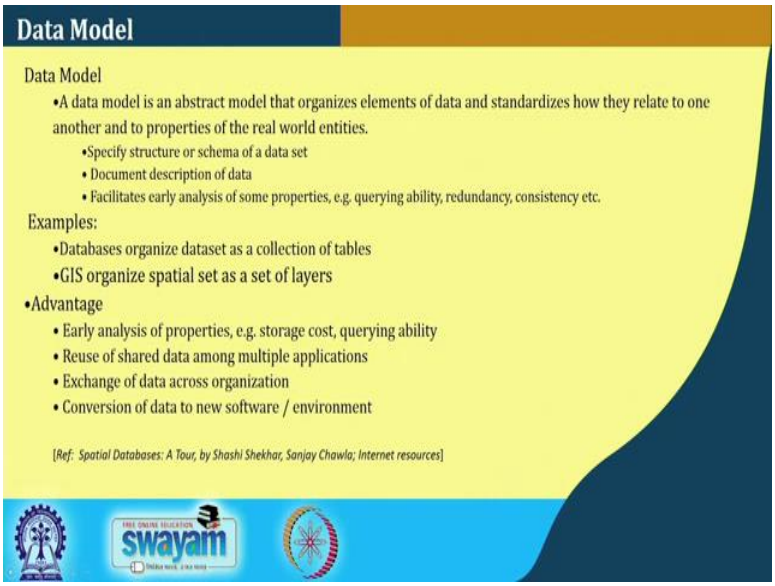
Spatial Informatics
Prof. Soumya K. Ghosh
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 02
Spatial Data Models (1)

Hi, today we will be discussing one very important aspect of this Spatial Informatics which is data models or so to say Spatial Data Models. We will have couple of a lectures on this and it is pretty important to understand that what does it mean and what we will see later on that most of the most of our this spatial query, spatial means decision making, data mining etcetera are primarily driven or why or to say that the foundation is this data models right.

So, it is important to understand these spatial data models. And if those who have already done data base courses you we you understand that this in our non spatial also called our standard database also, these data models are pretty important that is the major thing for designing any database or any information system. Similarly, for this sort of thing spatial information or geospatial information system also we require these data models should be robust that type of things.

(Refer Slide Time: 01:39)



So, primarily we will be looking at some of these aspects of our topic. So, when we talk about data model, so if we look at dictionary meaning or there are different internet

resources like, if you see like wiki etcetera. So, a data model is an abstract model that organizes elements of data and standardizes how they relate to one another and to the properties of the real world entities, very encompassing definition right. Its a abstract model right that organizations elements of the data and standardization how they relate to one another and the properties of the real world entities. Interestingly, this we are not talking directly about the content of that data, we are talking about that how they are organize, how they are structured and the relationship.

So, if I can define these data entities and their relationship appropriately represent them and I can do other things along we have means over this right. Like a specify structure or schema of a data set right what is the schema or structure of a data set? Document description of a data set how they are what are the different what different component means and so and so forth, facilitates early analysis of some properties right.

I can have a early analysis of querying ability redundancy consistency, like before say I want to have a student database forget about spatial database or library database. So, what sort of query I can do? Whether I can search by author, search by book name, search by partial book name, search by say data of prints or reprints or search by say query by say publisher and type of things?

Now, see these need to be thought a priori once the data came into picture and if I after the data entered into the book enter into a library, then I cannot think of the things right. So, a priori I we have to think those things, there are other things like whether there is a redundant information I am storing, whether there is a consistency of the data which I am storing or another important aspect we will see that the whether I can have a estimate of the space right or estimate of the storage right.

Say I have so, many books etcetera I am planning, so these attributes I will be storing. So, how much space estimate? This type of things becomes more critical when we talk about spatial data specifically geospatial data, where the data node is pretty high right. So, in appropriate models may lead to improper query or inefficient queries lot of redundant data set which is wastage of data say wastage of space not only space you at times you may there will be some sort of a if there is a redundancy, then you have to maintain appropriate consistency of the data set also right. So, all these standard database related problems come into play here.

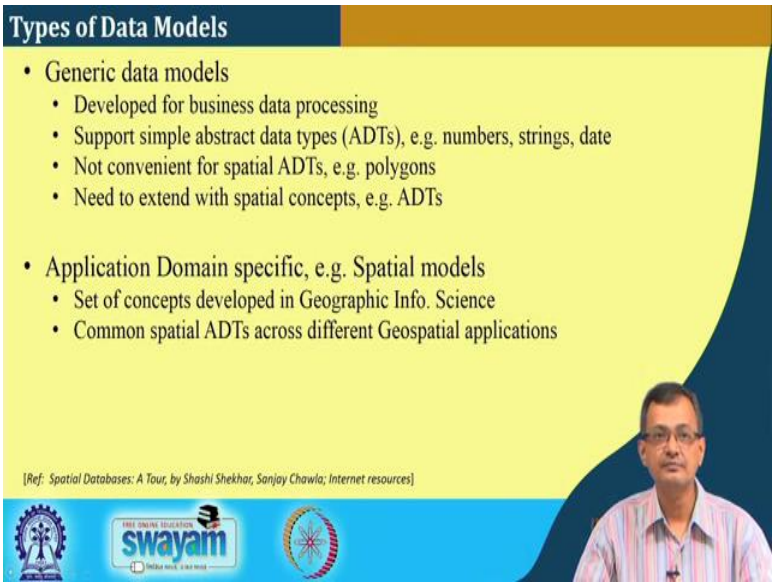
So, there are examples that you know that standard database is organized that it has is a collection of tables right or what we popularly called relations or relational databases. In case of a geospatial databases or GIS databases organizes spatial data has a set of layers; layers of information we will come to that. So, there are definitely there are different advantages early analysis of properties, storage cost, query ability etcetera. Reuse of the shared data among multiple applications right.

