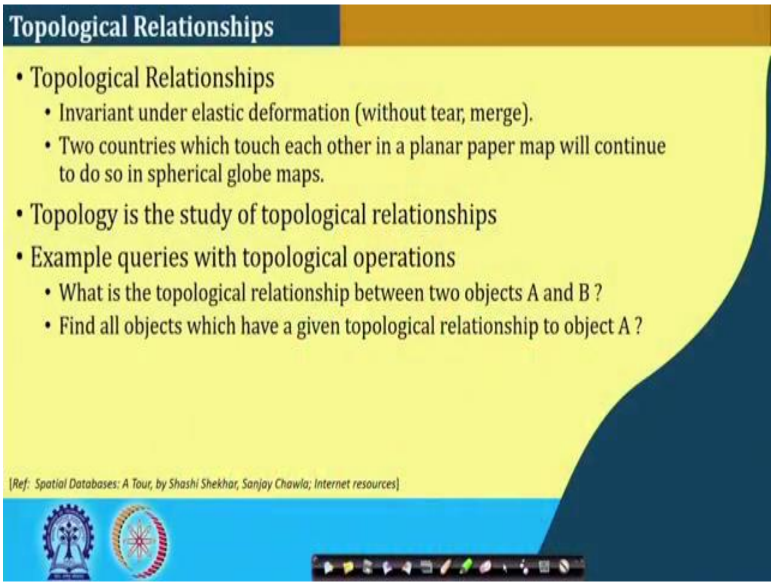
Spatial Informatics
Prof. Soumya K. Ghosh
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 03
Spatial Data Models (2)

Hi, let us continue our discussion on Spatial Data Models. So, in the last lecture, if you recollect we are discussing on the data models per se and spatial data models more specifically. Those who have background in data base or even working with data in different aspects the data model plays a important role in any information system right that how the data is model, how the data schema structured is extremely important when we try to extract information, try to interoperate between depositary is an type of things.

And in case of a spatial data, it becomes more tricky, because there are other representations, geometric representation along with whatever they are within our standard data, set spatial data sets come with more features which are not pretty straightforward to handle right, so that is why we are looking at the spatial data model. Rather in the last class where we what we are looking about some of the typical features of spatial data model, one was topological relationship. We will continue with that and see other aspects of the spatial data model.

(Refer Slide Time: 01:30)

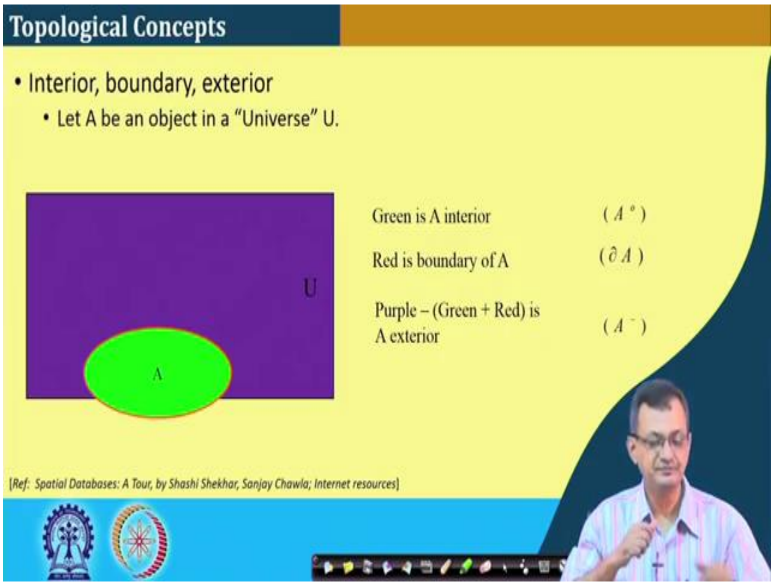


So, what we have what we have seen in the last class last lecture on these data models that is a there is a topological relationship right. So, this topological relationship are typically invariant under elastic deformation right, like unless you tear or merge something etcetera then the relationship holds. Like we say that that next line if you see that two countries touches each other right or that means, share a boundary, so that will that will that will remain whatever you do with the things like whether you put on a planar map or spherically you put on the spherical map that relationship still holds on right.

So, this is a typical feature of spatial data right. You cannot have like a number touching another. So, has to be there for the spatial features. So, you will see that this topological relationship with our three specific things like point, polyline and polygon. We argue that any type of features on a as of now on a 2D space, we can represent by this as a vector representation.

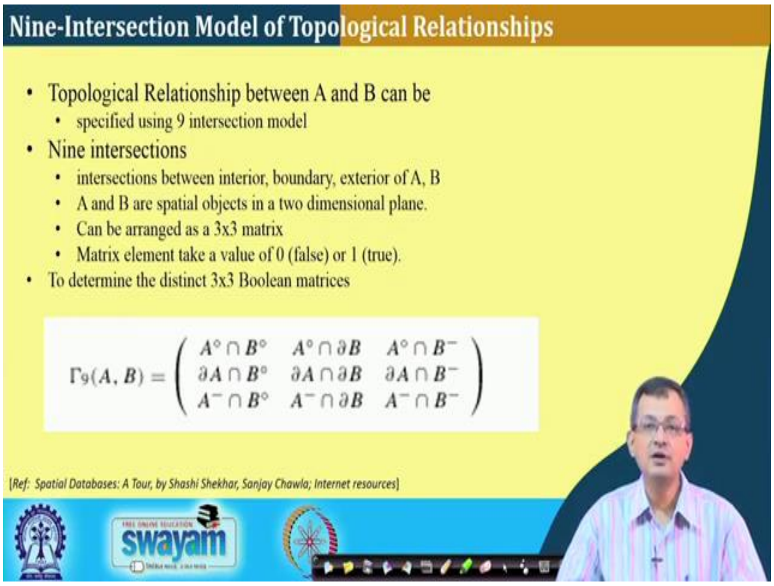
So, topology is a study of topological relationship and many of you with computer science or IT background have might have studied number of things on topology. So, queries like what is the topological relationship between object A and object B or objects A and B find all objects that are able to typical topological relationship to object A like as last I was telling that find all state which are sharing boundary or which touches West Bengal right or what we can say that find all state of states of India which has a seashore or touch touches the sea or something like that right or something crossing etcetera intersecting. So, those are things we these those type of queries make sense in this context.

