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

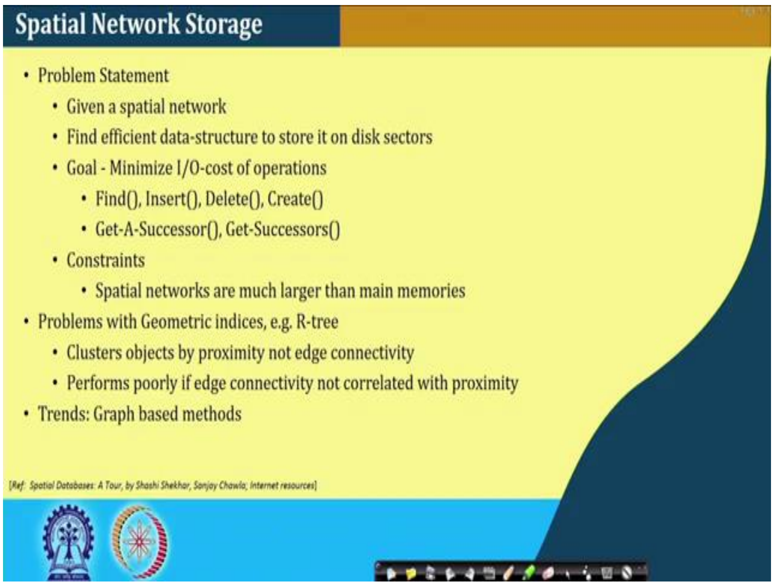
Lecture – 25
Spatial Networks – V

Hello. So, we will continue our discussion on Spatial Networks. So, last few lectures we discussed about that specialty of this spatial networks, how it can be handled, more specifically how it can be queried. Rather in our last lecture, we are trying to look at that how it can be stored so that the fetching of this data will be much in a much efficient manner.

Also in case of large networks, like say total a say a huge spatial network say considered a total transportation network of India or a even a portion of the things, then we need to have a hierarchical approach, because it may not be possible to pull all the data into the in your memory main memory and process the data.

So, we need to go for a hierarchical approach like we have seen that boundary nodes then finding the minimum thing from the building the boundary nodes and then expanding it from the source to destination. So, this type of approaches are possible. So, today we will continue with those things specifically look at how these storing the things etcetera effects the things.

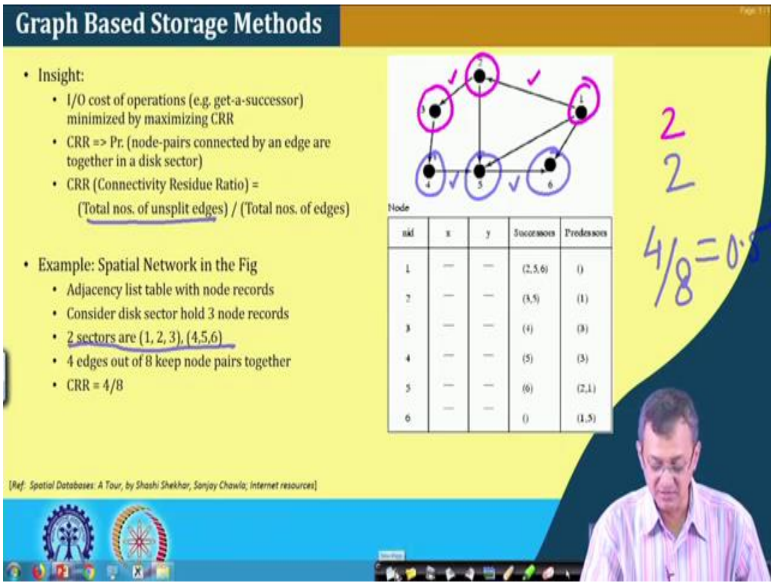
(Refer Slide Time: 01:38)



So, just to quickly little this overlap with our previous lectures so, what we have seen that our basically goal is to minimize the I/O cost operations right. So, specifically we are the type of operations what we look at the find, insert, delete, create and add this is a spatial network we look at the things called get a successor, get successor, get predecessor, get set of predecessor type of things.

So, it is a graph and given a node, I need to know all those things to do some meaningful calculation right. So, constraints spatial networks are much larger than the main memory that is the major constraints. So, that means, we need to have a lot of I/O access. So, cluster objects by proximity not by edge connectivity we discussed on that performs poorly if edge connectivity not correlated with the proximity right.