If I have a shared data I can use the multiple application like as we are discussing the introduction let us say there is a road network, it is used for say for railway road authority for expansion etcetera. A common people for finding a shortage route there may be or for some other city planner to see that the access road or industry setting up industry the access to the which are the roads etcetera. So, the same data set is access queried for different purposes.

Exchange of data across organizations, interoperability between organizations right there is a important aspects. Conversion of data into new software and environment that is another. If the model is known then if my software tool changes from say software x to software y, then what I need that first of need to know that what is the model whether that is supported by this new software y if it is there then there are changing will be there.

So, these are the some of the very important aspects which we need to understand very clearly for the our data models right. These are true for our standard databases and data models and these things are true for spatial databases also, along with that the spatial databases comes with different other challenges what we will see.

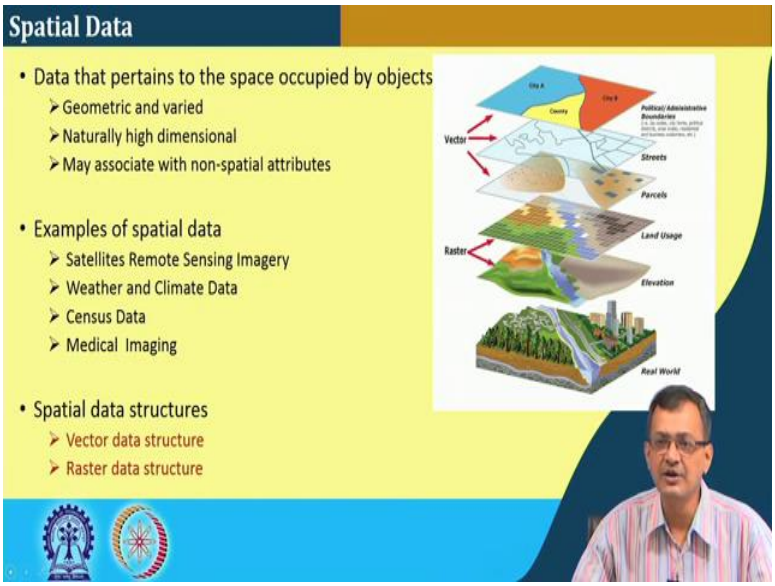
(Refer Slide Time: 06:37)



So, there are if we look at the types of data models. So, there are some generic data models they develop for business for data processing maybe one of the things. Support simple Abstract Data Types or ADTs like numbers, string, date etcetera. Non convenient for spatial data types like polygons, how do I represent a polygon, how do I represent a line, how do I represent a point? So, its not a simple data types need to extend to the spatial content. So, this ADTs etcetera, so I have to go for a spatial ADTs which you can support. And there can be different domain specific things like spatial data models.

So, it has a some domain to address like set of concepts developed in the geographic information science. Common spatial ADTs across different geospatial applications can be there. So, I have a one abstract data type which can be used by different applications for things. So, those are the things these are the challenges we are extension of our standard knowledge of database and data models to this spatial data models.

(Refer Slide Time: 07:49)



So, before we go a little bit on the spatial data models, well let us see that what is spatial data itself right already we discussed something on introduction, but its a data that pertain to space occupied by the objects. In our case for this particular course, we will be taking about the geospatial data or what we say that it is our space is earth surface primarily we will consider as a 2D phenomenon later on let us see that whether we can have one or two things of 3D.

But primarily a 2D space of earth surface, so any when you talk a spatial data it is related to something on the earth right. So, geometric and varied type of things may be there naturally high dimension there are a lot of things may associate with non spatial attributes also. May associate with non spatial attributes, I wrote road as a geometric right and he has non spatial attribute the name of the road or there can be other things like particular type of road metallic non metallic etcetera which add on the things right, so there can be non spatial.

Similarly a region can as a name has an other type of things as say particular see if I consider a district; district has a name, district as a say administrative authority like district magistrate it can have other blocks etcetera etcetera right name of the things. So, some spatial along with some non spatial; that means, to see very flatly spatial means which we can see from the top right.

If I go above the earth surface or whatever I am seeing around the things is the spatial data or the things very crudely if I say and non spatial we cannot see right. I cannot see the name of the road from a drone or flying on a aircraft neither I can see the name of the things right district. Now, that I know that name of the district magistrate or something right or even we cannot see the population count etcetera.