(Refer Slide Time: 03:45)



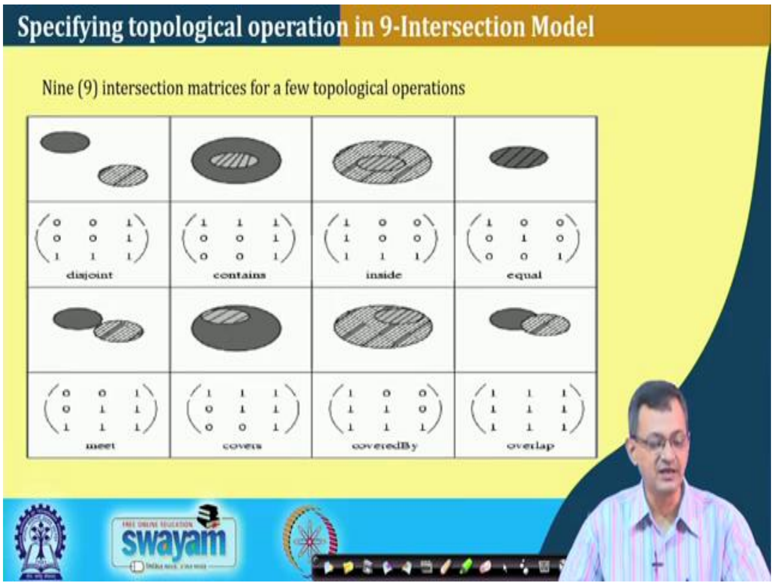
Also we have seen this picture I have changed that color of that orange, otherwise it mixing with the red. So, we have three things right say for polygons. So, green is interior to A. So, A node type of thing. And red is boundary of A, red is that red boundary of A, so that is represented by the del A. And this purple minus green plus red is the A exterior of red. So, universe, the whatever is outside this green and red is exterior of A right. So, this, these three A minus we put, so these are my three way how I can represent how I can see, how I try to see the other relationship with other things right.

(Refer Slide Time: 04:38)



Now, topological relationship can be between any objects A and B can be specified using 9 intersection models right. So, like intersection between interior, boundary, exterior A and B right. So, totally we have three, three, 3 cross 3 - 9 sort of model. So, A and B are special objects in a two-dimensional space can be arranged in a 3 cross 3 matrix, matrix element take a value 0 if false, 1 as true. And it is help in determining the distinct 3 cross 3 Boolean matrices by which can be represented. It will be clear if we take a one example or a set of example right.

(Refer Slide Time: 05:22)



Now, see here we have two polygons A and B, and there are different type of relationship, one for the first upper left this is a what it says about the disjoint. This is contained, one is contained in another. This is inside right, one is containing, another is having this inside this. There is a equal overlapping. There is a meeting, there is a cover it is meeting, but it is just touching insightly or covered by type of scenario or overlap. So, if you take the disjoint, so in this case of disjoint what we see 0 0 1, 0 0 1, 1 1 1.

Now, let us go back. So, A naught intersection A naught. What was A naught, it was interior. So, A interior intersecting with B interior, if it is disjoint this is 0, there is no inter intersection, no intersection between that. So, the first one is 0. A naught intersecting with del B. So, what was del B, boundary right and interior. So, interior with the boundary in case of a disjoint, it is again 0 right.

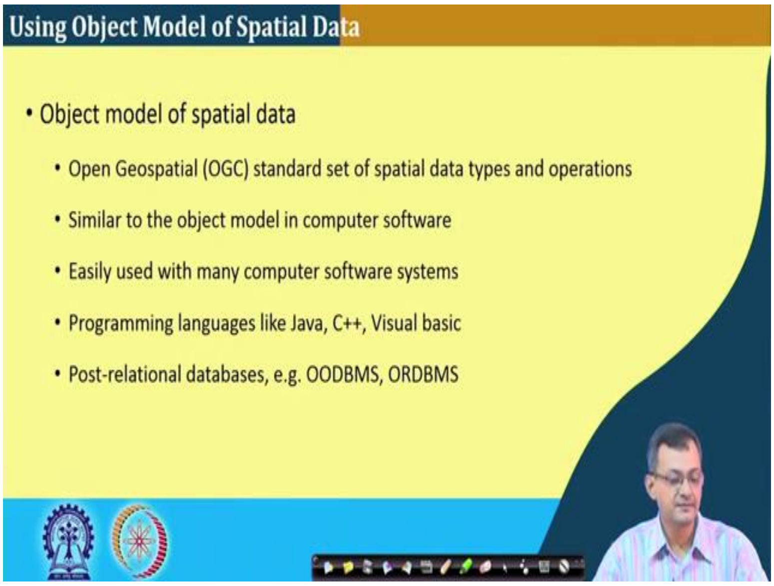
And then next one is 1 that is A naught intersecting with B exterior right or A interior will be exterior. Now, in case of B exterior, in this case it is the if this is B then the rest of the thing is B exterior, or if this is B extra of the things anyway whatever is what whatever is the B is, so that is that if it is intersecting with A B exterior in case of disjoint, there is a overlap right and that is why we say it is a the value is 1. So, there is some value when we interact intersect so that means, 0 0 1. And similarly you can if you look little clear minutely, then you will see that 0 0 1, 1 1 1 right.

Similarly, if I come to this other end at the overlap right, so in case of overlap A interior B interior overlapped in this particular scenario right A interior so 1. A interior boundary of B overlaps. So, A interior boundary of B overlaps; A interior B exterior overlaps. So, this is overlaps and so and so forth right. If you calculate go on calculate like here, A boundary B boundary overlaps right, it crosses each other that is 1. So, like that all things are 1.

Whereas in case of touch which is just touching, so it is 0 0 1. So, A interior B interior does not overlap. However, if you see A interior B exterior that is overlapping right ok. Or rather if you see this one that A boundary B of boundary are touching or intersection is 1. So, in this case, this is 1 right. So, you see what we mean to say that any this between two polygon, these are the possible situation, two planar polygon right, we need 2D space. So, these are the possible situation.

And this we can represent by some matrix right. So, if I have two polygon, I can represent by two matrix right. I can generalize with three polygons and etcetera, but nevertheless this helps me to map it using a matrix right we have a handler to do that right. Similarly, you can try for polygon line, line, point, point polygon, this type of finding that what are the different type of relationships are there right.

