects which are extensively important and bring this. And there are development planning activities, like say government things that have different type of development planning activities or schemes, how they are performing because there is a public money which is being invested, how they are performing, how many how much of the population is getting benefited and how to some sort of quote, unquote or audit them that what is the how much has been invested, and how what is the overall return on the things, not only on the work front, on the socio economic front and other things.

So, what we see that these aspects it is not only data, data related to earth comes into picture. So, whether it is necessary or it makes more sense if we try to loop the data in that perspective, right. How do I represent? Like how do I represent say this campus, IIT campus on a map, how do I analyze the map, how I can basically mine a map or such data, right. I want to mine the data for last say 10 years, 20 years and find out and try to make a predictive or forecasting model to see that what will happen next. So, definitely it involves acquisition analysis of data related to earth commonly refer references geospatial data set. So, spatial data and geo is primarily related to the earth.

As I was telling the spatial data can be anything on the space, but if we talk about more on our are, so these geospatial data says. And we will see there is a multidisciplinary activity. It requires several discipline. First of all if we if we look at any portion of the earth. So, if you see what other things are there if we commonly we say that how the land use is there, right, say consider our IIT campus or any campus or any city.

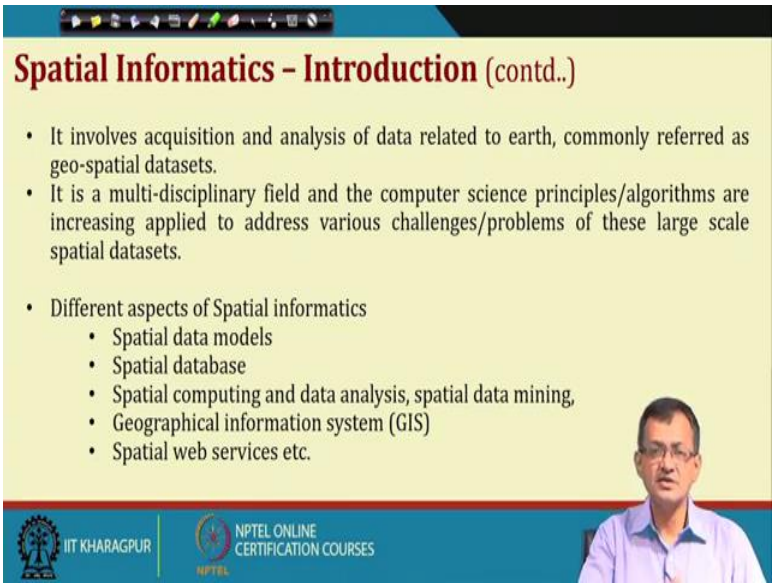
So, what one is that road network, right which may be a transportation which is under maybe a separate department or under civil engineering. There are things which are what sort of visitations are there, what sort of build up areas, what sort of soil structure, what is the population map, and anything like how what is the air pollution and so and so forth.

In other sense if you see there are different domain coming into picture what they are pinning to is a to a particular location, right. It is a IIT campus, this portion of the IIT campus or this is a city in Delhi or Kolkata, this portion of the city area, right. So, that is it is a multidisciplinary. Where, if you see computer science principles algorithms are increasingly applied to address various challenges and problem so of this type of large scale data sets. It is a huge data sets, it is enormous data sets. So, this these are where we computer science, ITE related expertise needs to be applied to find out meaningful information out of this data set, right.

So, here we come into picture. So, what the in this course we will try to look at those aspects, right, what are the aspects, how data need to be modeled, how to model a campus, how do I query something; whether our standard database query system will work or whether we require a some sort of a spatial query type of engine, whether how to mine this data. So, representation, acquisition, represent mining things are what we try to

look at. We will see that there are other several different challenges which come into coming into play which requires this our so called CS knowledge or CS different principles need to be applied.

(Refer Slide Time: 11:43)



So, that we talked about. So, there are different aspects of spatial informatics which some of them or several of them will be trying to address in this particular course. One is spatial data models, right how do I model a spatial data, right. Modeling a non-spatial, so called non-spatial things may be one aspect that is there is also important and that that to here some of you or many of you may be knowing. But what whether there is facility of modeling spatial data, right.

So, and whether there is how this spatial database looks like we are used to that DBMS or RDMS relational database models which are widely used whether the spatial DBMS how they differ from this our standardized database or what extra is they are into the thing, right. Then of course, spatial computing, right; I want to find out say what is the if there is a say this particular river is flooded how much area is likely to be affected, how much crop is likely to be inundated, right, how what should be my plan of relocation this population, how much population is there; that means, there is a good amount of computing is necessary. One is that querying and retrieving the data and computing the things and.