So, that, but they are all related to some coordinate or set of coordinate systems right. So, essence spatial data one is very popular or very routinely done is the satellite remote sensing imagery with our own Indian remote sensing satellite, which are launched and maintained by ISRO we have a huge volume of data repeatedly collected of the same region.

It helps in different type of things like one major aspects is natural resource management, there can be agricultural applications and so and so forth there are various applications. There are another set of data weather and climate data some are collected through different sensors and type of things like in our country primarily IMD Indian Meteorological Department they are the custodian of erody of the data and you have different nowadays to census to collect at different type of things. There are census data which are collected through different mechanisms like census team goes around and collect the data.

These are again related to region, but not directly visible from the things, I cannot see going up on the sky by that what is the population density of the Kharagpur. With high resolution you can say this is a highly populated less populated and type of things, but say what is the population count of the Kharagpur you cannot see right, but that is, but they are related to some region right. So, there is a spatial context everywhere and if we exchange it not this geo spatial you can have a medical imaging also like if I consider the space as my body. So, there are different things which are in the thing, so that is also spatial.

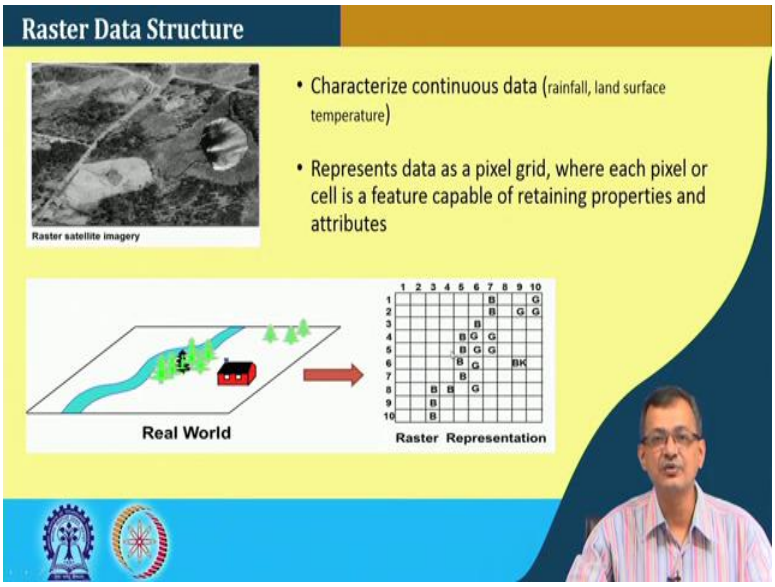
So, I can have universe in the universe some aspect, but for our discourse we restrict our discussion on the earth surface only right. So, most all our discussion things I will be there, but many of them can be applicable to other things as well right. So, if we look at the spatial data structure there are broadly two category; one we say vector data structure another is the raster data structure right. So, we will see that what they call. So, if you see

in this picture what we see before going to the raster vector, so this is our real world scenario right particular region of interest.

Now, this has different aspects; one is that this political and administrative boundary. So, some region city A, city B something county or something, so administrative boundary. Other can be something of different aspects of administrative boundary or there can be river, road etcetera which are drawn on the things right. So, there are land for sale so and so forth, we have some of the data like raster data like land use data I can have a raster data like, every pixel how the land is used right whether this is a forest, this is a agriculture, this is a road etcetera.

So, what is happening? I have that image of the data I take a image from the satellite or by something and then I put a gridded map over the things right; a grid over the things say graph paper transparent graph paper, then every box I say this is forest or I mark it something right. So, these things become a image or if I take this red its a raster a presentation right, say river. So, I point; point; point; point; point that blue color wherever the I see in the grid on the graph paper right that is a raster, but vector is I say I say that the river or the road is x 1 y 1, x 2 y 2, x 3 y 2 x x 3 y 3, x 4 y 4, x 5 x 5. So, these are the points and then I say this is a line and it has different attributes, if it is a attributes like it says that whether it is a full line road, name of the road etcetera are put on the things right. So, this is this these are the two way of representation of the data.

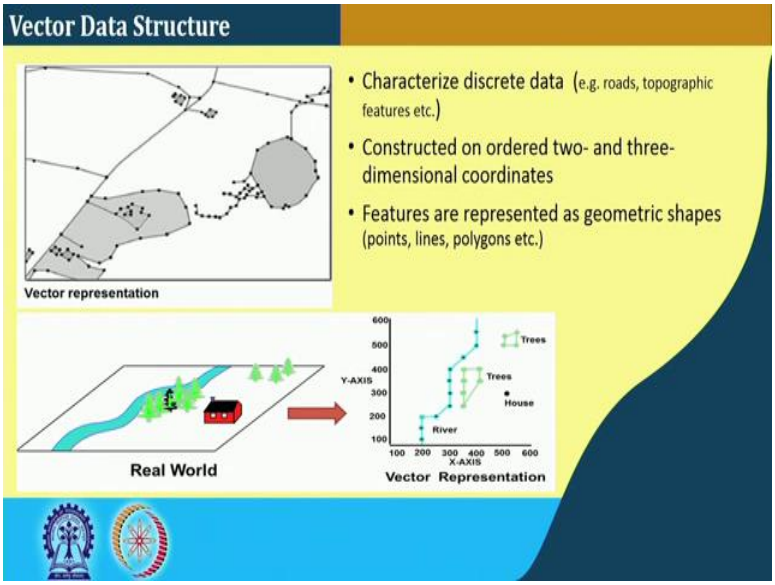
(Refer Slide Time: 14:01)