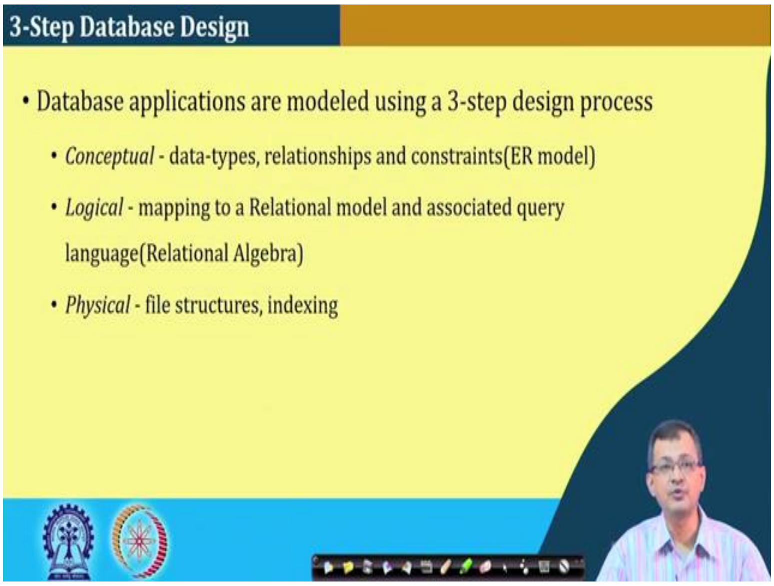
(Refer Slide Time: 09:38)



Now, we with this thing, we come back to this object model of spatial data using object model. So, object model of spatial data, one is that open geospatial consortium or OGC, those who are interested can see you www.opengeospatialconsortium.org standard that is a standard set of spatial data types and operations. Otherwise everybody will use something of their own, and it will be difficult to interoperate. So, there is a standard set of spatial data types and operations. Similar, to the object models in the in our when we say computer software or software engineering aspect easily used by many computer software systems.

So, as we see in case of our object models, in case of other field these also can be easily for any type of software and tools which deals with this. Programming languages like Java, C plus plus, Visual basics or there are Python and different type of languages in this. And we can have post-relational database object-oriented database, object-relational database and different type of things we can work on those type of things where this type of models.

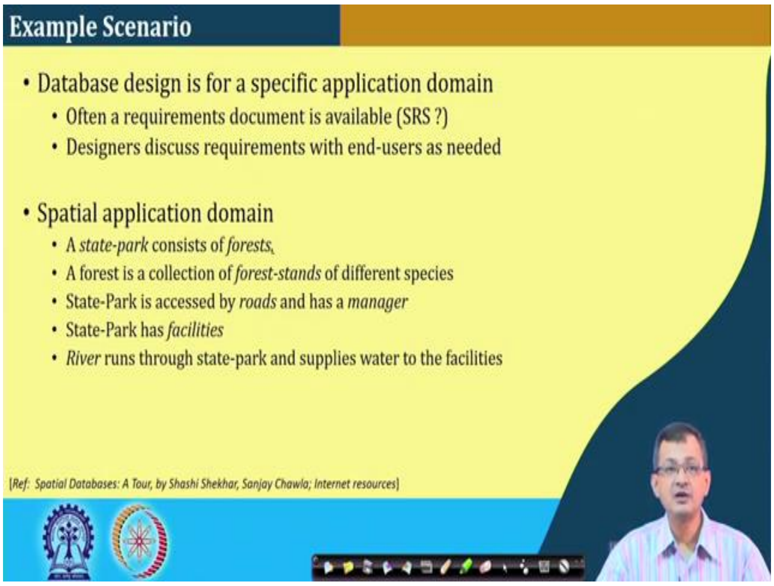
(Refer Slide Time: 10:57)



And if we look at a database, it is a 3-step design process. This is generally nothing to do with nothing special as of now with the for this actually for the special data, definitely special data needs more what we say there are more complexities and things are there, but never nevertheless we want to do a conceptual model data types relations and constant ER model. Logical mapping to relational model and associated queries, language – relational algebra, and so and so forth, those are the logical model. And we have physical model file structure indexing and type of things how things are stored; in other sense conceptually I want to find out that which are the which are the data types relationship and constant those who have background in data modeling, databases etcetera that that ER diagram is one of the popular thing that entity relationship diagram, so that is conceptual.

Then we wanted to map to the relational model or associated query language right like a relational algebra, because finally, we need to query on the things, so relational model that is more of a logical model. And then finally, how the file is structured within the disc is also important. It is this is specially specifically important for spatial data, because this sort of data are not only compute intensive, there data intensive also right. So, that means, the data how they are organized may dictate the performance of the query right we will see those indexing schemes etcetera sometime later in this course, but nevertheless these are the three standard way of looking at it.

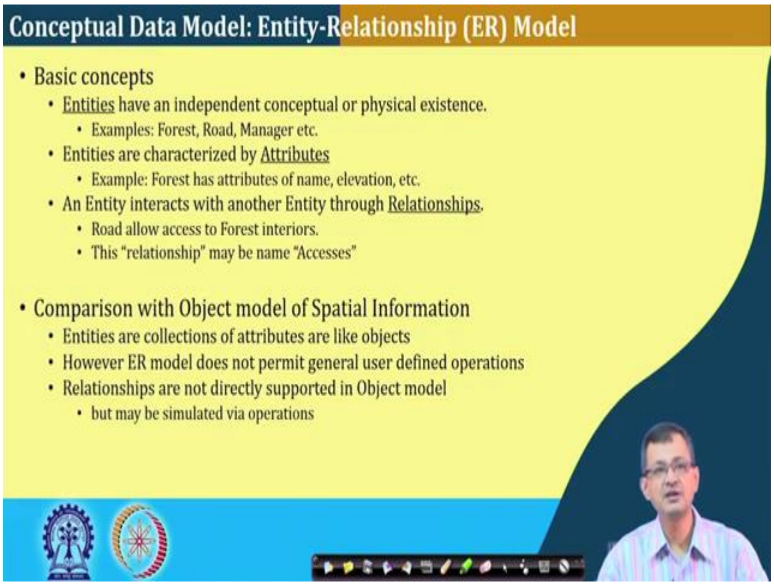
(Refer Slide Time: 12:32)



So, database designed for a specific application what is required the requirement document specification or say so to say that some sort of SRS type of thing. Designers discuss requirement with the end-user and freeze it as needed. Here also same cases. And in thus if we look at the spatial application domain with the same example scenario from that particular book, then we have a state-park consists of forest. Forest is a collection of forest-stands and different species. State-park is accessed by roads and has a manager. State-park has facilities. Rivers runs through state-park and supplies water to the facilities.

See these are facts right, a a state-park or in particular example anything. Like if I say IIT Kharagpur is collection of different academic region and residential region, commercial region, etcetera some part of the things academic region there are collection of departments and so and so forth. I can say that it is a a particular is department has number of floors etcetera and so on and so forth. So, I can define any spatial way things are there, definitely non-spatial or attribute data are also embedded into the thing right. So, for simplicity we take this one.