(Refer Slide Time: 02:33)



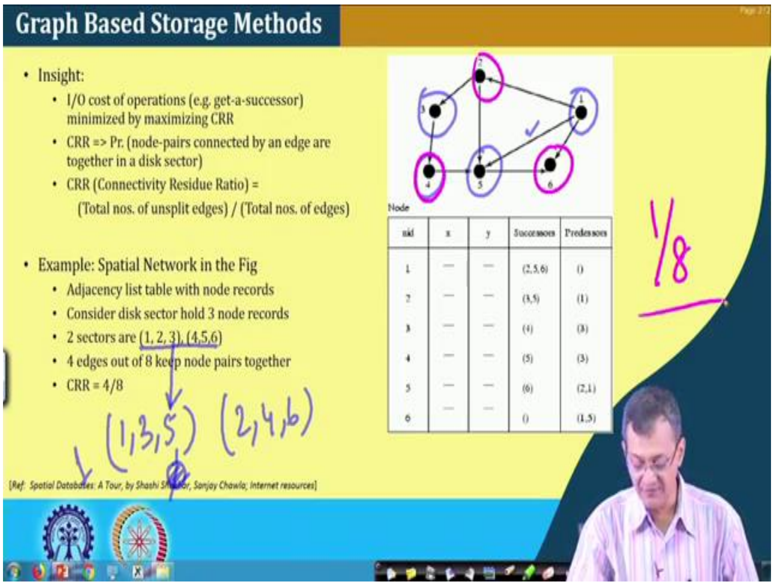
So, there is graph based methods are one of the favorites and this we were discussing right CRR, we defined or there is a thing called connectivity residue ratio residue ratio, total number of unsplit edges total number of edges right. Like if I have this typical graph right and this is the connectivity diagram and we have these are the different nodes and their successor predecessor, there was some typo in the things that is keeping a part. So, what we see? If I if a so, the in this example adjacency list matrix of the nodes considered two disk sectors let us consider two disk sectors with three node records each right; so, one is a 1, 2, 3, continuing 4, 5, 6.

Now if we look at that number of unsplit edges so, so; that means, I take this at one point of time 1 2 3 one 4 5 6, so, if it is everything within that 1 2 3 means could have been the best thing. So, number of see if 1 2 3, so, unsplit edges is 1 to 2 one unsplit 2 and these are the two unsplit edges like 1 to 2, 2 to 3.

Now, if I have other call is 4 5 6 so, unsplit edges is 4 to 5, 5 to 6. So, total number of unsplit edges is becoming 4 got it. So, if I take 1 2 3; so, 1 is this is my 1 2 3 right and once I take this is 4 5 6; so, number of unsplit edges in case of this is 1 to 2 there is thing so, this will be there this will be there. So, total number of this in this case is 2 whereas, the other case this is 4 to 5 and this is so, that is number of two right.

For other case 3 to 4 I this is a split edge. So, I mean I need to access both the disk right. What you have seen that considered to a both the sectors two sectors like this right both the sectors. That means so ideally it should be if I get all the things; that means, the maximize the CRR that one fetch I get all the all the things that could have been the ideal thing, but this we cannot do that. So, total number of thing is 4 unsplit edge total number of unsplit edge and total number of edges are 1, 2, 3, 4, 5, 6, 7, 8 by 8. So, our CRR is in this case 0.5 right.

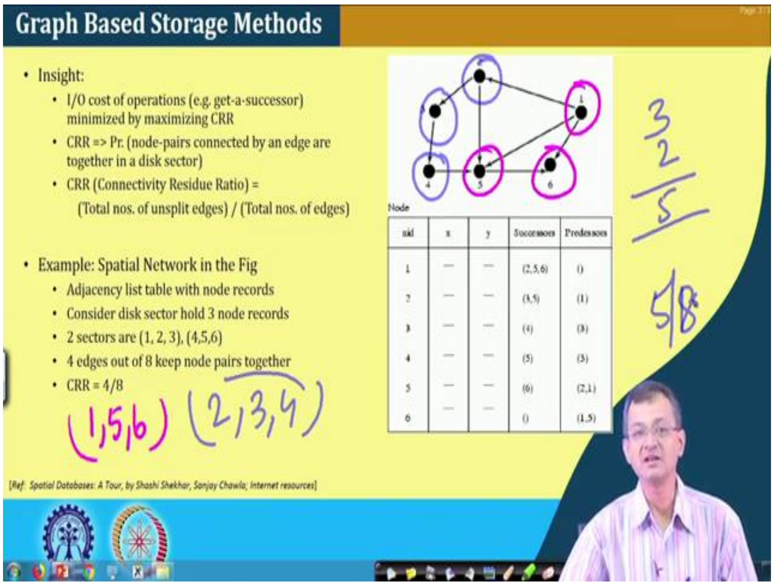
(Refer Slide Time: 05:47)



Now, if I say for arguments sake for calculating sake if I consider say one thing like I have instead of these two 4 5 6 these things if I have say 1 3 5 one sector, another say 2 4 6; that means, even so called even ID's one and odd ID's one right. So, in this case what is the unsplit edge? So, 1 to 1 to 2 is not there, 3 to 5 is not there right. So, what we have for this, I have 0 sort of unsplit edge right.