Set like as we are talking about if I have a grid like, so this was my typical a particular satellite image. So, characterized by the continuous data raster data structure represent data as a pixel grid as I was discussing where each pixel or the cell is feature capable of retaining properties and the attributes right. So, marking it that blue BBB may be the river structure, G is the greeneries and there are other things like it can be a some build up area which we can say as a big BK etcetera.

So, this grid give the raster type of s effect of the things right. The same thing I can represent by a vector characterized by discrete data set road, topographic feature etcetera constructed on ordered two or three dimensional coordinates. As of now again we restrict ourselves two to dimensional and later on see that if time permits we will look at some of the 3D aspects, but primarily it is a 2D two dimensional three dimensional coordinates and that you see now the previous represented and it was green like this or the river was like these represent on the graph paper.

(Refer Slide Time: 15:13)



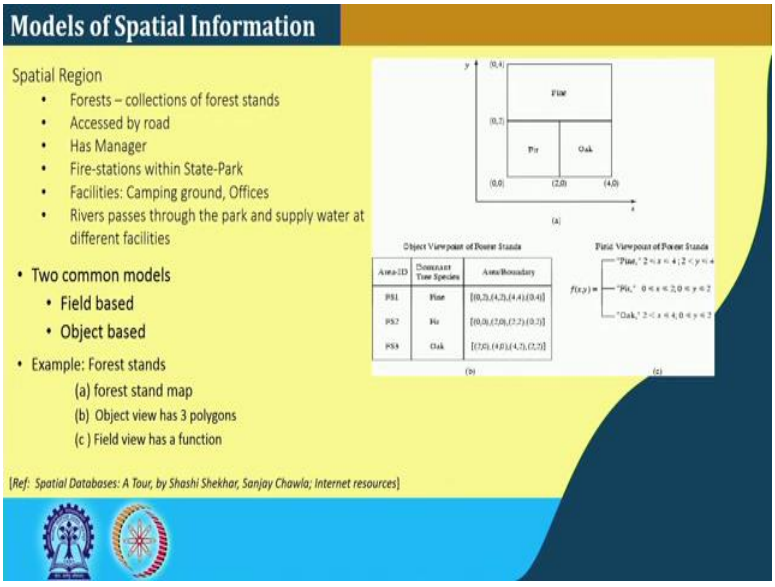
Here what we have a polyline right a series of XY coordinates which connects these things right. Similarly for the vegetation on that trees I have a polygon, so this polygon is you see, similarly for a house or a build up area things are like this. So, it what is required they are also in the raster also requires a coordinate system correct. So, if we think about our general coordinate system 00 and this type is I make it right.

Challenges come when you have different data mapped in different coordinate system, it is not it may happen that river is mapped by some organization say something hydrological department of India or something like that whereas, your road is mapped by a say highway authority right this your forest and etcetera are mapped by say forest department. Now, they are mapped or marked or put on the things in different formats and different scale how do I talk to each other.

So, that requires the standardization we will discuss later on that how things, but as of now we have two way of looking at it right; one is putting as if as a graph paper and marking it and take out the graph paper and it gives you a representation another things I do not keep like this I put a point like this right. So, again main majority our of our discussion will be on this vector type of datasets which we will see that how can we put on the data models and data bases, spatial data bases and so on and so forth.

So, construction on order two or three dimensional coordinates feature are represented as geometric shapes. So, three primary geometric shape point, polyline and polygon. So, these are the three geometric shapes.

(Refer Slide Time: 16:57)



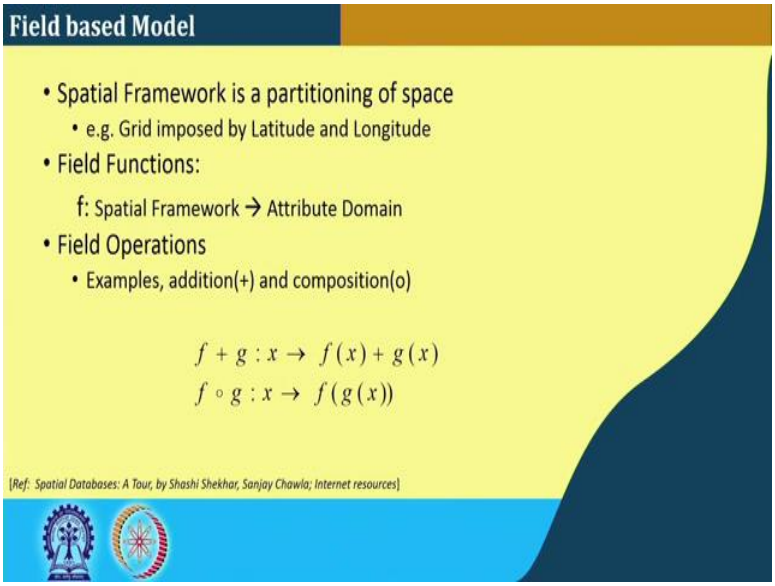
So, this is a example taken from that reference. So, this is a scenario of a particular spatial region something it is in the book it is mentioned as a forest park or something. So, it has some of the things like forest, collection of forest stands, accessed roads are there, it has a manager, fire stations within the state parks sorry state park, facilities