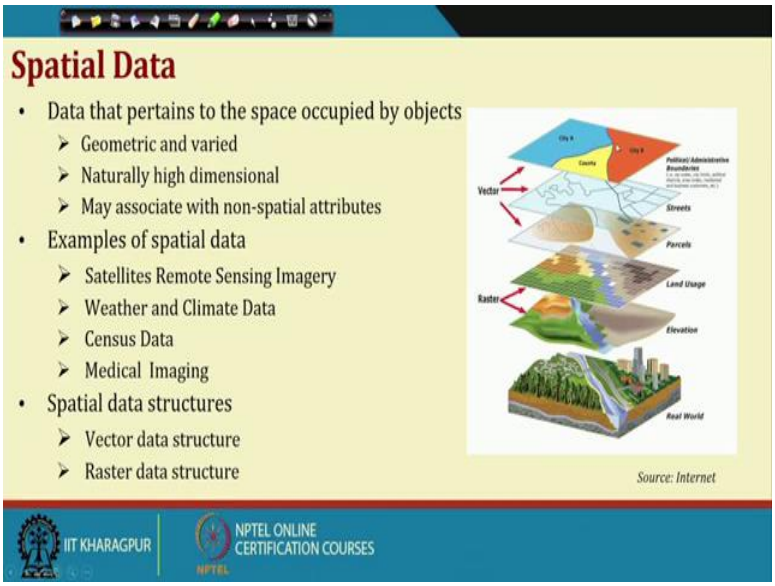
So, one is computing analysis of the data like I want to do a multi layer modeling of the different aspects and find out a analysis. And of course, mining the data, right with our standard data mining techniques definitely are used, but whether we require some special attention to the spatial aspects; so, spatial computing, spatial mining, spatial data analysis this becoming in popular. Rather, these days we see the word which is the jargon or which is more coming up in compressing all those things or maybe interchangeably used in number of places or number of cases is that spatial data science, right.

How to address this huge large scale data into the things? One another aspects of this is geographical or geographic information systems, right. So, it is the information system like any other information system, right every information which look into the geography of the earth and allow you to work on it, right. It takes care that the data said data database models are there, right and with that we can have a geo geographic information system to find out different queries or finding or find out different inference on the data sets, right. We will look into those aspects also.

Another upcoming area for quite few years is the spatial web services, right. So, whether I can give as a services, we are used to that, right. I want to find a location a to b, I do not have download the data set, look into the map and find out, right, previously say long back you need to buy the map and then find out which place to what. Now, we have a lot of dynamic things, right.

So, what I am what we are looking at is more of a web service related things, right web services, right. We are used to that. We are used to say business services or say e-portal, like your e-market places at those are services, right or buying tickets in airlines or railways etcetera those are services. So, whether it is possible to have specialized. So, yes, it is possible. We are knowingly unknowingly using it and that will be a good aspect to look at. Like, I have a services defined then when I require I call the services and look into the things, right. So, these aspects we will be looking into this particular course.

(Refer Slide Time: 15:56)



So, if we look at these data sets. So, this was my this down thing is my real word, right. It has a it has city county t, city a, city b county etcetera region wise which are more of a administrative boundaries which are not there actually out here, it is nobody has written right. We say this is a state of West Bengal, is a state of Maharashtra, or state of Bihar, state of Odisha these are administrative boundaries. So, this is a layer.

So, there is a layer which I can take out this elevation, like elevation you know that at which portion what is the elevation. We usually calculate with respect to the mean sea level, right. There can be a layer which is land uses. How the land is used? This portion road is there, that is a building is there or somewhere vegetation is there somewhere water body is there. So, this is how the land use a land is land uses or sometimes land use and land cover.

So, how is land use land is used and cover. There are things called parcels streets and there are road networks also. So, there are what we say though we have this is the real world data, I can segregate these different layers, right, use some of some or more of the layers to do some computation, spatial computation. Like, I want to increase the width of a particular highway say national highway, say national highway 6 we want to from 6 lengths to 8 length or 8 length to 12 length you want to make it. Now, the thing is that what should be the cotton field, like how much how much area I require. So, that is the estimation.

So, far for that I require a road network, I require a that what is the land use land cover of the things and I may require some other aspects of the things, right. Like if it is educational institute, a school, etcetera I may not be able to break it, so I may have a d two that combined may have different type of policies. So, those things will be there. You see, so I need to do a multilayer modeling and also based on different other constraint etcetera.

So, this if I can extract this thing this multi layer and store somewhere it may be possible for people or possible for decision makers to work on those things. So, so what we are you trying to this is I am trying to create a information base with a back end database, proper database models and giving them connector as a spatial say services by which they connect and get the services, right or they in the sense we will be working on those things or it can be hooked into other system to work on, right.

So, data that pertains to space occupied by objects, so geometric, geometric and varied. So, there are geometric aspect of the thing. Naturally it is very high dimensional. So, different aspects like you see there are multi layers and type of things, may associate with non-spatial attributes, right. I say this is a road this is the name of the road. So, there is a non-spatial attribute, right. Nowhere on the road it is written in the name of the things, right. So, there can be or this is a region there is a name of the region that number of say population count, it is difficult that from the things we can see that how many, what is the population density of the thing. So, all is in density. So, there are aspects which has some different attributes.

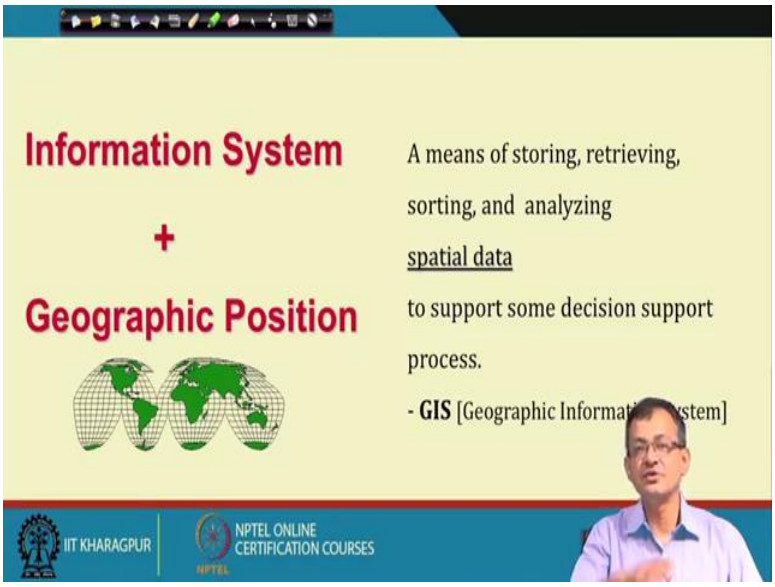
So, examples of spatial data one is the satellite remote sensing imagery. So, we get a huge amount of data from the satellite remote sensing imagery our own IRS satellite in the any mode sensing satellite which captures data very frequently. So, there is a huge data set. There are weather and climate data which comes from different sources meteorological offices like IMDs etcetera and type of things made offices. There are sensors datas which are collected at different point of time that is the particular authority, there are there it can even talk about the spatial, right can be medical imagery imaging data also, right medical imaging of the body where the it is also a coordinate.