Whereas for 2, 4, 6 only we have this so we have 2 4 and 6 right, 1, 3, 5 ok, here it is not 0 sorry this should be 1, because we have 1 to 5. This is 1 right so, 1 was right and other was sorry 1 to 5 is there, but 2 5 6, I do not have anything. So, total number of things is 1 divided by total number of is. So, CRR reduces. So, keeping this way things will may reduce right.

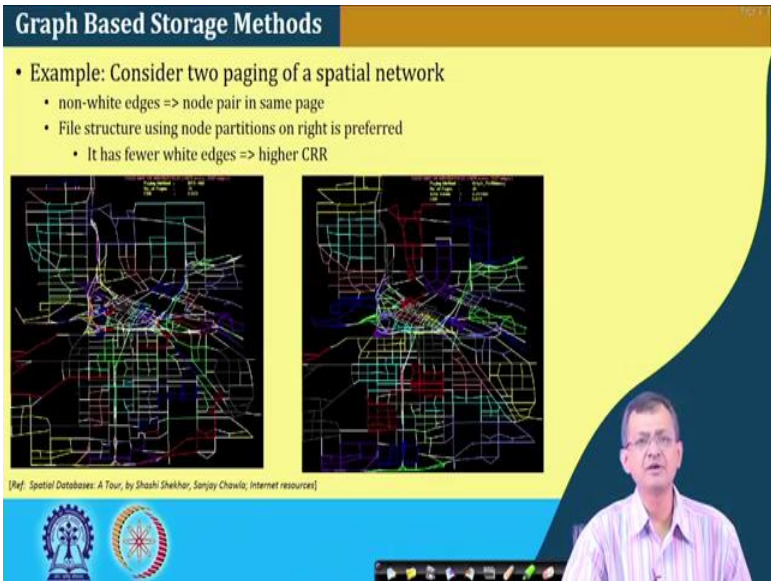
(Refer Slide Time: 07:12)



Whereas suppose I have another consideration, say I have 1 5 6 and another is another node is say 2 3 4. So, over this 1 5 6 so; that means, 1 5 6, other one is 2 3 4. Now what is the count? So, for 1 5 6 one edge two edge 1 to 2, 1 2 3 edges right. I have 3 edges and 4 5 6, 1 2 edges right, 2 3 4, 2 edges. So, total is 5.

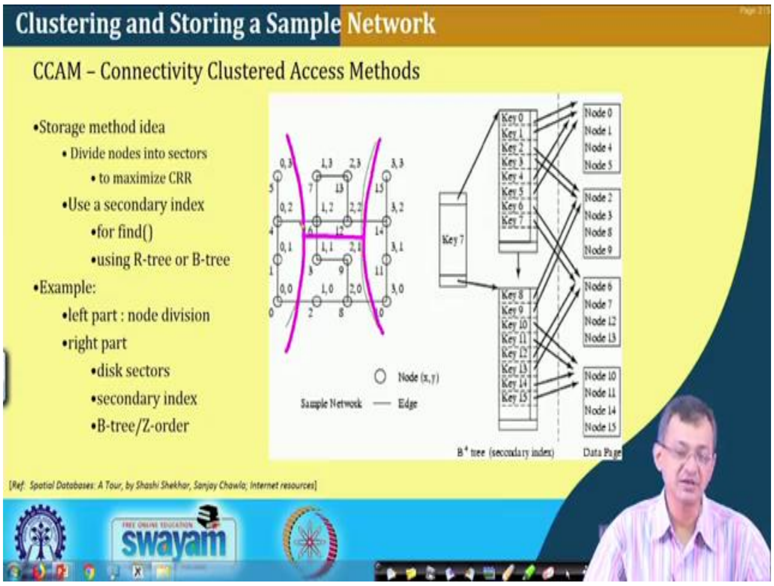
So, we have the CRR as 5 by sorry 8. So, the CRR increases. So, what it is say what it try to say that first of all we want to maximize the CRR and once we can have appropriate selection of these things that I can have a CRR. So, the partitioning of this particular graph while storing into this actual my secondary media or is important right so, that I can my appropriately CRR there.

