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

**Lecture – 04**
**Spatial Data Models (3)**

Hi, so we will be continuing our discussion on Spatial Data Models; rather we last two lectures we discussed much in details not we talk about that what need of this data model and what is the specialty of the spatial data models; just to have a quick recap.

So, what we see that a data model is a gives us a de facto way of mapping a real world scenario to migrate to a database type of things right; so that how to map those things; so data model is a handler for that. There are we have looked into ER model many of you or most of you have already done in several databases courses. And also we looked into things like relational models and for the spatial extension what we have seen when we talk about spatial things; then what comes into play is the spatial context like the first one is that what is the geometry type right.

One is whether it is a point, polyline, polygon type of features and there are other things like SRS like or what is the reference system there will be we will see that there are other scenarios like context etcetera. So, as we discussed as again just to repeat we are as of now considering a planar surface or 2 D type of scenario not the 3 D when you talk about 3 D there are another dimension you will come into play all right.

So, there are; so in today's lecture what we will try to concentrate that using those concept of data models, how we will try to look at that how it is applicable to the special context right. So, that we will try to look at because see now if we look at the spatial things again; we are mostly with the geospatial data set that means the data which is of the art surface.

So, if you look at it there are multiple what we have the weight you know there has a multiple layers of informations right. It is same area like IIT, Kharagpur region or as I say this is my region of interest. So, there are things like a land use map how is Land Used Land Cover; LUNC, land use map. There are context of say road network or if you say transportation network or road network right. There are issues of built up areas

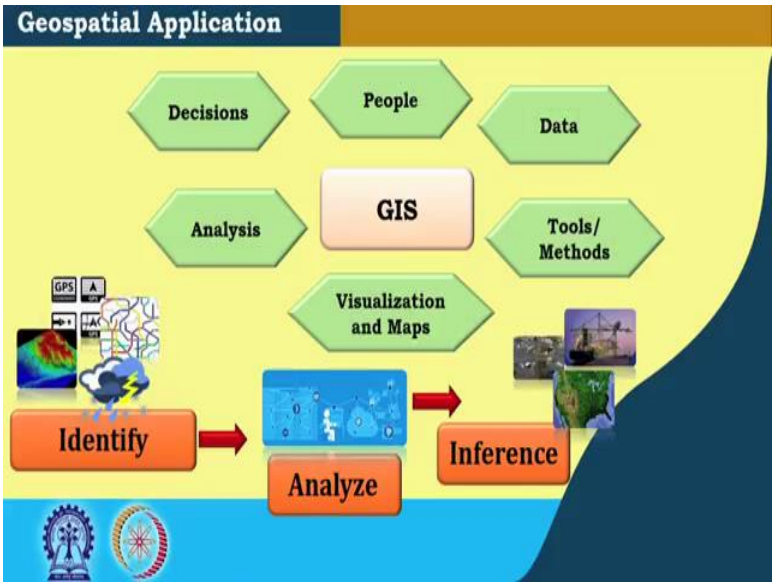
buildings etcetera, there are water bodies, there are several type of other species us right or other entities right.

So, some of them are point features, some are polygon feature, some are polyline feature and also we have seen the other context that some polygon may polyline may be a polygon in some other scale right. A road from finding a distance may be considered as a polyline, whereas or a road network, but when the road is taken up for construction say 2 lane to 4 lane; 4 lane to 6 lane or something; then it is a more of a polygon type of a things.

Like how much area for what is the how much material required then it is not based on the calculation of the line, but on the polygon feature. Similarly, a city or a particular say building may be a point and we need close down or other features then it is a polygon. And also we have seen there are attributed assets right, so a road has a name, a building has a name, a building has a location right. Though location is again a spatial context now building has a name building have different type of things; whether if it is a office or say academic building like ours that which is the departments housing there, if the department; who are who are the HODs or Head of the Department etcetera.

So, these are attribute or non special attribute datasets rights which cannot seen from the earth surface; so, those all those things need to be accommodated. Now the problem is now if I have diverse data set and wants to query; have some query which want to look into the different data sets, then how do I query them or how do I integrate them? So, that is one of the major challenge; when we see in case of a spatial data.

(Refer Slide Time: 05:07)

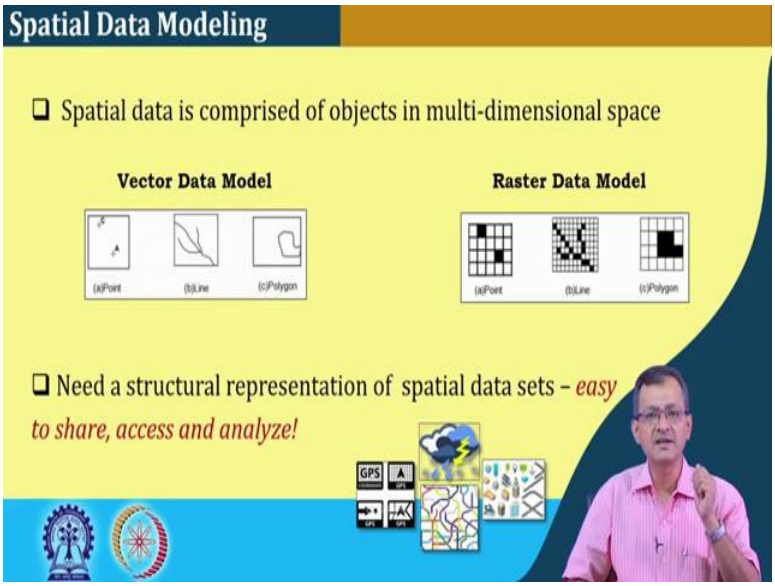



So, with this context let us look at our present scenario right. So, when we look at the spatial context; what we see that there are different aspects of the thing right. One is geographical information system or geographic information system is one aspects of the thing or whereas, other things like people, data, tools and mechanisms, visualization all those things are other aspects of the things.