camping ground etcetera, rivers passes through the park and supply water at different things.

So, there are different features some of the features are like what we say forest lands. So, there are type 3 type of forest things its a what we say sort of artificial park you can say pine, far or oak type of threes are there and it has a access road to access the thing, it has some of the things like facilities like a campground offices and type of things has fire stations, a river also passes through this forest park right. So, looking at this data model part say we have two type of model; one is field based another is object based we will see.

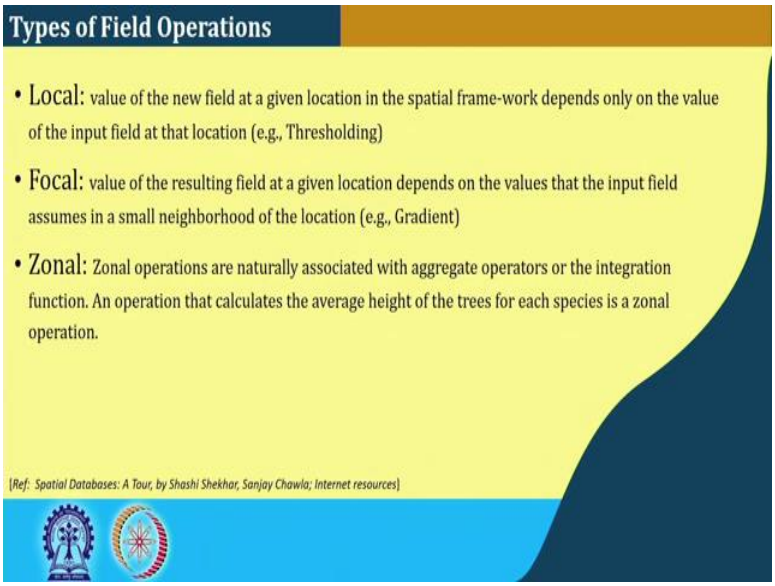
So, what we have forest stand map object view has three polygons and object view has 3 polygons like you are having 3 polygons pine, far, oak right. Field view as a function type of things we will see that how it happens.

(Refer Slide Time: 18:33)



So, field based model spatial framework is a partition on the space right. Grid imposed on the latitude by latitude longitude or any SRS or what we say Spatial Reference System. Field function, spatial framework to the attribute domain and field operations like addition, composition and type of things are there like I two function f and g if it is addition that implies some things right and then composition and so on and so forth.

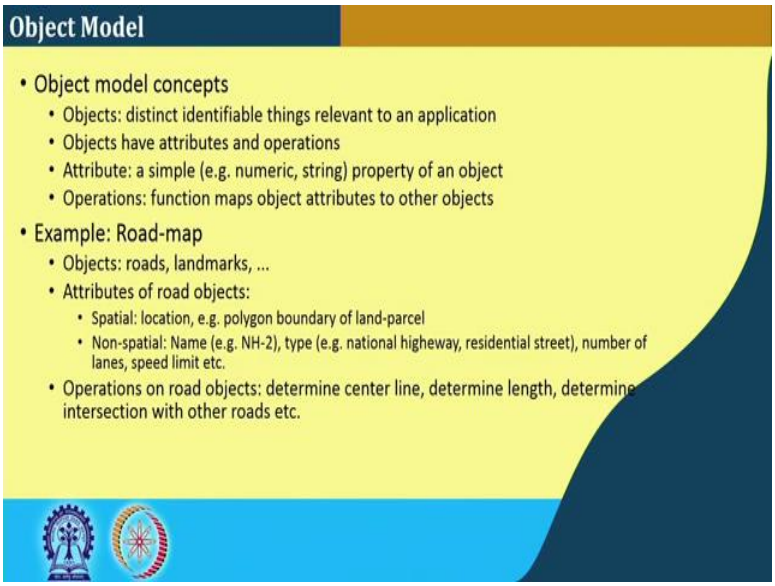
(Refer Slide Time: 19:01)



So, a field operations can be primarily three type; one is local, value of a new field is given location on the spatial frame work depends only on the value of the field which are at that location. So, there is some sort of a thresholding that is local phenomena all right. Like I say a particular elevation or something with respect to the height of the things is with respect to something which is local all right or even I can have different aspect like; one is focal.