(Refer Slide Time: 08:31)



So, this is a typical graph of a of a real network considered two paging of a spatial network non-white pages node pairs are in the same page and file structure using node pairs on the right is preferred, because it is a fewer white pages so higher high higher CRR you can achieve CRR. So, that this is by the coloring try to see that because in a real network this type of making this 1 2 3 4 etcetera may not be feasible right. So, I can have different edges which are storing in different colour and type of things.

(Refer Slide Time: 09:08)



So, there is a another concept. Again the basic goal is to maximize the CRR that is the connectivity clustered access methods right. So, the basic idea is that divide the node in sectors to maximize CRR right. Use secondary indexes for find using R-tree or B-tree right. So, these are storage method idea divide nodes into sectors to maximize CRR and to use secondary indexes so, I can have different secondary indexes into the things.

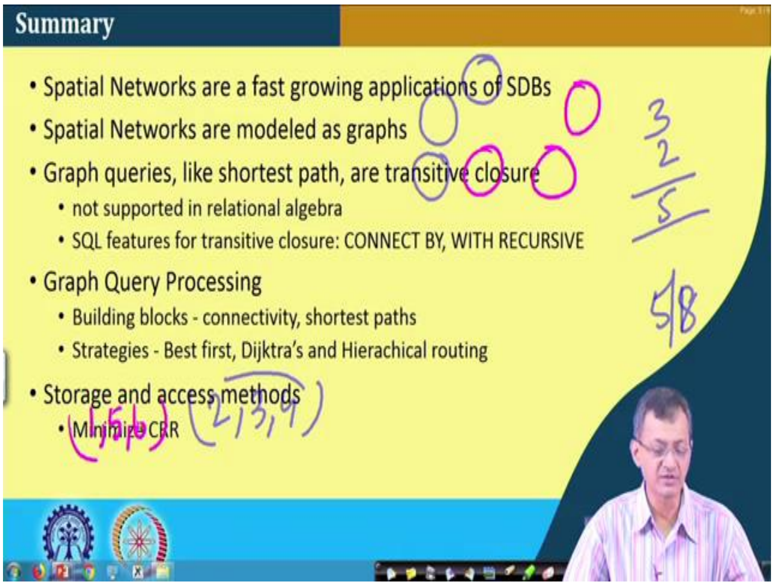
So, in this particular example so, left part is this is what we see there are node division. So, it is so, the division is something like this, one partition is like this, one partition is like this, one partition is like this. So, every node has a node ID right and it has a coordinate also or location x y coordinates type of things right. Every node if you see 5 is 03, 4 is 02, 1 is 01 and this is 00 and type of things right.

So, these are different coordinate points and actual and then we have different type of things like this particular partition is kept in node 0 1 4 5 is in one particular data page, then this 5 6 7 13 12 13 or 13 12 twelve thirteen is one data page and there are in the different data pages right.

So, these are partition in this fashion and then we this can be again kept in particular order either the z order or b order type of things and then I can have different secondary indexes. So, that it helps me in searching. So, I can have different secondary indexes which connect to these type of things right. So, I have different these data pages and then we have different secondary indexes which connect to these type of things.

So, in so, the this partitioning so, this is connectivity clustered access method or CCAM helps us in partitioning in appropriate manner so, that my overall CRR CRR is maximized. So, I have mostly the data which are the connectivity data in the same particular data page. So, that my number of access of that particular sector is reduces or the for finding a particular path. So, these are the things which are there and if we different type of a techniques are followed which are again looking at that storage mechanism part say looking that how is overall storage things can be handled in appropriate way.

(Refer Slide Time: 12:44)



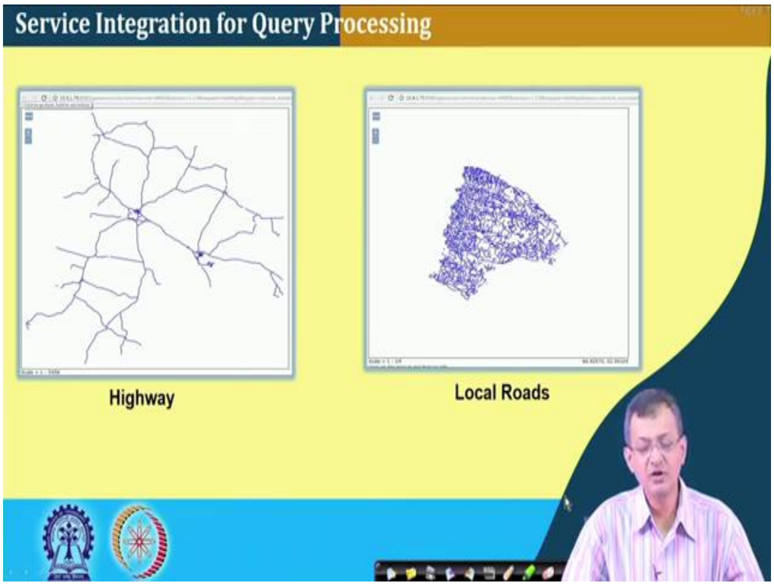
So, if you look at the overall spatial network different different type of application so, this is a wide applications, like spatial network are fast growing application for this difference spatial databases, spatial networks are models as graphs, graphs queries like shortest path, transitive closures are not supported in traditional relational algebra SQL feature of transitive closure connect by with recursive things are what we have seen.