So, what we see that a decision; one is the core tool or the geographic information; there are data sets, there are analysis, tools method, visualization and respects of the things right. So, what we want to do with these data sets? First of all, we may want to identify something right, we want to analyze something, we want to draw infinite something.

These we want the data sets to be in some form of other in a integrated fashion right; which we want to have a integrated view of this whole data sets right or diverse data sets. So, our objective or our focus of this things that how we can achieve this sort of interoperability or integration between the different data sets right; so, that is our one of the goal of these things.

(Refer Slide Time: 06:40)

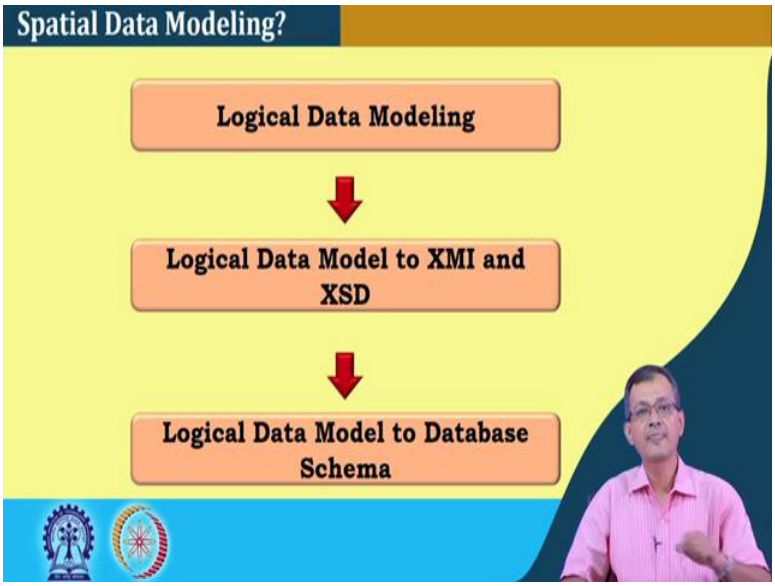



Now, this also we have seen right that spatial data is comprised of both your vector data set and raster data set right. So, the vector as we have seen, vector and raster that how one is that grid basin thing, another is your more of a coordinate based systems right.

So, what we require a needed structural representation of the spatial data. So, is to see our access and analyze and for that we require this spatial data model. And if we again recollect our previous lectures on these things; we have seen that one of the de facto standard which is followed is our UML modeling right; Unified Modeling Language right; how I can unifiedly model the data set.

So, this is the de facto standard we follow in any data modeling. So, here also we want to follow this UML and try to see that how UML help us in this doing so right.

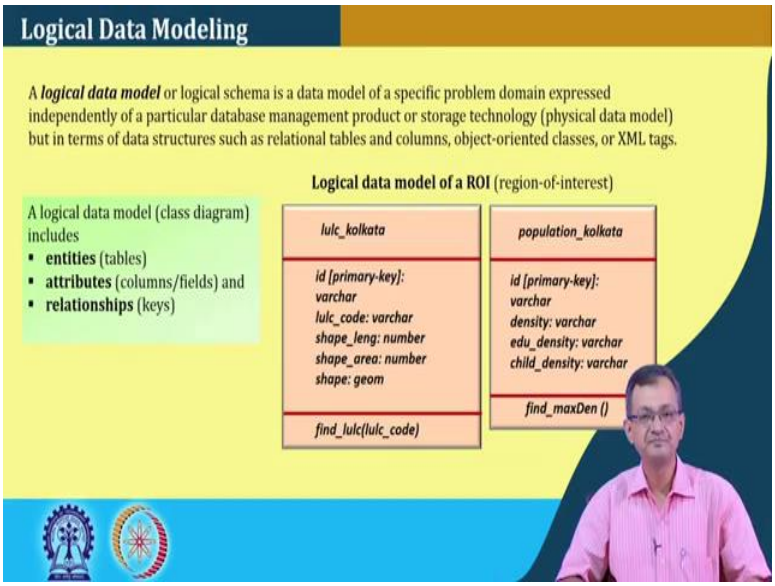
(Refer Slide Time: 07:47)




So, again this is the way we model any data sets, logical data model; logical data model to XMI and XSD you see that what is that and is from the logical data model to database. Or in other sense logical data model XMI, XSD and data database schema right. If I can come to the database schema, then my the problem is I should not say solved, but the problem is much in a easier to handle; now I can fire SQL type queries and etcetera to handle those type of things right.

So, somehow this logical data model or UML need to be seen that how that XMI or XSD; we will see that how what are they and then this transforming to the database schema. So, UML, XMI, XSD to UML schema; hey sorry the to database schema. Once the database schema is there then I can populate the data and then inter work with the datasets all right; one the schema is there of the database then we can have lot of handling on the things.

(Refer Slide Time: 08:56)




Now, if we just recollect all of you might have; radiological data model or logical schema is a data model of a specific problem of a problem domain right, expressed independently of a particular database management product or storage technology physical considering physical data model, but in terms of the data structure relational tables and columns etcetera.

In other sense what we tried to say that how I can probably; how I can present it in a some generic form right. One of the very standard way of representation; representation is our UML right, Unified Modeling Language. So, there are three things come into play again we have already discussed entities, attributes and the relationships right. So, how they link to each other, and if you those who have done with the class diagrams and sort of other things.