So, it is idea is that if I can put on a basically a graph paper if I can find out that where this position is that at that position what is the thing, right. So, at in this particular course

we will be dealing most primarily on geospatial data sets data related to that not on the medical imaging etcetera.

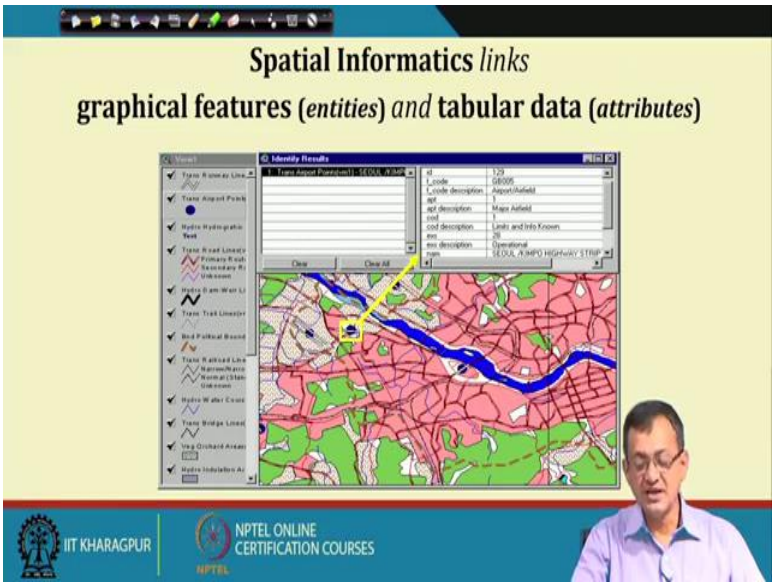
Now, broadly see there are two terms vector and raster was coming up, we will come to that, many of you may be knowing. So, broadly the data structure at a gross level is two thing, one is the vector data sets and is a vector or is a raster data structure. We will see that.

(Refer Slide Time: 21:04)



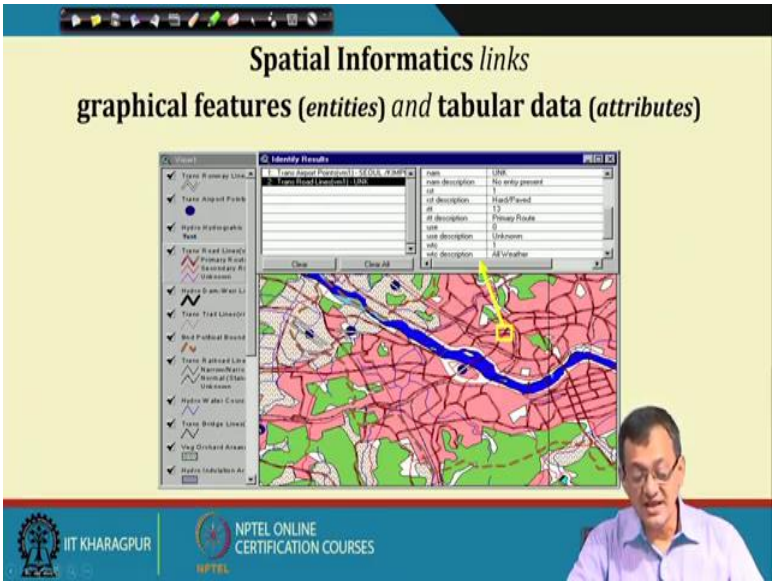
Now, before just what we are discussing, one is the information and the geographic position, so it allows us to handle another separate set of information system is geographic information systems. So, we will come to that aspects, already we were discussing, right. Mean of storing, retrieving, sorting, analyzing spatial data to support some decent or sorry there are two supports, so decision making processes.

(Refer Slide Time: 21:31)



So, this spatial data again I must say that many of these images are taken from different intended resources a must of the case many of the cases I try to refer, otherwise it is absolutely for academic purpose we have I am using it. So, you see this is a data which has a attribute data, right which is the different type of aspects of the things, right. So, this is data which is different description.