(Refer Slide Time: 13:49)



Now if we look at the basic concepts, again this may be a repetition those who are already used to it, but this is for our benefit for larger community, we just have quickly browse to this. So, there are entities have an independent conceptual or physical existence like every entities has a independent conceptual or physical existence like student database is a student is a entity or say faculty is a entity in a academic data database or data model type of things and so learn like in a in a in library we say that book is a entity or authorized is a entity and so and so forth.

So, examples in this our particular example case is forest, road, manager, etcetera are entities. Entities are characterized by attributes right. So, entity, how they are different from other entity, because their attribute sets are different right. How a student entity is different from the faculty entity because of that the attribute set for the student is different from the faculty attribute. So, entity interacts with another entity through relationships. Road allows access to forest interiors, these relationship may be name of the accesses and type of things. So, I say entity interacts with another entity with relationship right like I say teacher teaches a student or I say in this case road allowed access to the things. So, this access is a relationship right. So, this is accessing. So, I can define different relationship, I can have some standard relationship and so on and so forth.

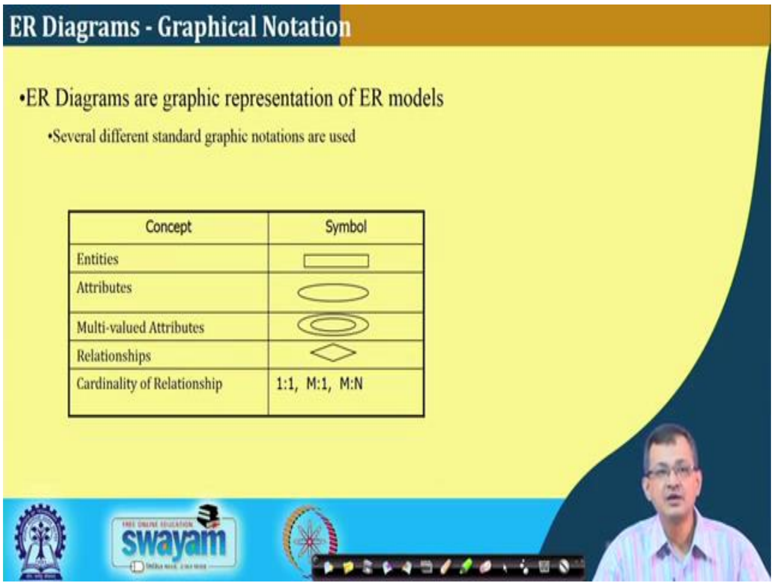
So, with the specifically for the special object model for spatial informatics what do you see, entities are collection of attributes like objects. However, ER model does not permit general user defined operations right. Typically ER model does not permit that general, general is the define operations; relationships are not directly supported in object model, so but by simulated by operations we can do that.

(Refer Slide Time: 16:05)

Now, any relationship which characterized first of all the cardinality of the relationship, we will come to that, many of you must be knowing. Other properties either number of participating entities etcetera, binary relationship are two entities. So, if you consider a binary entity, there are two relationship A is related to another object another entity B right, but there is A cardinality.

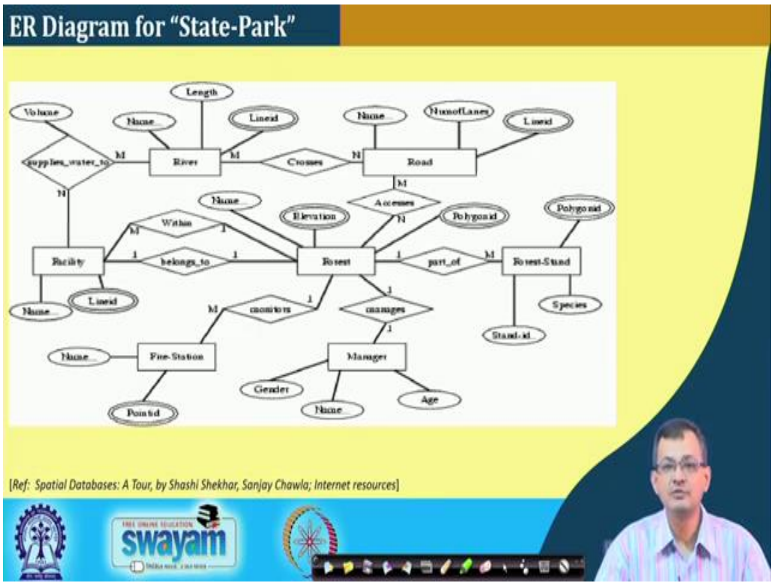
So, what is cardinality? I can have a one-to-one relationship an instance of an entity relates to a unique instance of another entity right. So, I can have a typically one to one relationship of a particular between two entities. Many-to-one relationship many instances of an entity relate to a instance of another entity right. So, number of instances are related to entity or many-to-many instance number of things are taking number of relate to multiple instances of another. We will come to, I will so you the different type of examples, many of you may be knowing.

(Refer Slide Time: 17:12)



So, typically in a ER diagram as you know that these are the representation usually right entities and like these and attributes, there is a multi-valued in attributes, we have multi level of entities, and there is a relationship right. So, and the cardinality one is to one, m is to one, m is to n type of cardinalities. So, these are the different representation of a typical inter entity relationship model.

(Refer Slide Time: 17:45)

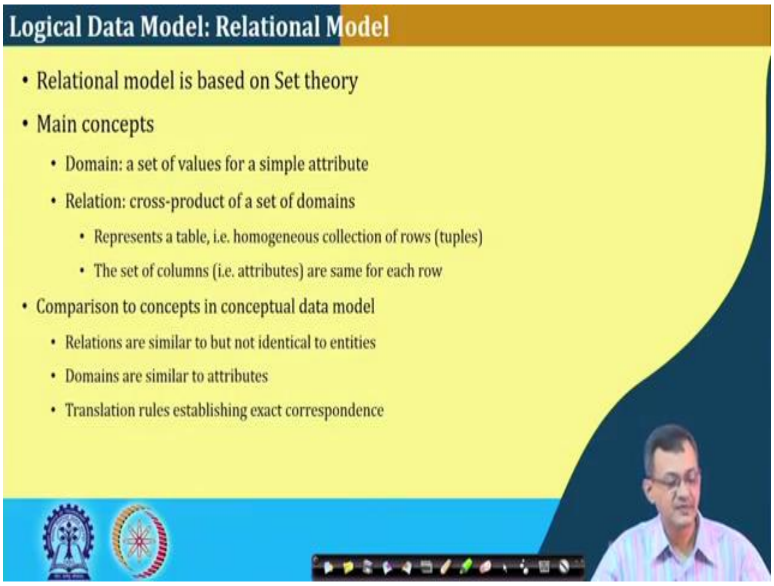


With this we try to now represent that the same state-park of of the thing right. So, there are there are different entities like entity called river. It has attribute name, length, some line id right, it is it is for the spatial purpose. And there is a road name, name of the lanes, line id etcetera. There are entity forest, entity forest stand, entity facility, fire station, manager, these are the different type of entities right. And they have their respective relation a attributes like manager has a gender, as a particular name, a particular age, and you can have lot of other activity, he or she can have a lot of other attributes right.