Then there are graph query processing the building block, connectivity, shortest path strategies like best first best path first or Dijkstras, hierarchical routing these are the different type of things and we have also seen that storage methodology where it tries to sorry here should be maximized the CRR value. We want to here it should be maximize the CRR value what we are looking for right.

So, these are the things which we want to achieve. So, if we look at the again just to repeat that the spatial networks per say is have a different dealing than our normal other special databases; one of the major reason is that here where the connectivity dictates over proximity. So, it may be proximity is near or about the connectivity will be in a different way the how the particular network is connected.

And also other different so, the dealing of this spatial network is more of a graph theoretic approach what we have followed or what is a follow in the things and it is it has different type of application for different location based services like, different maps and type of things. So, it is require a special consideration right. So, we will see some few more example queries or a scenarios or which will help us in looking at those networks.

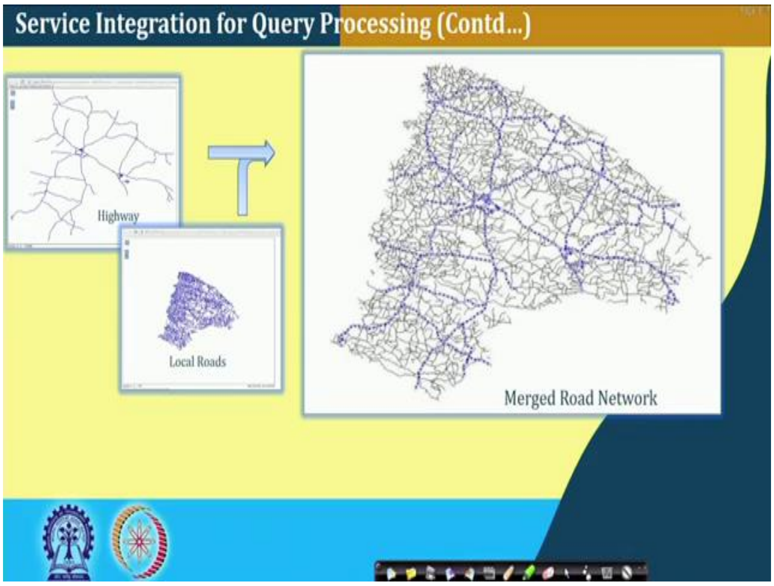
(Refer Slide Time: 15:08)



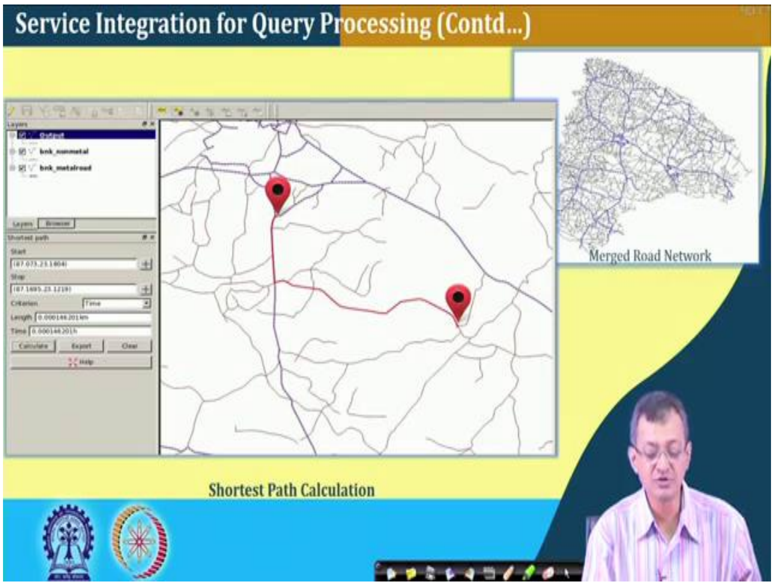
So, if you recollect we have seen this particular diagrams previously also. So, there are two networks right, one is highway one is road network. So, these are fetched using some services spatial services. So, it is a spatial service followed by map service to display the things. So, this is mostly the highway network right. So, this is a area of a Bankura district Bankura district in our West Bengal state and this is a highway network and this is some local roads right it may be village road and type of things.