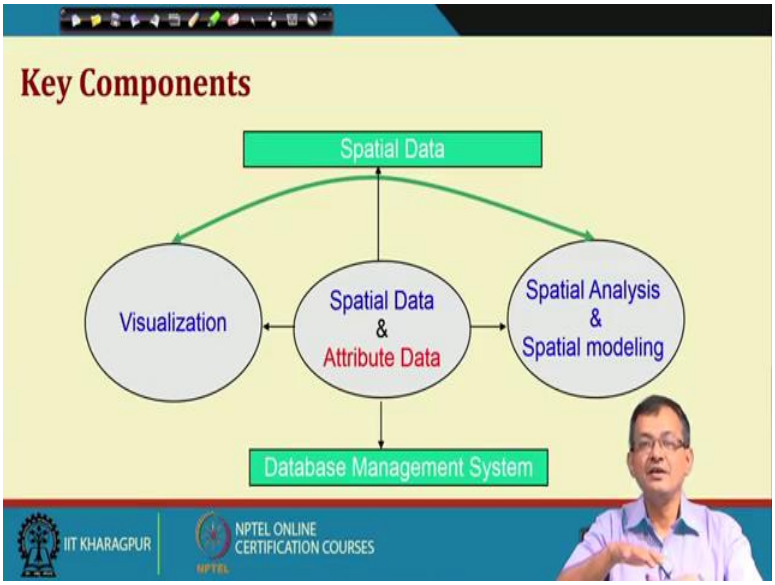
(Refer Slide Time: 21:58)



So, this is a image, but which refers this is a space spatial image or spatial data which refers to different attributes of the aspects of the things. Like it says that hard or paved

way; so, it is not written on the things, right it is a attribute to the thing. We will come to those aspects.

(Refer Slide Time: 22:19)



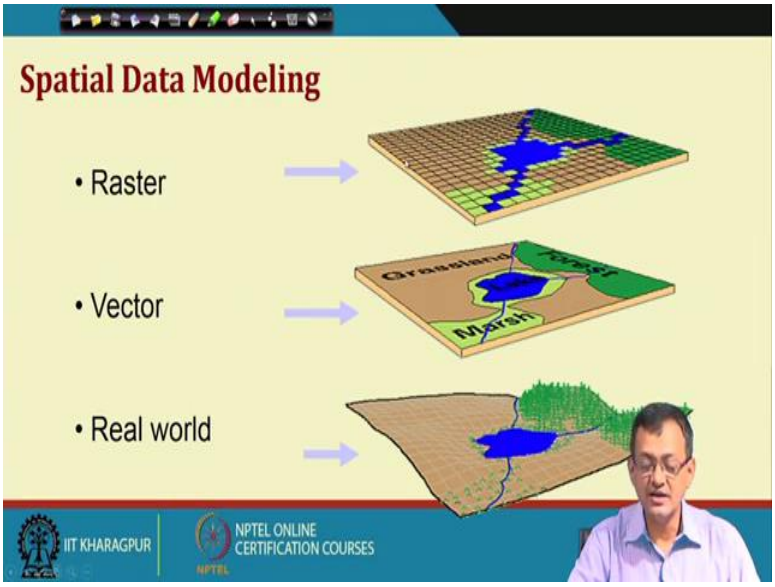
So, if we look at the spatial data one end is definitely the database management system. Unless there is a strong database management system handling data is a problem, whether it is spatial non-spatial anything, we require a some database management system, right. So, in order to handle that we require the spatial data attribute data.

So, spatial data is the data which is on the it can be seen or very gross detailing. If I go little up in the on the over the earth surface if we look at what is there take a picture what we see is the spatial data, right, right. It can be from the satellite, it can be from a drone, it can be from a pares say a balloon where by which it can take in photo graph, but what we see on the things, right. There are different challenges, we will come to those challenges so, spatial data and the attributes.

So, attribute is what is the meaning of the data. This is a road, this road is say IIT main road, it is not written on the when you take a picture you have to give you attribute other things, right. Even when we take the picture of our things a group picture then we have to say this person is this, this person is this is this is attribute of the things we say there are if we on the broadcase these are ladies, these are gents this is putting attributes the things, right. So, those are to be marked.

So, one side we require a spatial modeling and the aspects of the things other side finally, we need to visualize that. The visualization can be different aspects. Like the way I need to see for my purpose you may want to see the data in a different purpose all together. I am a transportation person, I do not block at the road in more detail, other things are there. You are somebody may be a say climatologist, you want to look at the climatic or weather conditions, right some maybe water resources part, look at the water aspects of the things. So, the same data sets can be visualized in different aspects.

(Refer Slide Time: 24:27)

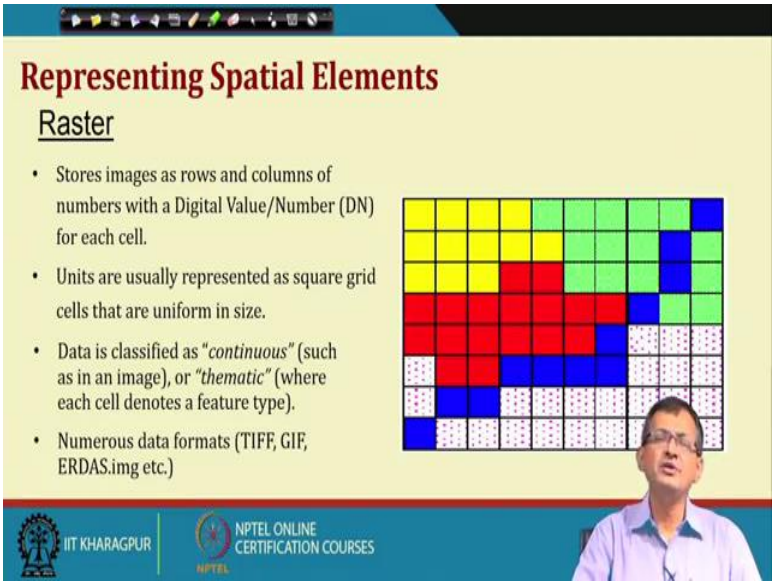


So, if we come back try to come back to the same thing. So, one is that this is our real world we have seen that type of picture. This is what we say a vector data we will come to that what is there, so vector data. This is a raster representation of the same data, same real world raster. So, this raster is as if putting a graph paper on the things and then coloring that this is this, this is this, this is this, right or labeling giving a number, number 10 is water body, number 4 is this, giving a digital number to the thing. So, sometimes you say dn values to the things.