A value of a resulting field at a given location depends on the value of the input fields assuming in small neighborhood of region, so that is a particular gradient of the things right how it is. So, what things are there that will be focal which is respect to some dependent on the things right. Zonal; zonal operations are naturally associated with aggregate operations or the integration function right, so it is a larger space or zonal. A operation that calculates average height of the trees of a each species in a zonal operations right. So, this three type of operations in field operations we are having local focal and zonal that can be what can be done.

(Refer Slide Time: 20:19)

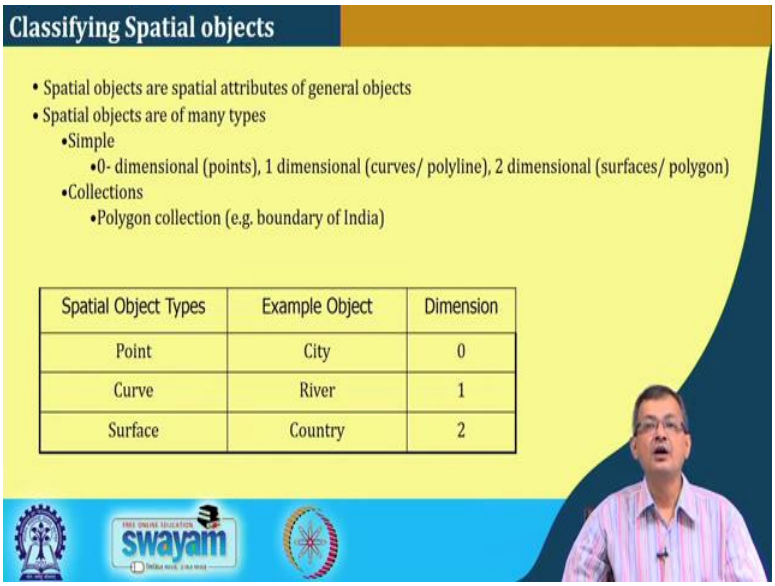


Then we have object models, so object model concept objects distinct identifiable things relevant to an application right and objects have attributes and operations right. Attributes a simple numeric properties, operations function maps object attributes to the other things right, so it is a object type of things. Like example is if we consider the road map, so object is road, landmarks are the objects right.

Attribute, spatial attributes or location polygon boundary of the land parcel etcetera can be the spatial attribute. Non spatial attribute name of the road like national highway 2, national highway 6 and type of things or national highway 34, type a whether it is a type of things then whether it is a residential street a residential area road or sorry street road or highway and type of things.

So, number of lanes, speed limits even and these are the non spatial attributes. So, these are the attribute attached to these particular road. An operation on the road objects determines centerline, determine the length of the things, determine intersection with other roads is these are the different operation on the road objects right. So, what we see here we have a different attribute sets and different operation on those attribute set right like, if you refer the other things also I can have different attribute says for different example in the different forest stand and then I can have different operations on those at on those particular objects.

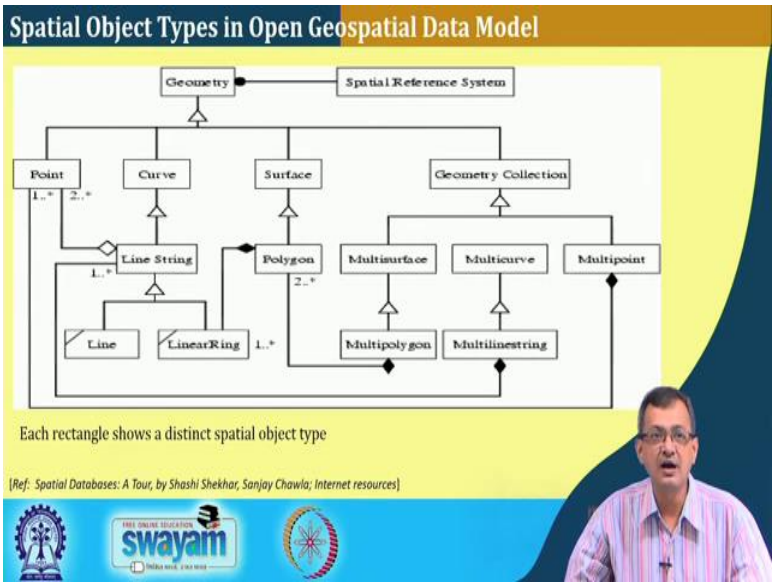
(Refer Slide Time: 22:03)



Now, if we look at broadly, so, what we have thought what we have discussing in a 2D space, we have three type of representation right. One is the point like with a 0 dimension curve or polyline with a dimension 1 and surface or polygon with a dimension 2 2D type of things.

So, spatial objects are spatial attributes of general objects right spatial objects are of many types simple. So, 0 dimension or points and dimension 1 dimension a curves or polyline 2 dimensional a surface or polygon. So, that there can be collection of the things like polygons collection, boundary of India if we want to construct here I can construct it from collection of the boundary of the states of India right. So, it is a collection which gives me the whole polygon right; so whole boundary of a country and type of things right. So, this is I can handle this way. So, this is in a 2D space we have three type of representation.