Whereas some of the things like say a facility or a thing called river is as an another feature called line ad. So, these are spatial features like. And there are different relationship a river crosses road right. So, this is aim into many, many rivers can cross many roads right, the same river can cross two roads, a road can cross two rivers and so on and so forth. So, M is to N relationship. Whereas, if you see manager to the forest is one is to one, that means, a manager can be can manage only one forest area if forest can have only one manager right. Whereas, this facility to forest can be also one is to one relationship, like a facility belongs to a forest if forest can have one facility right or a actually facility belongs to the forest is one is to one.

But a forest can have multiple places one is to map, that means, within the forest there can be multiple facilities. Similarly here also forest stands and so on and so forth. So, this cardinality plays an important role in defining these things. Now, see the those who have used I believe most of you have been used this ER diagram give you a very nice pictorially present things right, you on the on the fly you have a now feel of the things that these are the entities, these are the attributes of the entities, and these entities related to this entity by this relationship. This is excellent way of conceptualization of the whole problem right.

(Refer Slide Time: 20:20)

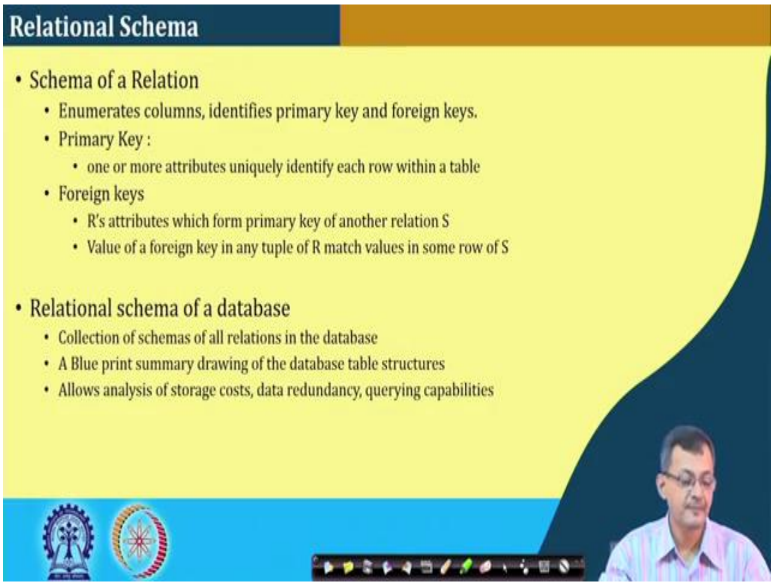


But for database working or the going to the database schema, we require something more concrete what we say relational model which is the are the prevalent model for any database type of stuff. So, relational model is based on set theory right. So, whatever the set theoretic approach is that there is a relational model. Main concepts, domain - a set of values for a simple attribute represents table that is homogeneous collection of rows or tuples, a set of columns or attributes for the same each row. So, that means, it is a two-dimensional table, columns represent that attributes right, and the rows represent the different tuple right. We have done for student database right, student name, student roll number, name, then phone number, department, call, etcetera, etcetera come into play then we have different students in the things.

As it is a set then I cannot have two tuple repeating right. So, it is only one tuple in that typically in a relational table right and all other set theory things will come right. So, so we have a set of attributes and cross product of set of domains will constitute represent the table, and the set of the columns.

Comparison of the concept in conceptual model, relations are similar but not identical to the entities. But it is a similar to the entity entities are there. Domains are similar to attributes. Translation rules establish the exact correspondence, how it will be translated that how these are mapped is the that rule is important.

(Refer Slide Time: 22:11)

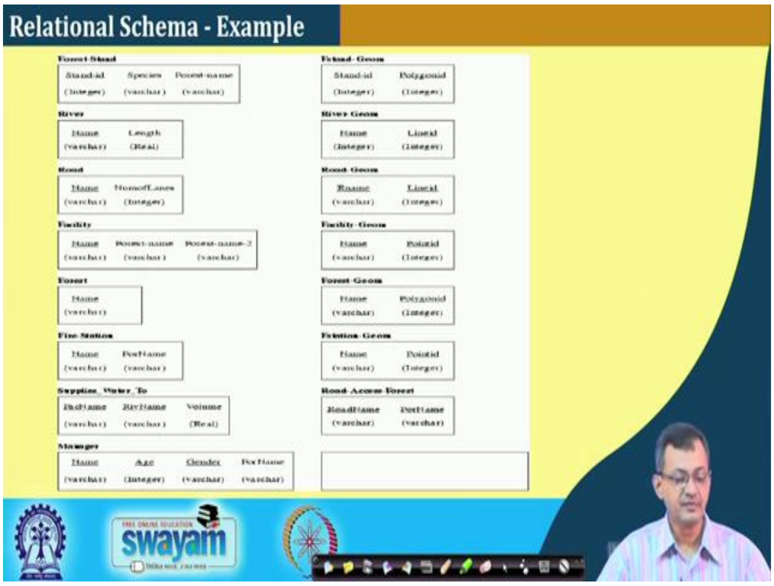


Relational schema, schema of a relation. So, it enumerates column identifies primary key. foreign key, foreign keys right. What is a primary key we know all, one or more attributes uniquely identify each row within a table right. So, I say primary key is the student roll number ok. So, it uniquely identify every tuple or every row in the student table right. It can be if the I could I can say the student name and the address may be the primary key right, so that means, uniquely identify.

Foreign key, R’s attributes which form primary key of another relation R. So, a relation R’s attributes which from the primary key on the relation R. This is important, but because when we query multiple relations or multiple tables this becomes a important factor values of a foreign key in any tuple R must match the values in the, in the say some row in the S right. If it is a foreign key of S, then some row it should be matching right. So, I can have a foreign key like departmental code may be a foreign key in the departmental table. So, there should be a existence of the departmental code right. I cannot give a code with that of a known existing department.

Allows analysis of storage data redundancy query capability and other things this also schema of the database sorry. So, what we are talking about relational schema. So, based on these I can have the relational schema of a database is I am not looking at the data at all. Now, I am to find you know the what is the structure of the relational schema of the database, collection of schemas of all relation in the database. So, all the relations and the schema is the is the collection of the thing. A blueprint summary drawing the database table structure, so it is a blueprints summary or the blueprint for the database table structure. Allows analysis of storage cost, data redundancy, query capabilities, etcetera for the database right. So, this is important thing.