Raster is more represented by what the algorithm if the 2D space we say that 3 things can represent every or all objects like a point, a line or polyline, a polygon, right. So, it presents the things. So, instead of doing all these things that if this is the number is 3 3 3 3, I say that this is a polygon of feature this grassland, right, this is a polygon of lake, this

is a polygon of marshy land and type of things, right. So, different advantages are there for different type of applications, so both raster and vector type of things.

(Refer Slide Time: 25:42)

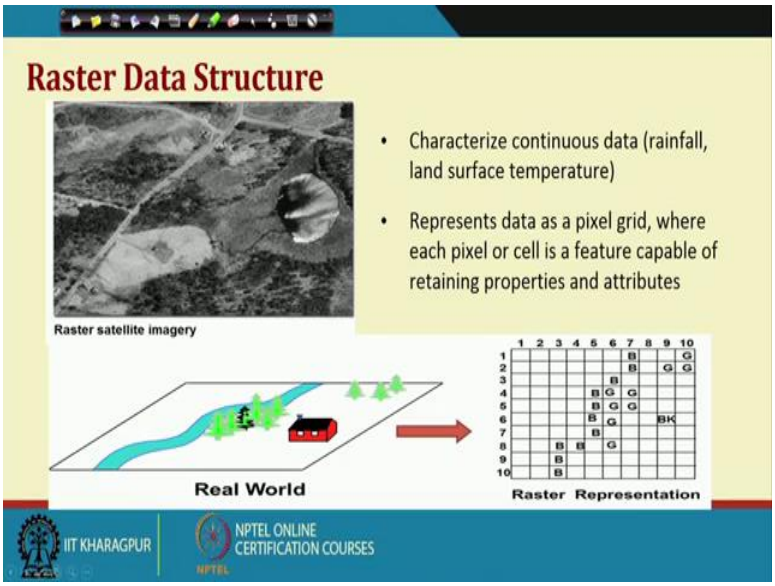


So, as we are taking there is a representation representing spatial element. This is the raster represent. Stored images as rows and columns as we the direct example is the graph paper with a digital value or number, right. So, this yellow is 4, green is 5, red is 1 and type of thing, so it distinguishes. Units are usually represented a square grid that are uniform in size that is the standard representation. It can be variable, not square rectangle etcetera, as of now with things are square grid and type of things.

Data is classified as continuous, there is a continuous stone of data such as and in an image, when you take a photograph it is more of a raster, right it is a continuous stone of data. We even take a photograph of a natural image, it is a continuous stone of the data or sometimes called a thematic where each cell denotes a feature type, right I guess a thematic layers. I say I i for a particular theme I say soil layer of IIT, Kharagpur is a thematic layer in the theme is soil.

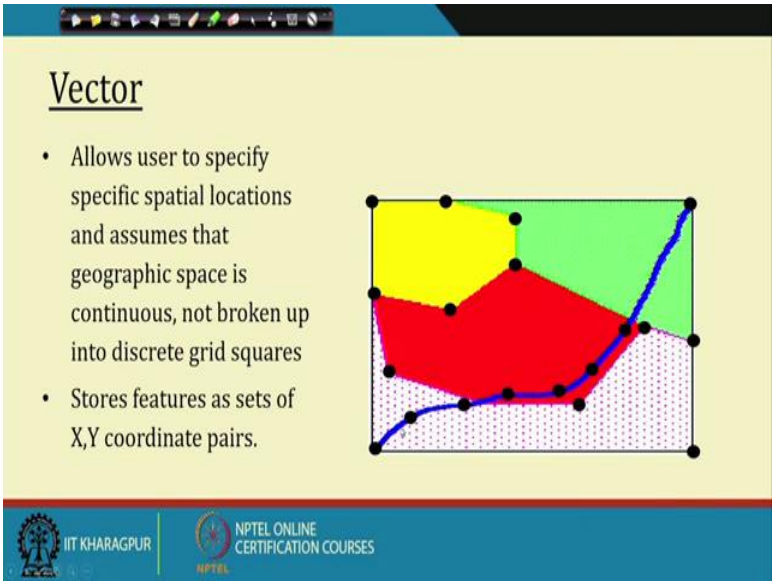
I say that overall say land use land cover or road networks that is a transportation network, it is a thematic layer. I say build up one layer is a thematic layer where the buildings are there and type of things. So, it is a themes. And there can be numerous data format TIGG, GIF some type of things which are open source there can be some proprietary formats also.

(Refer Slide Time: 27:23)



And, if this is the particular image so, we say that is a photograph. So, this is a raster image characterized continuous data, rainfall, land surface temperature. Represent data as a pixel grid, where each data cells at feature capable of retaining the property is an attribute. Like we have seen this one, at each pixel or this square we say each pixel grid, right. And, so is a real world I can represent by this blue, green this can have a values and type of things.