(Refer Slide Time: 23:21)



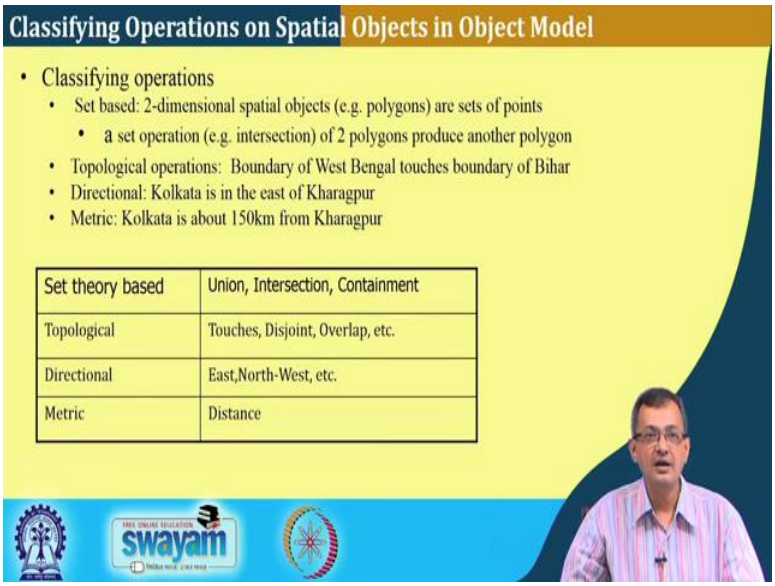
Now, if we have spatial objects in a say open geospatial data model or in a what, so to say spatial data model following open geospatial standard. So, what we have that there is a geometric class right. So, everybody has a geometric class like point, curve, surface these are the three primary things or I should say point curve or polyline, surface or polygon are the three primary representation and then we have a class of geometric.

So, they bring on to some geometric class and these this is also should have a spatial reference. So, whatever we are doing with respect to a spatial reference system. Now, if you look at, so there can be a line string which is a part of a curve right, line string has two points at two ends like a line string has to have the two or more points at the 2 ends right. So, that it is the erode is one either these two points or these; these; these; these whatever is going.

Similarly, for the polygon also, it has a means bounding area which is a line string, its basically polygon if we can see its a line string only where your starting point and the ending points are same. Similarly, we have a geometrical x n, so multi surface, multi curve, multi point things are there right. It is a collection of things right, if I have multiple polygons, multiple point, multiple lines and type of things right.

So, this is the if I see that if I my representation in a 2D space is point polyline and polygon with this with help of this I can represent or I can map or represent any type of spatial data scenarios.

(Refer Slide Time: 25:25)



Now, if we look at hmm classification operation. So, it is state based 2 dimensional spatial objects that is polygons are set of sets of points right. So, a set operation interaction of 2 polygons produces another polygon right, if I say to any set theoretic operations or set operations on two things. So, one is set based or set theoretic approach of classifying operation etcetera like there are all that whatever interaction and union it set the whatever the set operation are made.

Other is the topological operation this is very interesting right and this is this sort of things that you do not find in case of standard datasets right standard hour data sets like bank or library and type of things right. So, like I can say the boundary of West Bengal touches the boundary of Bihar.

So, there is a this data model there is a concept of touching one touches one another like there is a directional like Kolkata is in east Kharagpur right. So, if that is thing. So, that is directional, it can be metric say Kolkata is about 150 kilometer from Kharagpur right this is a metric type of things. So, we have to classify this spatial operations into different type; one is set based operations, other is topological operations, another is directional operations, another is metric that how much things are there all right.

So, topological touches like set theory based is that union intersection containment these are the set theory based, they are topological which touches, disjoint, overlapped etcetera are topological based operations and the metric may be distance and there are different

type of calculation etcetera will come into play when we talk about metric type of operations.

(Refer Slide Time: 27:43)

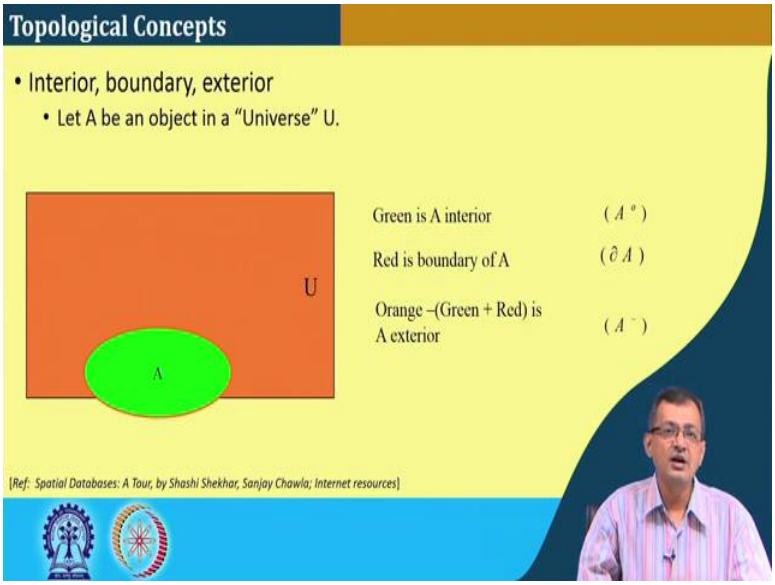
This topological relations are interesting right. Invariant under elastic deformation, all these topological relations are invariant under elastic deformation right. In other sense if I say if two regions are touching each other whatever you do with the things they still touch it each other right unless you tear it off right, you stretch it this way that way thing the touching will be there.