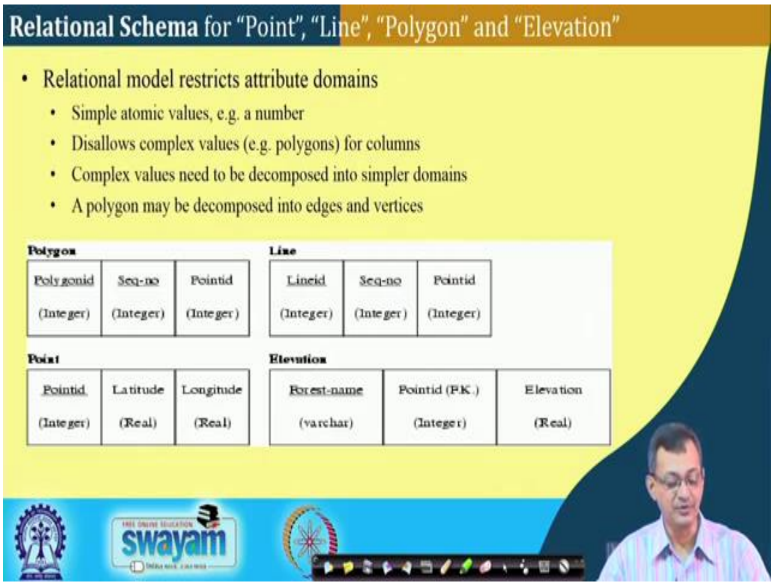
(Refer Slide Time: 24:31)



So, if we see that different things like here the forest stand, stand-id, species, forest name or say river, river name, length of the river, road name, name of the lanes or say we have different type of things right. And see these are things which there is no geometric class or no so to say spatial extent. Here we have the geometric class like say river geometry, name integer, line id right. So, it is a it is a what we say multi value, they are different number of line ids into the things right. Similarly, if I say forest geometry, it is a polygon id right. So, it is a polygon. So, this is the geometry class.

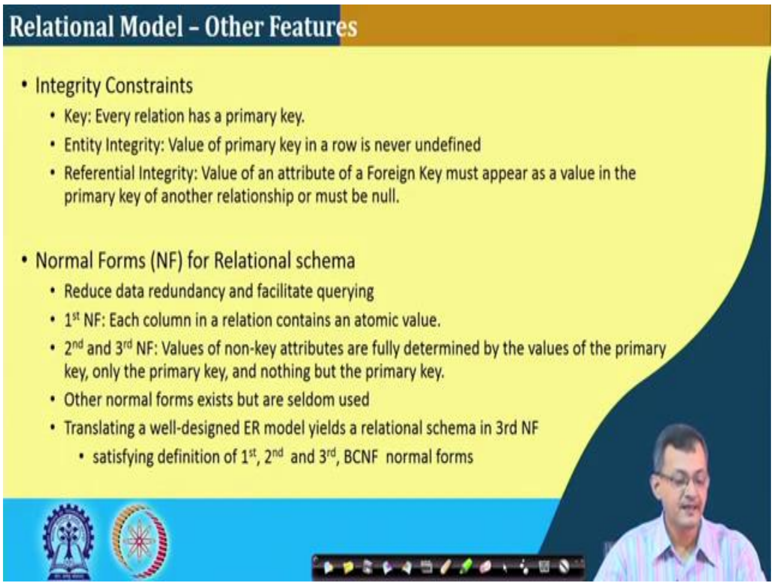
So, my database schema is able to handle this geometry class, but at the same time the database schema should be also should be able to have a implementation in the database right or in other sense what we say that spatial database should have these features that polygon, polyline etcetera need to be represented probably.

(Refer Slide Time: 25:47)



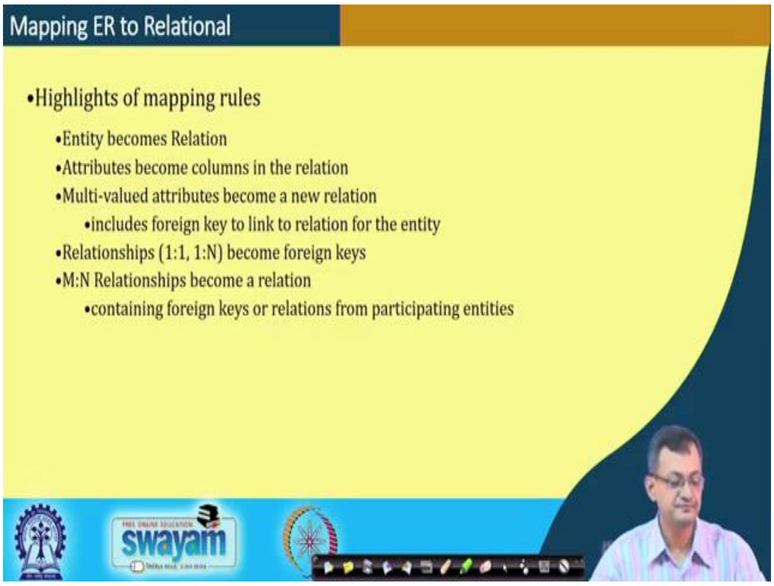
Like we say relational schema for point, line, polygon, in this case another elevation, if we keep about the elevation as of now, so polyline, polygons, sequence number, point ids right. So, it is and if you see the line, it is also sequence number, point id, in this case the last source and the destination point id are same. So, point is latitude-longitude x y. So, I can map that like say I have these are the points for my polygon, these are the points for my line. And this every point has a xy coordinate right lat-long or any coordinate systems. So, this is this way I can keep the track of the spatial extent of my data.

(Refer Slide Time: 26:29)



These are some of the other features already you are knowing this. Just to mention, one is the integrity content like every relations has a primary key. Entity integrity, value of the primary key in a row is never undefined. Referential integrity value of the attribute of a foreign key must appear the value in the primary key of the another table relational table, these are the integrity constant there are normal forms that means how to reduce the redundancy. So, first normal form, second normal form third normal form BCNF, so these are to reduce the redundancy, so that the overall efficiency increases or consistency easy to maintain the consistency. So, it is reduce redundancy and facilitate efficient query processing.