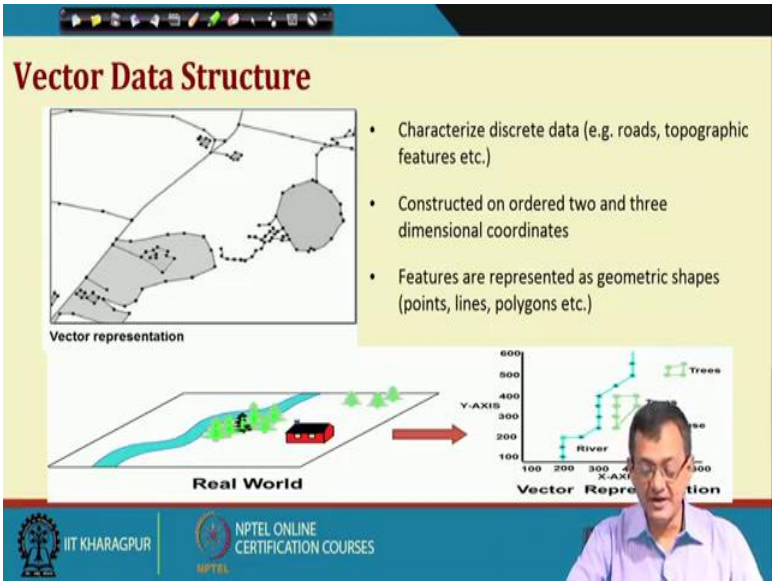
(Refer Slide Time: 27:53)



On the other hand, in case of vector we represent by these things. Instead of storing every row and column we store this particular polygon as this one, this polygon as this one and so and so forth, right. So, this is this in case of a vector data set. So, in other sense say I want to store this line I store this point 1, point 2, this point 4 these xy coordinates, right. If it is on the earth surface the base known xy coordinates what we know we have learned from the school days is the lat-long, right. I can have other representation. If a graph paper we say 3 4, 3.4 and 4.4 and if we based on the granularity, right.

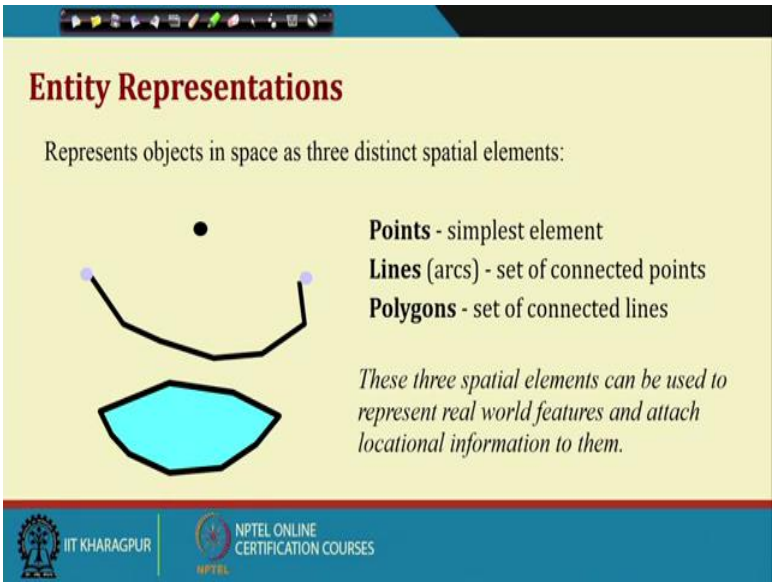
So, allows user to specify, spatial location assume that the geographic space is continuous not broken up into discrete grid square, right. Stores feature as a set of xy coordinates, so it is a series of. In a polygon I say there is a series of xy, where the starting xy and the ending xy are same, right. In a polyline it is a see again a series of xy point. If you have point, only one xy point. So, what is what is what we argue what is argued that in a vector represented if we can represent point, polygon and polyline we can represent any object.

(Refer Slide Time: 29:17)



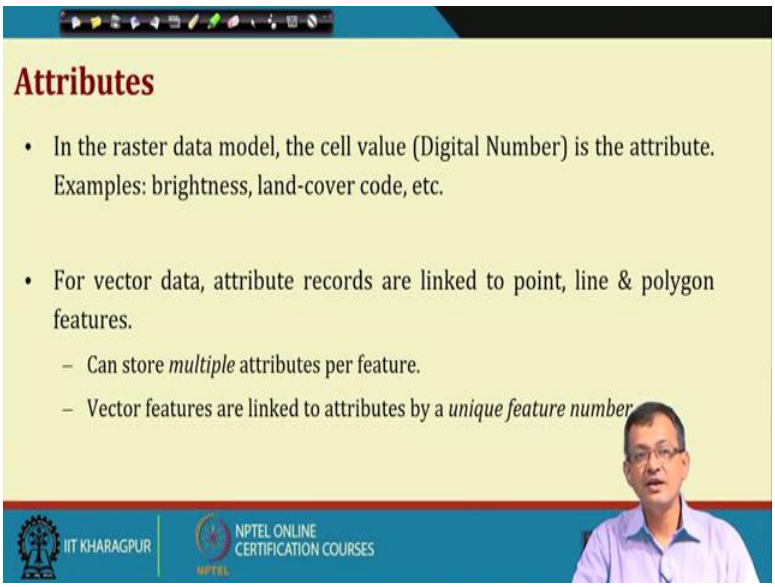
So, these are two representations like vector data set, same data if you just see how we represented here. This, same data set if it is in the vector it is represented by this.

(Refer Slide Time: 29:31)



So, as you are discussing, so there are represent objects by points, simplest of the element lines or arcs or polylines and polygons set up this. These 3 elements can be used to represent real world feature and attach location information. I say this line is a IIT street road, this point is a ATM, this polygon is a water body. So, this is the attribute can be attached.