So, like this is a typical example scenario, this is land use land cover of Kolkata data. So, it has a primary key; it has a name lul cover etcetera then different things; one important thing is the geometry; other part are same of any a non spatial data one is geometry. Similarly there is a population map of the Kolkata where different things etcetera there, but if you see there is no as such the population do not have a geometry class right.

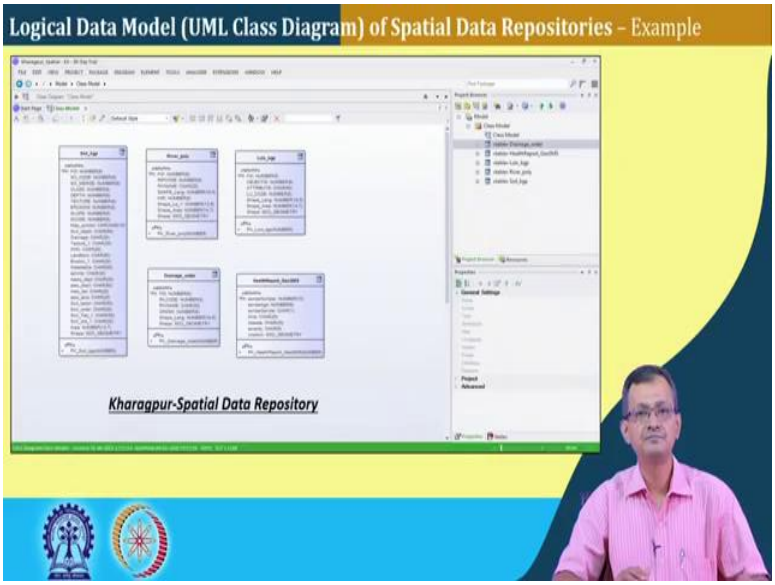
So, there can be different function like here find lulc; is the given lulc land use land cover code I can find out that which lulc etcetera they are find the maximum density population density etcetera things are come into play. So, for a particular region of

interest in this my in this case our region of interest is the Kolkata we have taken several type of small example to make it more meaningful.

So, but for see the population as such is not; that data as such is not having any spatial context per say. But in the other sense if you see the population is of a region right; population of a city, population of a mouza, population of a block districts etcetera. So, there is a special context to the things right; it is not like that it is out of context. So, there is a; that this what we say this is not geo referenced data or not geo coded data, we can have a in arrange geo referencing into the things right; so that is possible.

So, that that should be possible some way or other that we will try to look at how things with there; if it is there then I can map; if a had it been there; then I can map that population map, I have some population map of the some region either whole Kolkata or Kolkata divided into difference parts at the population; then I can have a layer for the population. Nevertheless, in order to do any of this sized of things; I need to modelist must; first of all I mean I understand what are the different type of features and then try to and have a type of a class diagram; the basic way of looking at the model sorry.

(Refer Slide Time: 12:17)



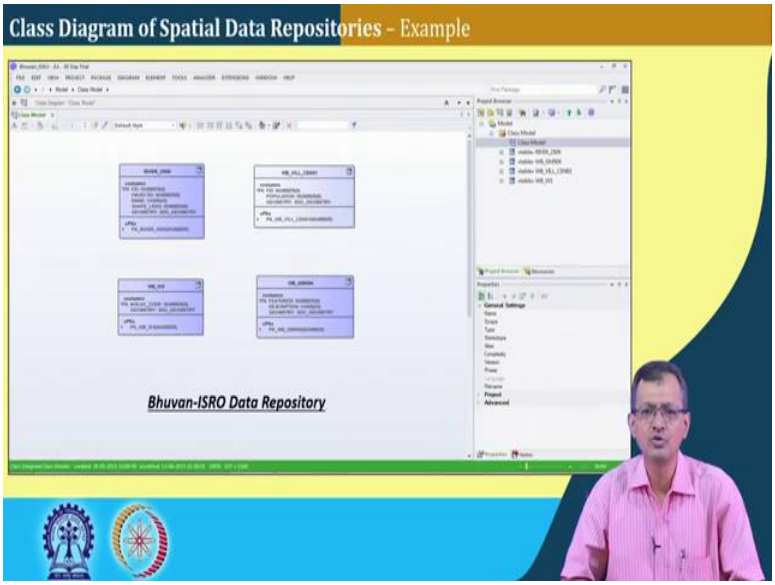

Or in this case we have used a popular tool for which this used a popular tool for having this class diagram. Now this things may not be very visible, but that is the features are not that important rather than the concept is important.

So, it is UML class diagram of spatial repository. So, example logical data model of the things you using a enterprise architect one popular tool we have used. So, there are different type of things; this is a again a Kharagpur spatial data repository, a special data repository of Kharagpur region what we are using here in the say particular case.

So, we have a soil map, a river polygon, a lunc of KGP, drainage type of thing health care type of data. Collected to different means in number of cases may be collected by different stakeholders per say right. Now if it is there; then we have to see that how they integrate but nevertheless we are considering that this is a Kharagpur spatial repository, they have taken care of those things; so these layers are the Kharagpur spatial repository.

And we have created this class diagrams and these are the feature says which are there; everything has a concept called shape which say that what is the geometry type. Like polygon, poly is geo geometry of this river is a polygon type, this is a soil type, this is polygon type, river is a polyline type, drainage network is a geometry may be polyline and so and so forth right and health report or health centers type of things which can be again a point feature. So, any way different type of feature sets are there.