So, the topological relation is invariant of elastic, so without tear, merge etcetera. Two countries which touch each other in planar paper map will continue to do so in spherical global map also right. So, if you put on the globe then also touch each other they are deformation may be there, but this topological relationship will not go out right, this is a very important aspects of the spatial data models and spatial database and how to handle this type of things is a major is a challenge right need to be handled you know when you talk about spatial databases.

Topology is a study of topological relationship, so when you talk about topology many of you might have studies. Example queries of topological operations; what is the topological relationship between two objects A and B? Find all objects which have a particular topological relationship with A? Like find all state which touches West

Bengal, find all states through which river say river Ganges flow? Right. So, this these are I require something which is which comes with a some topological relationship right.

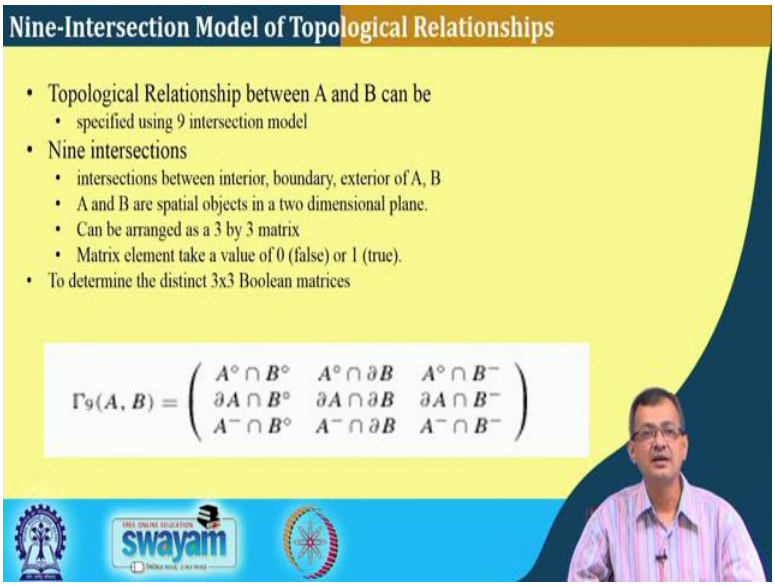
(Refer Slide Time: 29:27)



So, one aspect we will see this, so when we talk about two regions consider poly 2 polygons, it may be polygon line; line; lin; line point anything right. So, any region for that matter interior boundary, exterior boundary; interior, boundary, exterior three things will be there. So, let A be object in a universe U, then what we can say green is A interior, the green color is As interior, red it is not too visible due to this color problem this boundary of this is the rate actually merge to it this orange color.

So, that, so this is the red is the boundary of the things orange minus green plus red is the exterior of it right. So, this is our definition. So, we give by these symbols these symbols into the aspect, so these are the three things.

(Refer Slide Time: 30:39)



Now, with this I can define any type of topological relationship involving any things, it is generic any objects right like here it is we want to see that topological relationship between A and B can be specified using 9 intersection models right intersection between interior, boundary and exterior of A and B two objects A and B.

So, they can be 2 polygons, polygon and polyline 2 polyline, point, polyline and say anything right, but this 9 A, B are spatial objects in a two dimensional space. Can be arranged by a 3 by 3 matrix; matrix element take a value 0 false or 1 true to determine the distinct 3 by 3 Boolean matrix right. So, you need to determine the 3 by 3 distincts matrix when you are defining the things right. So, by this we will be able to define the things right.

So, what we see, what we try to see or what we tried to discussed today's class is that one is that there are two broad model of vector and the raster we are primarily now considering the vector type of models and try to see that what are the different way of representation of the data of these models right. So, rather from this model we will be going to how it can be a spatial database and etcetera can be build up which can be used further for queries and other type of operations say data mining and type of things.

So, that is our basic bottom line of looking at these data models. So, this particular discussion of this topological relationship and what are the different type of matrices we will continue in our next class, where we discuss about more about these data models

and then we will see that how this data models, how this relationship can be established like what is the entity relationship things can be entity and relationship between those entities can be established. I believe that most of you are accustomed with the ER type of model, relational database model and etcetera.

So, we will touch upon, but there is no scope of going detailed on those models. So, those who are not so conversant I encourage you to go through that any standard database book that what is ER model, what is relational database models and so that it will be good to discuss on those aspects with spatial extension. So, let us stop today's discussion here and we will continue in the next class over the things.

Thank you.