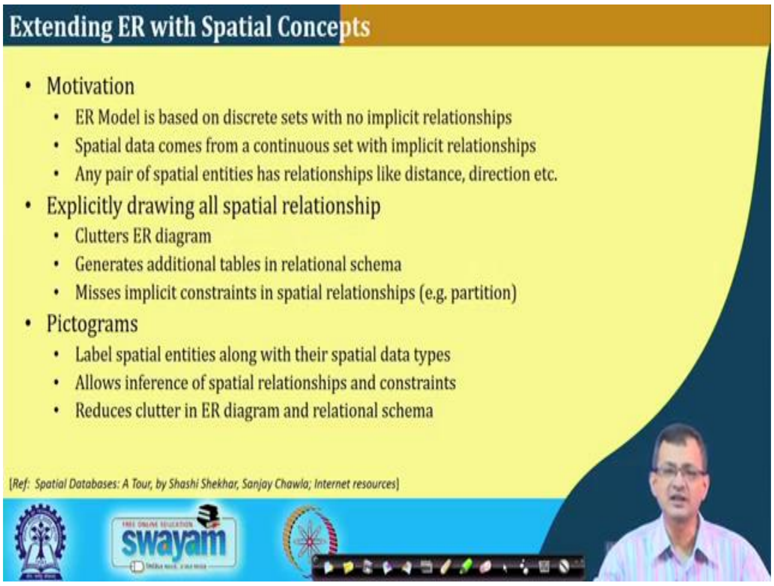
(Refer Slide Time: 27:16)



So, mapping ER to relational schema or relational model is another aspect and then only we can go to the database. So, entity becomes relation so to say. Attributes become column in the relation. Multi-valued attributes become new relation where things are as a connected with the things includes foreign key to lean to the relation to the entity right. I can have a line, polyline or say river and which is connected to the number of points on the other relation.

Relationship 1 is to 1 is to N etcetera are the foreign keys right, 1 is to 1, 1 is to N, M is to N, these are defined by the foreign keys. M is to N relationship become becomes another relationship right 1 is to 1 and 1 is to N, we have handled by the foreign key, but for M is to N relationship we make another schema. So, it says that M is to N relationship containing foreign keys or the relation from the participating entities. So, it is participating entities things are there right. So, I have two things which are participating entities.

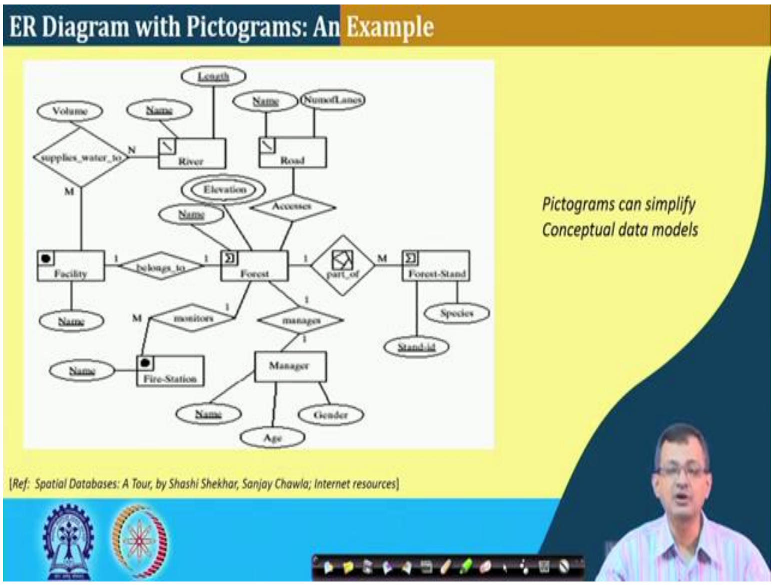
(Refer Slide Time: 28:25)



So, these are again known facts. So, extending ER to the spatial concept with that those things along the motivations the ER model is based on discrete sets with no impair implicit relationships there. Spatial data comes from continuous set and implicit relationships that is the challenge, whatever the ER diagram for the standard models what we have seen. Any pair of spatial entities as relationship like distance, directions, etcetera right. So, explicitly drawing all spatial relationship will clutter the ER diagram right. Generates additional tables in the relational schema. Misses implicit constraint in the spatial relationship like partitioning of the data sets and etcetera. So, just implicitly we cannot just map it will not become a, it will become a misses so to say.

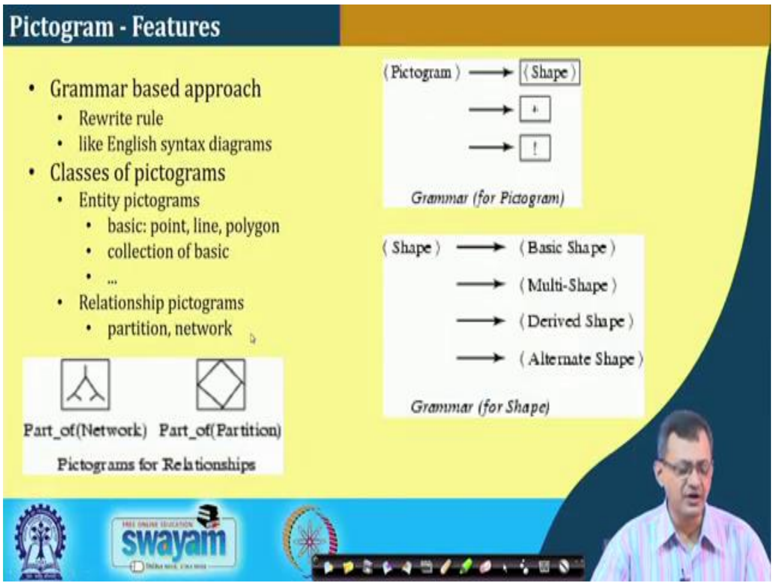
So, there is one approach of going towards pictograms right like label spatial entities along with their spatial data types allows interference of spatial relationships and constraint, reduces cluttered in ER diagram and relational schema. So, these are the things which are they are in the pictograms.