(Refer Slide Time: 14:15)




So, that may be a one depositary; this we have taken another ISRO web that is one portal. So, that is which is again if you Google or you can search that and their whatever their publicy, it is from there.

If we see that we can have a different type of that river, village, senses and different type of datasets right. So, say river network; it may be over Kharagpur also right. So, what is my objective I want to extract those portion into the thing right; I want to have a query which are different type of things. Now the first one is that I need to individually model these datasets; if it is successful then I am now able to do other type of things right.

(Refer Slide Time: 15:02)

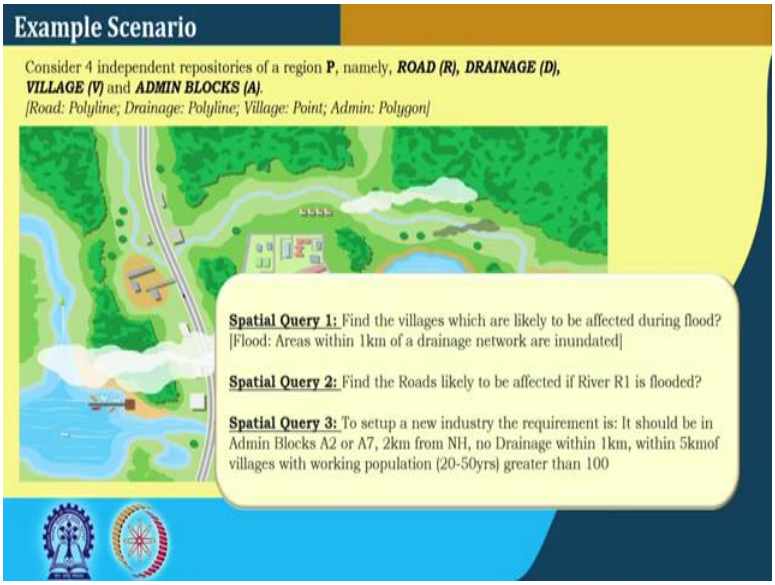

Like again a high very hypothetical scenario; like consider four independent repository of a region of interest P right; namely road R, drainage D, village V, admin block A correct. So, this is the real world scenario means as if a picture; this is a graphical representation, but something like this and then what I like taken from a higher drone or something like that.

And then what we are seeing that there are different type of things road, drainage, village, administrative boundaries. Now, incidentally this could have been collected by say as many as four different stakeholders; somebody collecting road related data like highway authority etcetera. Drainage maybe something that water resource or something hydrological department etcetera, village and etcetera maybe something other administrative things right.

Or in rather or later on we will discuss that will be more than one things right; organization 1 also collect this data, organization 2 also collect this data; then you have to take a that which one I have to take this data right. Nevertheless, so this is my scenario

and these are that way the data is there; somewhere or other this is the picture looks like and I can emulate the things.

(Refer Slide Time: 16:30)




But the that is not my the thing finally, I am looking for some special queries right; like find the villages which are likely to be affected during flood right. So, this is the query we want to see it.

Now there may be a definition of flood what is flood; so for flood I require the elevation and type of things, but say simple what we say that areas within 1 kilometer of the drainage network are likely to be inundated. So, that is my hunch on the thing right. So, that way I can have; so 1 query, query 2 maybe find the road likely to be affected if the river R is flooded right.

So; that means, now from the river; I need to find out the flood network and the flood network; I need to find out the a buffer by which I can have the overlaying on the things. See find the roads likely to be affected if a river R 1 is flooded. Now R 1 flooded means I my definition of flood is 1 kilometer buffer of the things; now this becomes a polygon; now this polygon need to be overlaid over the road network right.

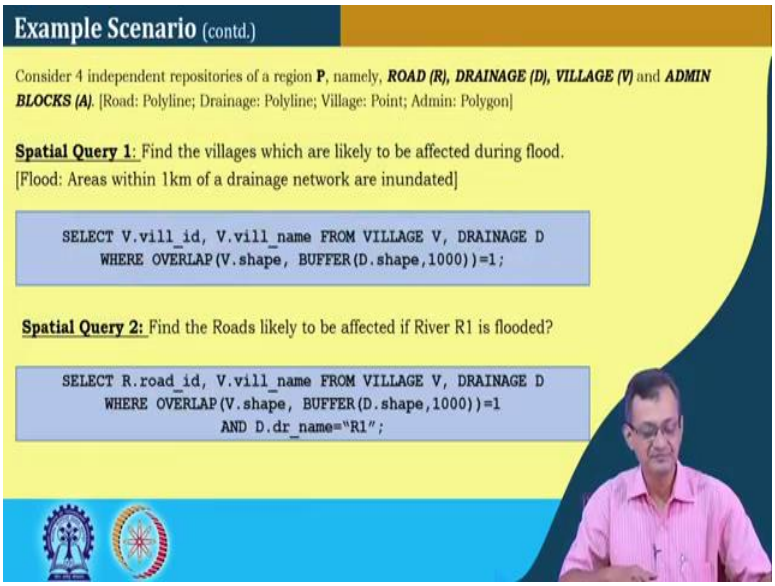
So, once it is overlaid over the road network; then it is then I can find out that which are the river, which are the roads over which this buffer is this polygon is affecting, the flood is affecting right.

Now, you see there are multiple operations which you need to go into the things and there may be another thing; to set up a new industry requirement is it should be in a particular administrative block all right. 2 kilometer for any national highway within 2 kilometer ah, there should not be any major drainage within 1 kilometer.