So, these two roads are there, so one of the predominant application what we look at is finding the minimum distance, sorry shortest path shortest path between two locations. Now, by looking now usually what happened these two networks are maintained and stored by two different authorities this maybe district authority or the state authority there is a local node and whereas, this is maybe the national highway authority. So, two databases should merge and my query should run using those things.

(Refer Slide Time: 16:27)



(Refer Slide Time: 16:41)



So, this way we can integrate this particular thing into a merged road network right so, that two are fetched and merged into the things where the area of interest right. So, this is the merged road network. And then I can which shows that there are two categories of things which are one is Bankura non-metal and Bankura metal road right.

So, these are two types of roads there I can identify querying service querying on those I want to look at those road network only. Say I want to send a consignment or a particular vehicle which we need to move in this two category of road also or some categories so, I can extract. So, what we are doing we are doing all those things in a spatial web service mode.

Now, saying that we now have to have a shortest path between two locations. So, one location select this one and another location select these two points. Now, given these two points I want to shortest path between these two locations right. So, this ideally may fall into a particular road network or if it is not falling into the road network the nearest road network will be considered nearest point of the any road network will be considered the point which is the which is the source and the destination right.

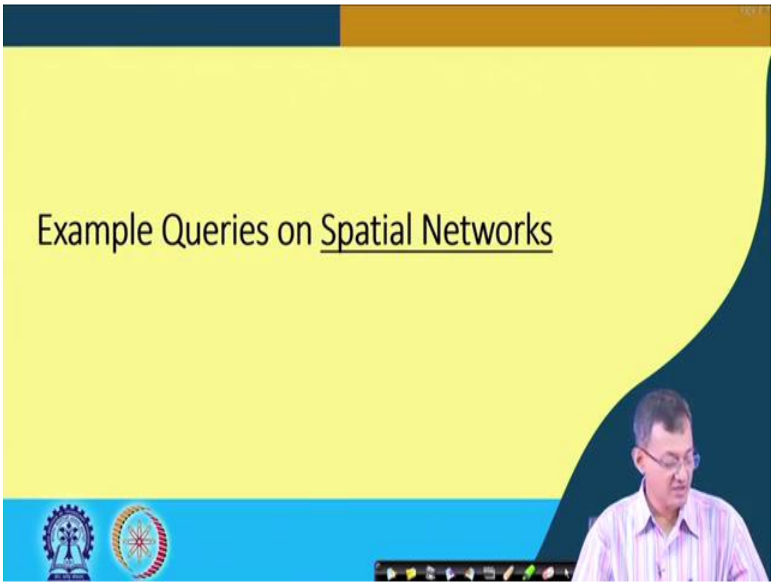
So, this is pushed into their so, this is my source and destination. So, these are two points right and it finds the shortest path between them considering the network under consideration right, which contain some portion of some say highway which is from one source and some portion of the village network or district network which is from other source. So, two source data are combined together and on a service mode and I can find the shortest path between the things right. So, not only shortest path, we can have different type of constraint or we have been different type of network based analysis on the things right.

So, this is how practically we can or how real life we can have some implementation of those algorithms into the play right into the play or looking at the things. Even if you look at that we discussed about the things called connectivity etcetera so, that we can also find out whether there is a connection between two nodes.

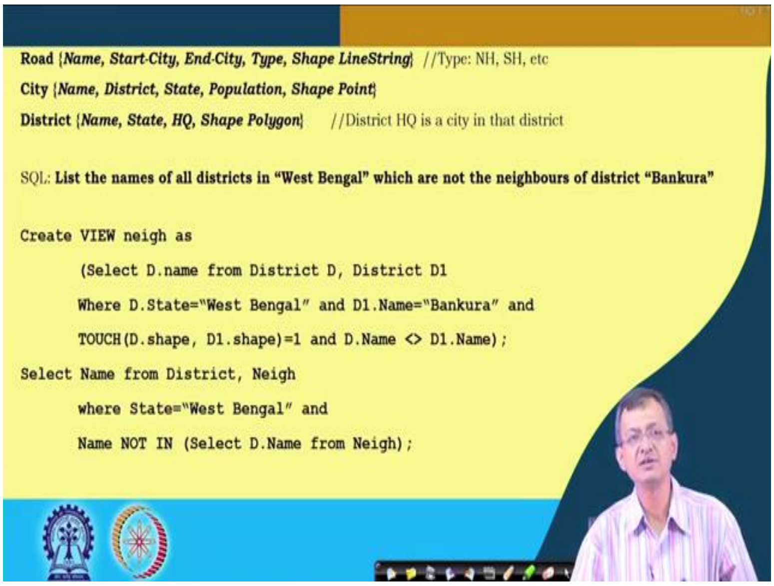
And or what I have time specially in disasters condition like flower etcetera we want to find out that whether there is a connectivity putting that over a inundation layer and then whenever the road as submerged then cutting those road and try to find out from a particular point say rehabilitation center or that is flood management center, disasters management center to particular village centroid or whether there is a connectivity or not right.