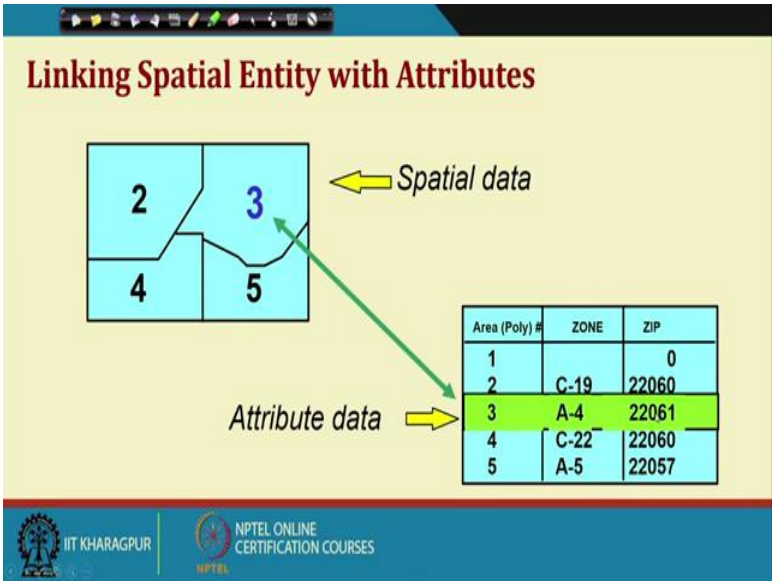
(Refer Slide Time: 29:59)



So, that is the role of attributes in a raster data model. The sale value region is the attribute which can attach to other files, brightness, land cover, etcetera. For vector data,

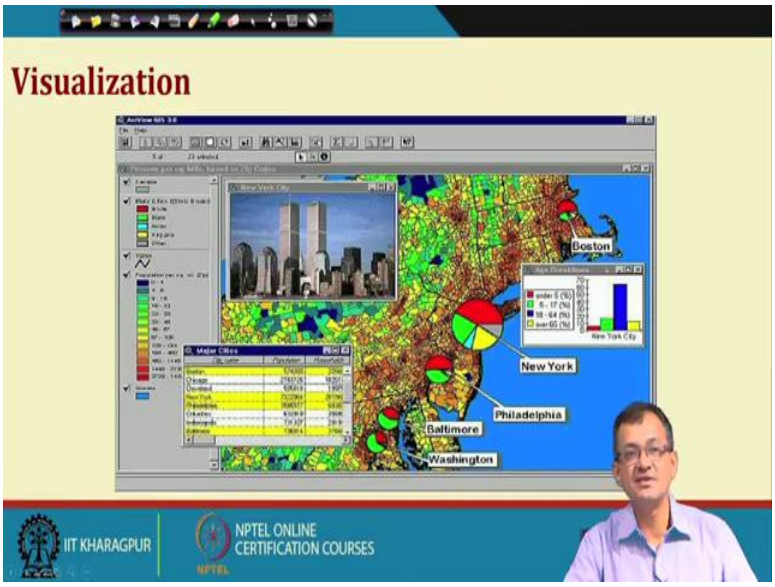
attribute records link to the point, polyline and polygons can store simple attributes as feature. Vector are linked to attributes of unique feature number.

(Refer Slide Time: 30:18)



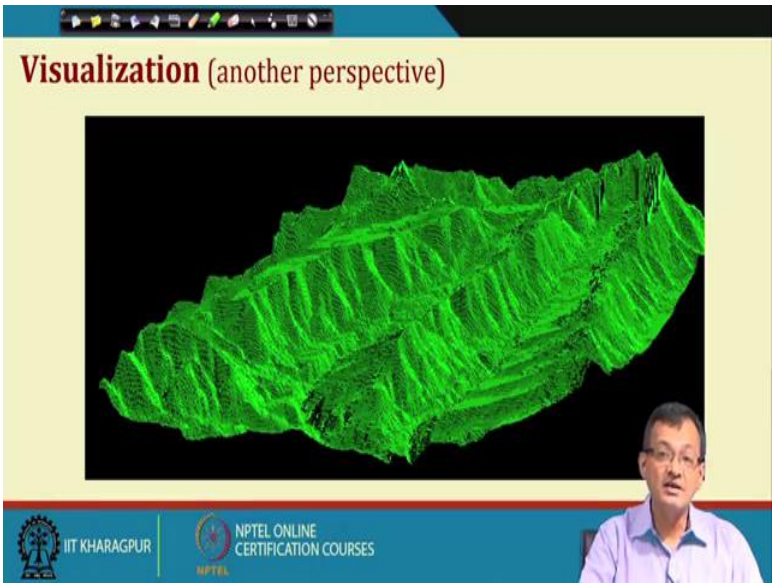
And, we can say that these are the different attribute representation, right. So, this attribute this polygon is polygon number 3, which is zone a 4 with some zip code, right. So, I have a separate table which is attribute data, right which is linked with a this spatial representation. So, we will see that how. So, this way we try to link the things.