And within 5 kilometer of the villages with worker population of 50 etcetera; see there are multiple party come into play. One is that administrative block, may be one stakeholder there is drainage network, there is a road network and there is a population map; I require the population map because I require a skilled labor or unskilled labor whatever right.

So, these requires a multi layer modeling; now these data are being collected by different parties they need to be model individually. This model needs to talk to each other and then I am to find out a schema or structure which can harvest this data, then I can extract the data in an and harvest the data right; then I can achieve this resolve these queries. We may not be solving these queries immediately, but what we want to put this there is the need of looking; modeling the data sets like this right; we will be resolving queries into the when we go for spatial query resolutions.

(Refer Slide Time: 19:38)



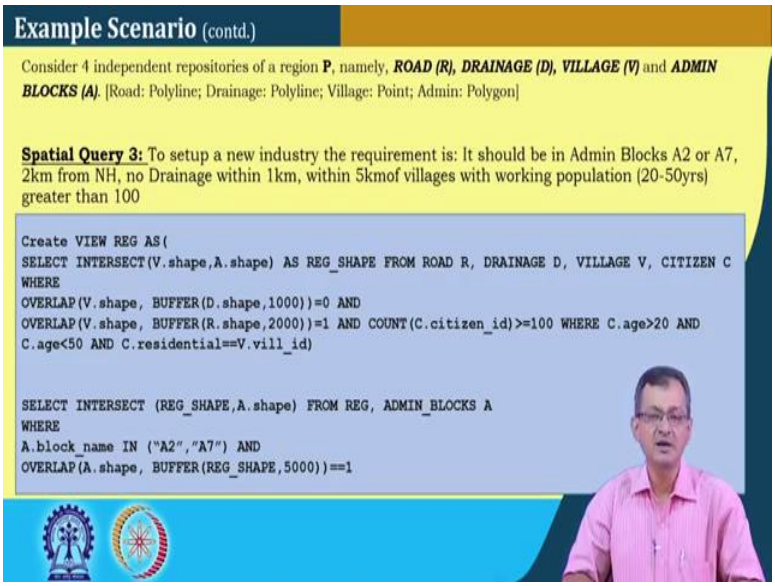

But nevertheless, if we try to quickly look at that; so, spatial query find the villages which are likely to be affected during flood like select V dot village id; V is the village

table, V dot village name from village V drainage D; where overlap V shape buffer this equal to 1, see this overlap is a function topological function right.

We have seen last two last lecture I believe, but in this one. So, overlap is a topological function. So, we a particular V shape village shape; if the village is a village is a polygon and the buffer of the D shape drainage along with that buffer 1 kilometer or 1000 meters buffer; if it is overlapping then we are telling that that village is affected by flood right. So, spatial query 2 find the roads likely to be affected if river R 1 is flooded right. So, if the if river R 1 is flooded then it is. So, the query is select R dot road id V dot village name from village V drainage I, where overlap V dot shape and buffer D dot shape equal to 1 and the drainage name with the R 1; thus the drainage should be R 1.

So we see that we have different depository come into play how do I; I can do this once that is in the my database right. If it is in your database and if these things are supported like overlap or overlap function is supported; you could have queried, but these are with different repositories or different type of database. So, I need to first their model or the schemas, schema level integration, then create a database schema and put into it that exactly what we are trying to put; impression phone.

(Refer Slide Time: 21:31)

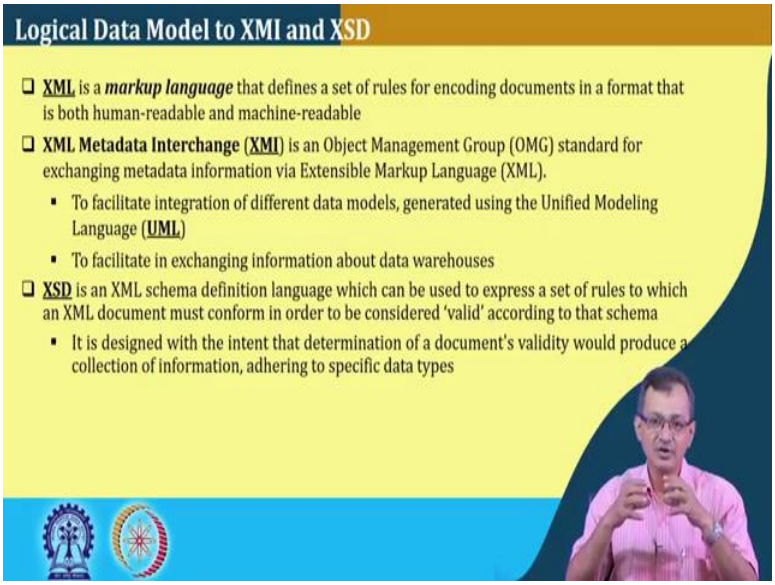



And this is little bit complicated like the third query which it create a new industry; it should be admin block 2 or 7, 2 kilometer per from NH, no drainage 1 kilometer

etcetera; this is little complicated query. But nevertheless if it is in a database; we can create different way of handling this right.

So, it not only require overlap, it requires things like count because there is a citizen count is there; there is a intersect of a particular operation. So, there are topological and other type of SQL operations which you need to do.

(Refer Slide Time: 22:17)




But the overall what we say moral of the story is that I need to have these different data models and somewhere other need; need a way to integrate this data model otherwise how do I have a connection between these data models right. So, as we have already know just to have XML is a markup language that defined set of rules for encoding documents in a format that is both human readable and machine readable right this is the XML.