So, there is analyzing the connectivity of the process. So, these are this is how practically I can use this and other we are used to with you different type of map services finding the destination, expected time of arrival which is congestion or not based on different map services which are available on our mobile devices.

(Refer Slide Time: 19:55)



Refer Slide Time: 20:11)



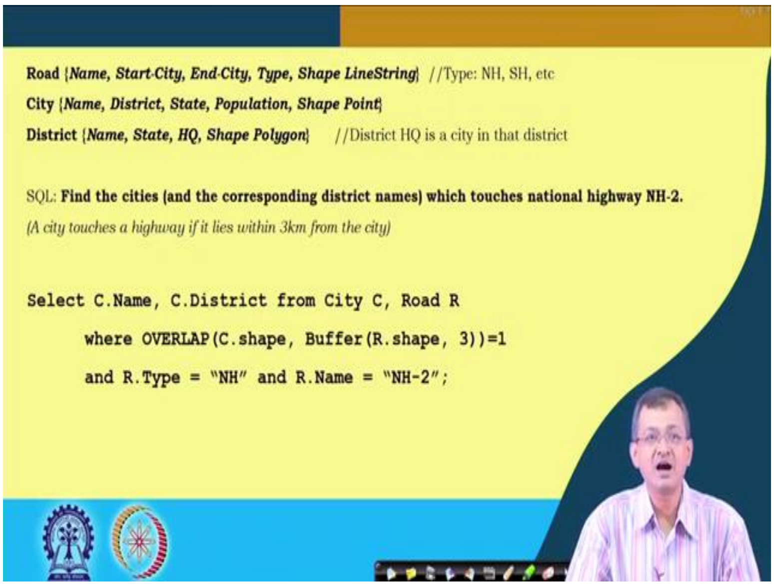
So, next though we have discussed a several queries I thought that will be it will be good that we can have few more queries so, that we can little bit discuss on that more. So, some of the queries which we set something that query like we have schema like road name, start city, end city type, shape and shape is a line string. So, shape is of the line string. So, I have name, road is a name of the road, from where it is starting where it is ending type of road shape is line string.

In case of a city name district state population shape is a point. So, city is taken as a point so, we can take that city centre or the centroid of the city at the point whichever the things and then it is have a it has belongs to some district, belongs to some state it has a population and it has a shape that is geometry feature. Now if I have the queries like list the names of all districts in West Bengal which are not neighbours of the district of Bankura right. So, in other sense I define that neighbor as that it is touching that Bankura district.

So, there can be different variation of the solutions, so, one solution like create view neighbor as select d name from district d and district d 1; where district d state equal to West Bengal, d 1 name equal to Bankura and touch d 1 d state and d name is not equal to the d name. So that means, we want to find out those things which are not touching the thing right. So, this is one consideration we are trying to give.

So, then actually, so I create a view of that things, which is neigh and then we have a select name from district and neigh right. So, should these two things you can have ideally the this capital n or this is small n to have that same your referring where state equal to West Bengal and name not in district name from the neigh. So, all the neighboring state, neighboring district which are touching and then I have a now exclusion states. So, we have all which are not in that particular set right. So, this may be the way of handling.