(Refer Slide Time: 30:47)



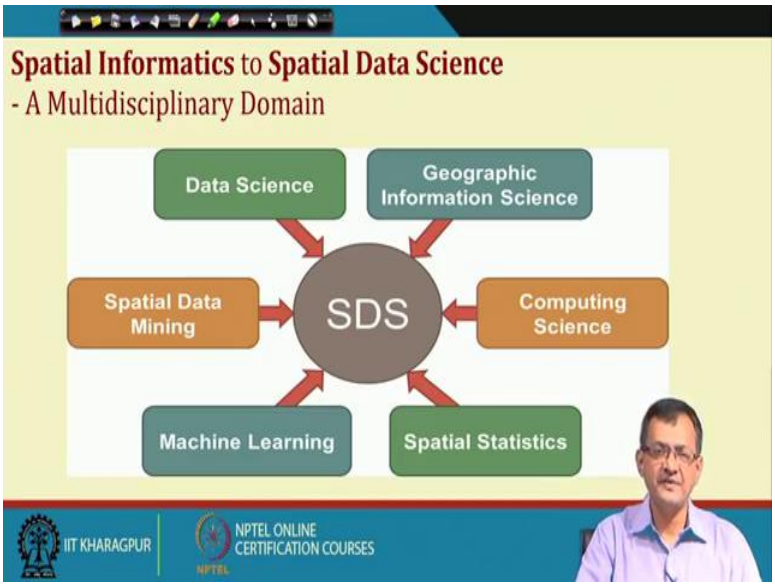
This is this is another aspects which we will cover at some somewhere end of the course is the visualize. How do I visualize this data? So, one visualization is this is again this is taken from internet. So, we can have this data with different type of visualization.

(Refer Slide Time: 31:07)



You can say that I want to visualize the digital terrain model; it is a 3D model of the things. The same region I can visualize a terrain model. So, it depend that what is the use of the data, right how do I use the data, right.

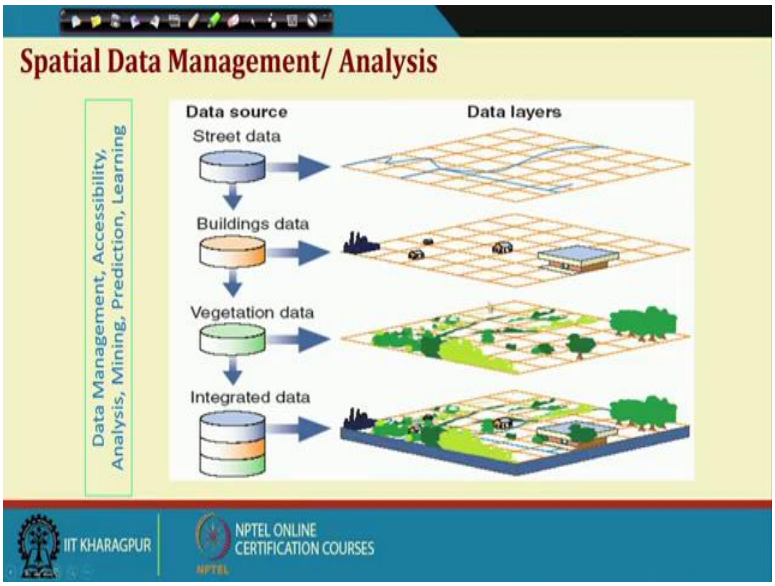
(Refer Slide Time: 31:24)



And we will deal in different aspects I thought that in the introduction, there is it will be note it will be good to mention. So, there is a concept of now more spatial informatics or is more what we say looked as a spatial data science, right. So, there are argument, whether the same thing etcetera, but we are not going to those means classes or arguments, but we would like to see it in the it is a more of a data science problem, right.

It is a multidisciplinary problem, where we require definitely computer science, spatial statistics, machine learning, spatial data mining, so these are all CS related stuff along with that which is missing here we require a domain expert part. A transportation problem has to be handled by a transportation person, right a soil by a soil scientist. So, that domain expert needs to be embedded into the things to get a meaningful into information into things. So, in this course may not be able to deal with all the aspects, but we will try to look at some of these things which helps us in how this multidisciplinary aspects can help in doing solving several problems which is related to or several challenges related to our day to day issues.

(Refer Slide Time: 32:45)



So, finally, but again coming to that picture back. So, this is our integrated data. We have different type of layers, right and we can take one or more layer for managing, analyzing the data. So, data management, accessibility, analysis, mining, prediction learning and dot dot dot which need to be there. And if you if you by you will go gradually, but you

understand that these data set has a little bit or some a behavioral different behavior than our normal standard vanilla data sets what we deal with, right.

So, how say even how do I say that a two polygons touching each other, what is the definition of overlap, right. Overlapping one student with one student let us say has what meaning has to be looked into, but here are that how two polygons overlap, whether this road is overlapping with something. So, there are different aspects of things what we will see. So, there are different topological relationship etcetera which come in to fair play where we require this sort of handling of these data sets.

So, what we see here this is a other way around. So, if I have the separate layers etcetera then I can I can make this integrated data, right or in other sense if I have different layers and take a decision things, like I want a prediction model I differ if I have a spatial temporal data set I predict the some aspects etcetera. Then, I want to look at to see the down the line 10 years, 20 years down the line how this overall IIT campus is likely to be there. There are there are a lot of ifs and buts into the things, but I can do a trend analysis and type of things looking at the different aspects. These are being extensively used for metallurgical datasets with looking at different aspects of climate change phenomena and several mechanisms, right.

So, in before ending today's talk, so what we are in this course we try to look at this different aspect like as we mentioned in the beginning how to model this data, what are spatial database, what are the query problems, whether there is a spatial expects of some of the things like weather if I say that spatial networks whether there is a spatial aspects or spatial dealing for that. How to mine this data and how to decimate this data, how people can use this data whether the spatial web services even when if time permits we will look at the spatial clouds and type of things, how this data can be can be harvested and can be used at by different domain and by different domain experts.

So, with this let me conclude today's talk, and we will continue in our next lecture.

Thank you.