So, XML meta data exchange; a meta data interchange or XMI is an OMG standard for exchanging metadata information via XML right. So, see initially I for schema generating schema integrating schema; I require the metadata not the actual data. So, this XMI has been that, to facilitate integration of different data models generated using Unified Modeling Language called UML right.

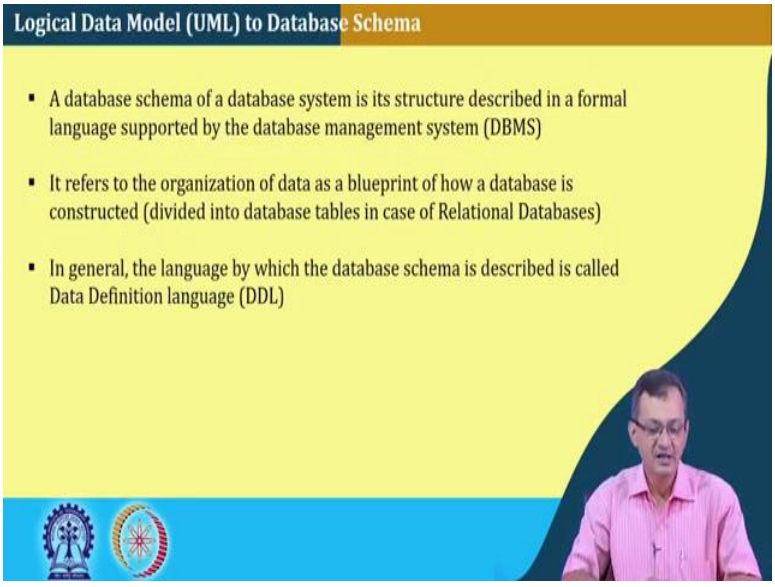
So, UML I might we might have used UML for making the model that is the standard. And its facilitating and exchanging information about the data even data warehouses; we

are not going to this spatial data warehouse there, but says are no, but it also facilitates. XSD is a XML schema definition language that can be used to express set of rules to which the XML document must conform in order to consider to be a valid according to that schema right.

So, what we say that the like database has a DDL of that Data Definition Language or the structure of the things; like here also we have the schema of which it defines the stuff like it is designed with in with the intent to determine of a document validity etcetera. So, we have I am not going to the XML, XSD and any other things I believe you have already done, but otherwise we will we will look into that lecture; we will with service oriented architecture right.

So, in other sense we require a XML type of stuff or more to say we will see later on that is a Geographic Markup Language or GML type of tough and which can handle this type of things.

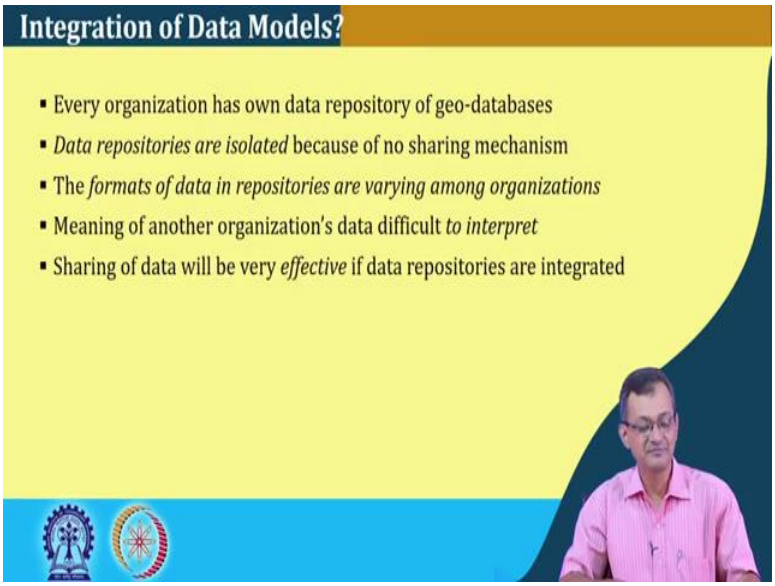
(Refer Slide Time: 24:33)




Now, if we have a logical data model UML to database schema is our basic objective. So, database schema of it of a database system in a structure described format language supported by the DBMS; it refers to the organization data as a blueprint and how the database is constructed like.

So, it gives us the as a blueprint and in general the language by which the database schema is described is the DDL of the Data Definition Languages that we all know right. So, from the these data models or EML; I need to find somehow come to this a unified DDL, where we can fire all those queries what we have seen couple of slides back.

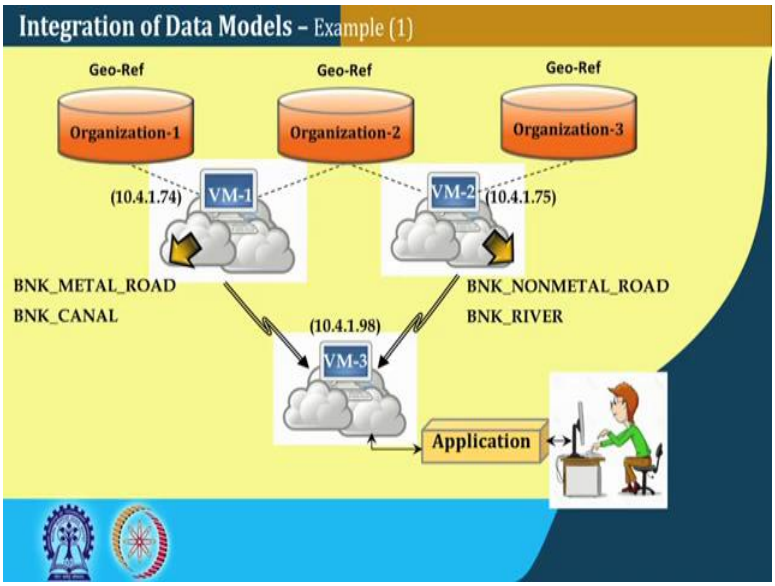
(Refer Slide Time: 25:21)




So, integration of data models; so it is a big challenge every organization has own depository and geo database. Data repositories are isolated because of no specific sharing mechanism of an isolated; not always the isolated. The formats of the data repositories varying among the organizations because of their own organization things meaning of another organization data difficult to how to interpret finding a that what does it mean that particular data set etcetera.