(Refer Slide Time: 29:42)



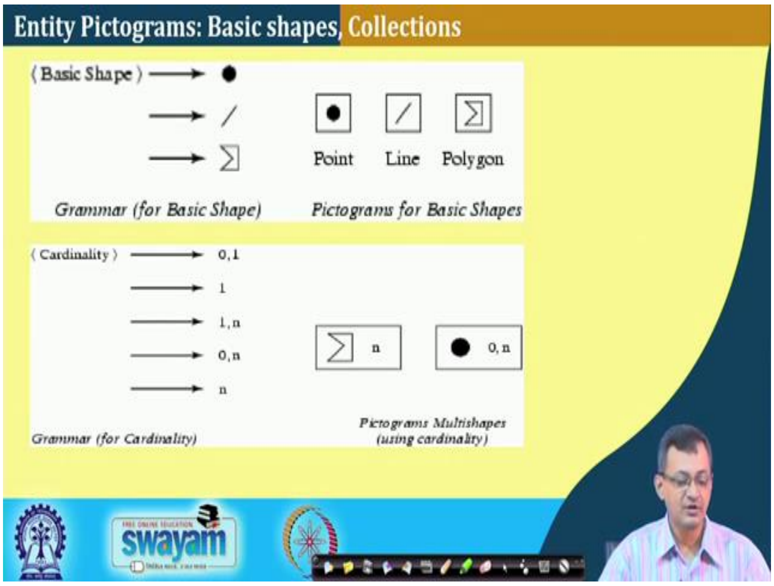
So, typically pictogram looks like that right. Along with that if you remember or recollect that entities along with that it shows that what sort of like river is a polyline, or line, feature here the facility is a point feature, forest is a polygon features. And we can also say that forest is a part of forest-stand. So, this is comes as a another relationship right. And this is easy to handle and see and check things like that.

(Refer Slide Time: 30:08)



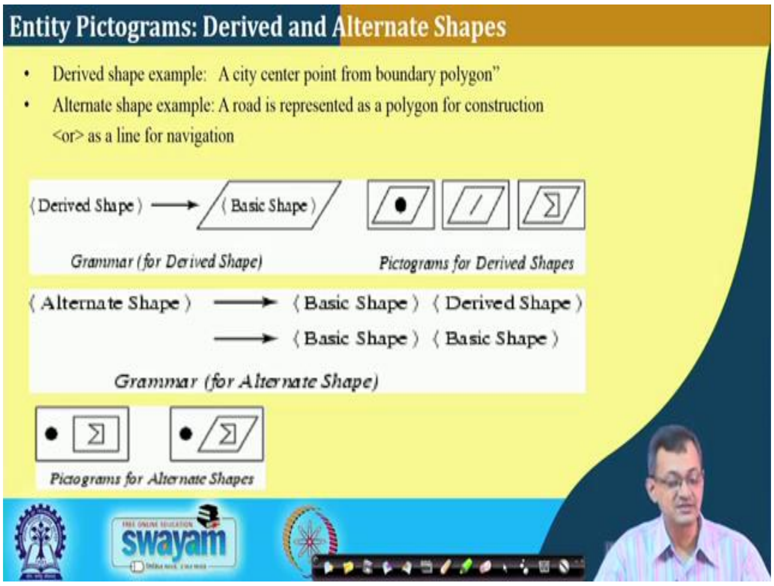
So, there are lot of things we can because pictogram also defines say grammar because there should be a grammar to represent those things. So, there are this part of network, part of partition, this pictogram for relationship there are shape, there are basic shape multi-shape derived shape alternate shape and there are different type of relationship shapes and etcetera we can work on right.

(Refer Slide Time: 30:36)



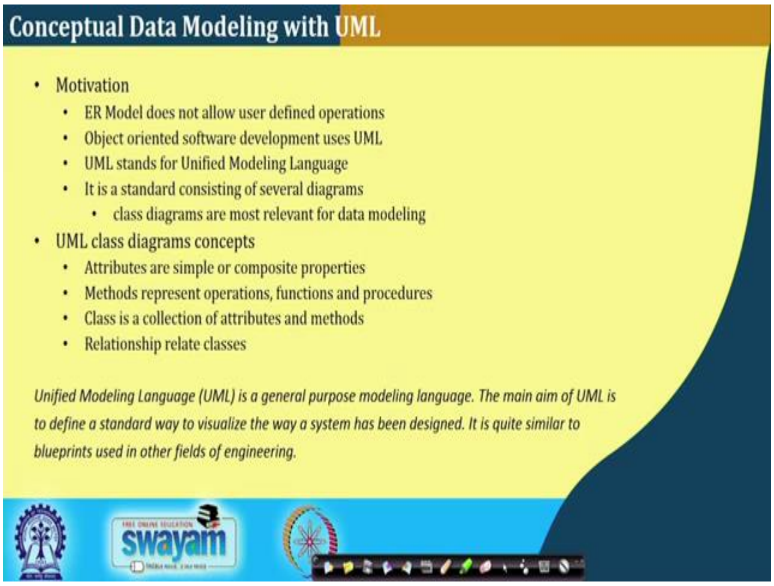
Like we have basics like point line polygon, they are a representation of the cardinality 0, 1, 1, a to n, 0 to n and type of things. So, pictogram will multiple shapes using cardinality right. Now, you see there are situations like shape may have different represented and different context like say if I want to find a point A to B that is what is distance then the road is a line right, but suppose we want to expand the road or make some modification construction on the road then the road is a polygon right. So, similarly on a large map cities may be a point, but if you expand it for doing some analysis then it is a polygon right. So, that there are different way of representation, so the same thing as different represent a different context right.

(Refer Slide Time: 31:36)



So, entity pictogram derived an alternate shapes, derived shapes a city center from a boundary polygon. I can find out the what is the city center. A road is represented as a polygon for construction or as a line for navigation as you are discussing. So, it is a alternate shape. So, this pictogram has the capability of showing these different type of shapes right, so that those things you can represent in that pictogram enhanced ER diagrams.

(Refer Slide Time: 32:07)



Finally, we should mention we actually will in subsequent class we will discuss about this there is a through UML, UML. I believe most of you are knowing that unified modeling language right is a general purpose modeling language. So, main aim of UML is to define standard way of representation or visualize the system that has been designed right, any UML not only for spatial data anything UML right, it is quite similar to something a blueprint of the whole thing right. So, the way of handling these is that motivation the ER diagram does not allow user defined operations object oriented software developed using UML is easier.

UML stands for that what we have discussed. It is a standard consisting of several diagrams, class diagrams are the most relevant or things which we use. Class diagram concept, so attributes are simple and composite properties. Methods represent operations functions and procedures. Class is a collection of attributes and method. And relationship relates to the classes. So, these are the properties of the UML that will look into the things and these things we want to use for modeling a spatial data. Given a spatial situation I want to make a UML model of the thing. If I can make a UML model we have the tool from the UML to the schema we can go we can migrate definitely interventions are required a different phase, but there is a possibility if I can design it.

And as you know UML is a universal unified modeling language, and a de facto way of representation a particular system, then easy, easy to interoperate easy to what we say interact with others and type of things right. So, with this what we seen today is more of that spatial data model aspects ER, and relational model how they fits into where the UML things come into. And the next class we will carry out our discussion on this data model mostly looking at more at spatial data how the how UML data modeling can help in different aspects. And we will also look into some of the aspects of spatial databases also.

Thank you.