Sharing of data will be very effective if the repositories are someway integrated logically right or if I can integrate at that meta level data level or at the schema level that will be much effective.

(Refer Slide Time: 26:05)

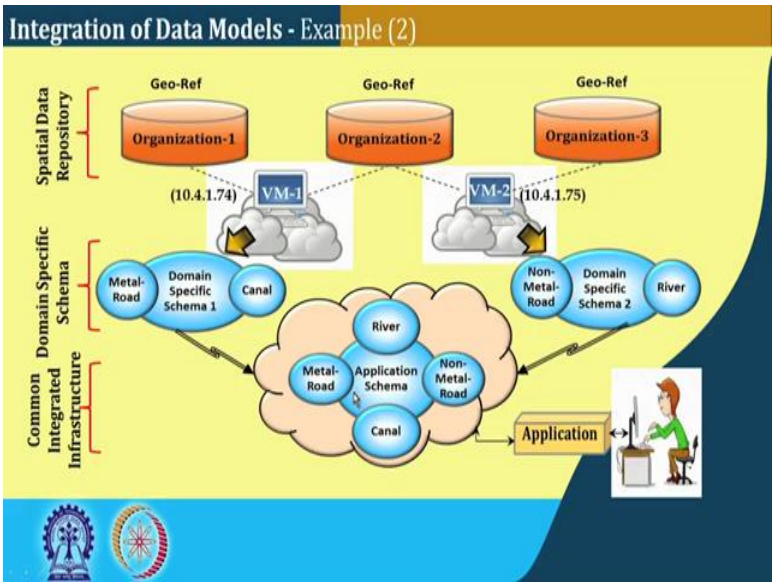



So, we have organization 1, organization 2, organization 3; these are basically a again a example scenario what we have taken; some of the things what we worked at or working at a IIT, Kharagpur we have taken.

So, the data geo reference means it is the coordinate systems is in place and this is some of the we hosted them in some of the virtual machines. So, that is why V M 1; this can be normal machine also; this has some IP etcetera. And these are some of the data sets we have taken from some reference data sets etcetera like this is a metal road, canal of a particular region; this is a non metal road, river of one.

So, this is these repositories have those things and it had it can be from different organization and I require all those data set. And finally, queries will be done through application on this particular back end data set; like which is the shortest path from something a particular block to particular block right.

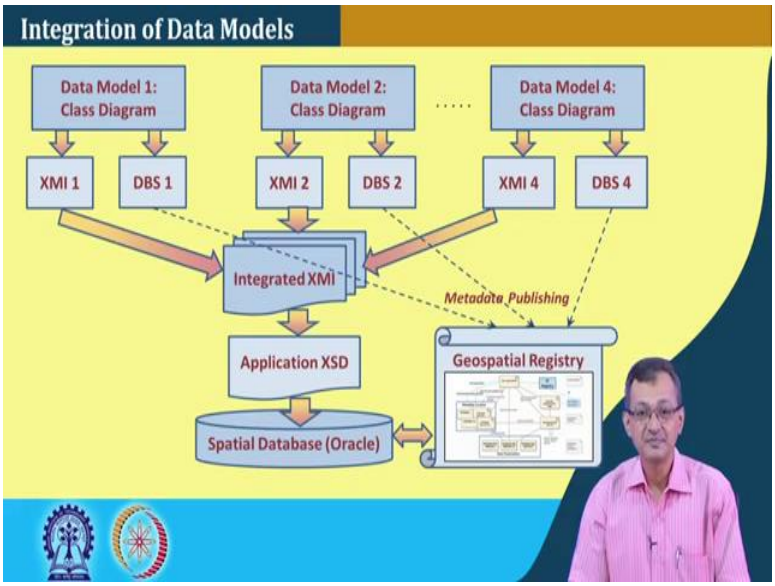
(Refer Slide Time: 27:06)




Now, just to same picture in some other way let us say; these are the repositories right spatial data repositories these are domain specific schemas right. So, for that particular repository they have different domain specific schema and we need to achieve some sort of a unified schema to query on these things right. So, these are some domain schema 1, domain schema 2; the in general there will be n number of schemas.

So, somewhere I; I need to have all our let us consider as done through standard modeling like UML or so and then you need to achieve this type of integration right; so that that particular application can done.

(Refer Slide Time: 27:53)

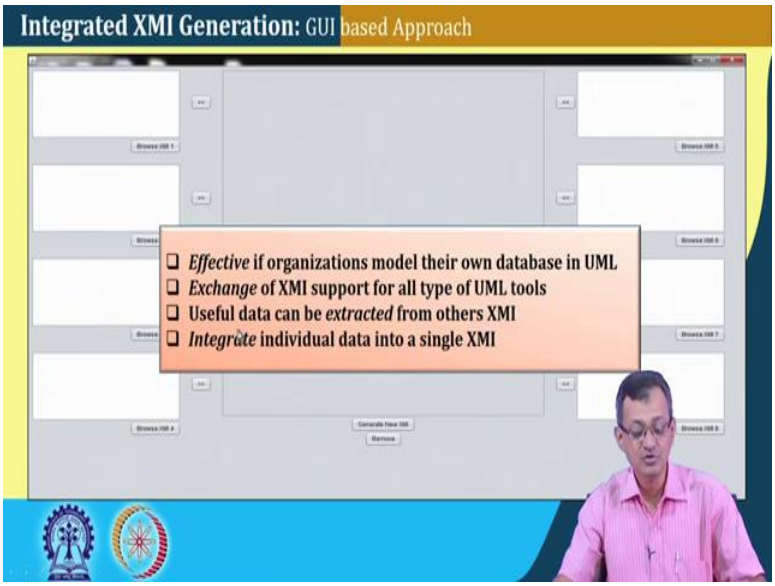



So, coming little bit down into the things; so it is has a class diagram all right like UML class diagram everyone has a class diagram. We have a XMI conversion of this class diagram; UML to XMI and also corresponding that database system 1, 2, 3 and then we what we are looking as this XMI as the integrated XMI; so that it how says this.

Then we have a application XSD; application in XSD mean which this particular system has a integrated way can understand the things like; which the user looking at that integrated X; XSD or the application XSD can work on and we can have the spatial database. It can be spatial oracle, post craze if any spatial database etcetera all right.

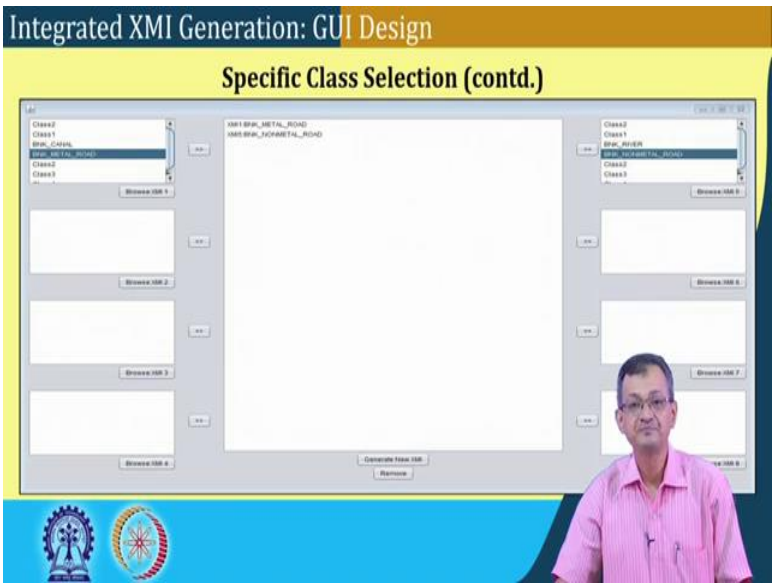
So, and we another thing which require we are not coming immediately written into nothing; there is a registry populace populating a registry will be a required right; how do as the user find out that who is hosting which data and type of things. So, that is a registry service we should come into play. So, we want to achieve this from a UML model to this XMI to XSD and integrate the things right.

(Refer Slide Time: 29:18)



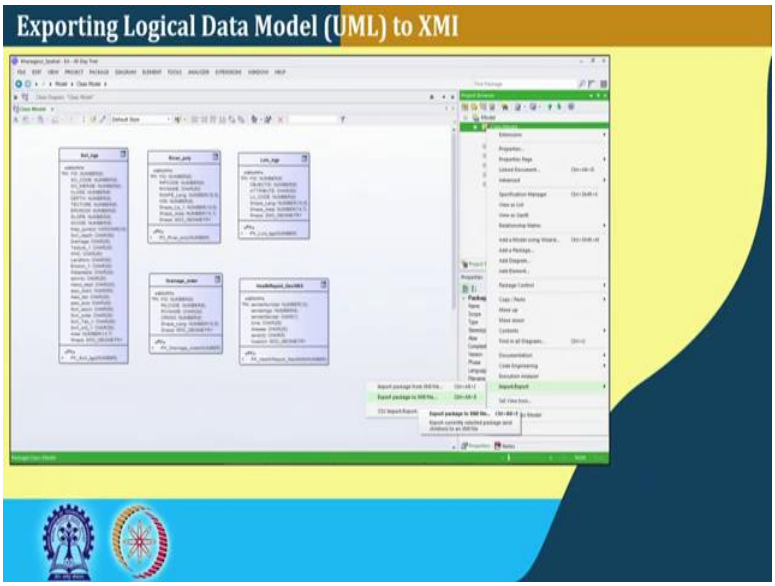

And there can be different way of integrating XSD; one what we there are some of the user interface wise. So, effectively organization models, their own data by UML like the domain expert knows how things will be there. Exchange of XMI support all types of UML model, useful data can be extracted from the others XMI and integrated the data into single XMI.

(Refer Slide Time: 29:42)




And it can be like this, so this is one model this is one data model and we extracted these other things to as a unified data model all right.

(Refer Slide Time: 29:54)




So, we if we have this again we came back to that same previously that UML model of the Kharagpur. So, if we have this then we can generate this model. So, what we require at the organization is this how to generate the UML models. So, what we are trying to discuss today is primarily how I can model the data, how these different data models can be integrated by using technologies like XMI and XSDs and then how to generate a schema out of it right.

We will continue this discussion in our next class, where we will see that how whole process goes on into the things. What are the different steps; it may not be going means very detailed into the process, but I believe that that is the definite steps how it will goes into the thing that are modeling to the schema will be helpful those who are interested in working into the things.

Thank you.