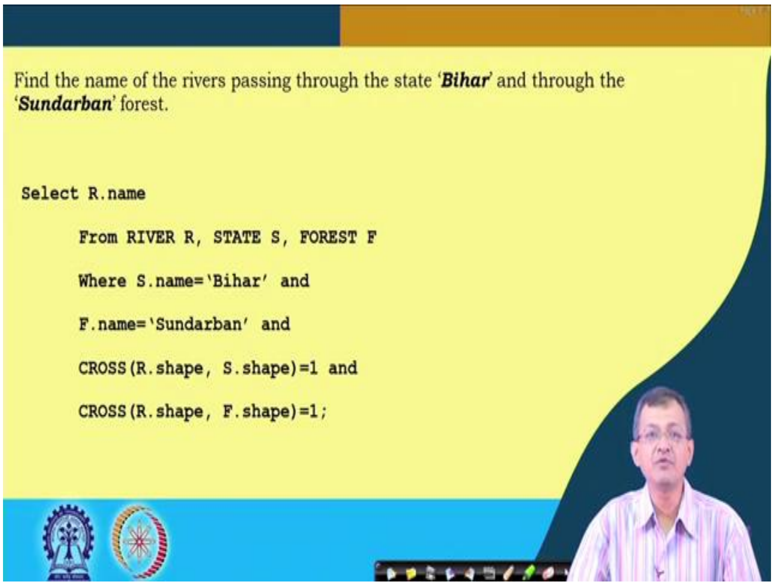
(Refer Slide Time: 22:32)



It this is one maybe one sort of a query, other considering the same type of scenario we want to find out find the cities and the corresponding district names which touches national highway NH 2 right. So, a city touches a national highway if it is lies within a 3 kilometers, we say the definition of touch here is it is within a 3 kilometer. So, like select c name c dot district and city c road r where overlap c shape and buffer of 3 kilometer is 1 right; that means, true and r type equal to NH the road type is national highway and r name NH 2 right these are the thing.

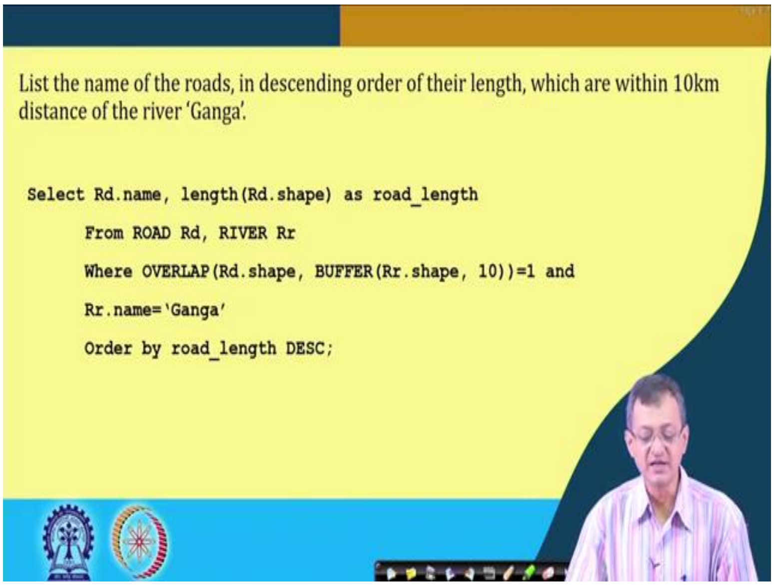
So, here what we see this is my spatial operation in the previous case this was touch was my primarily my spatial operations right. So, here we want to create a buffer and see that it is overlapping or not if it is overlapping then we consider those things right.

(Refer Slide Time: 23:43)



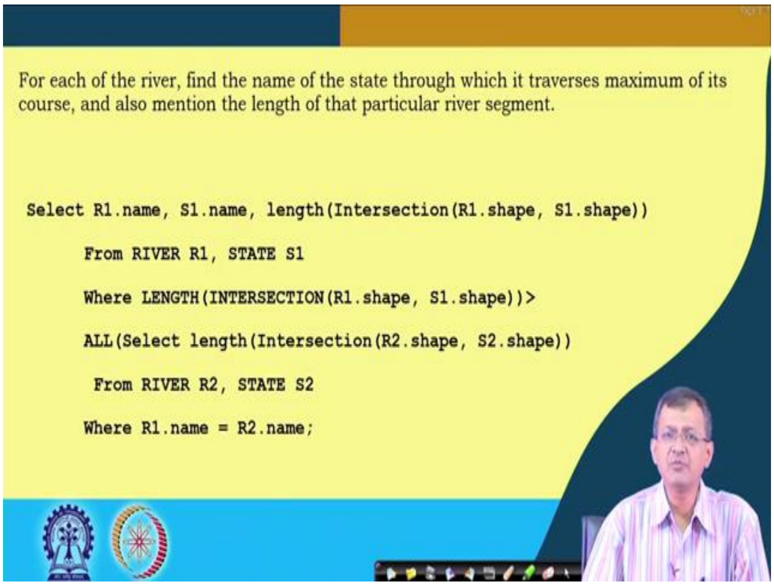
Find the name of the river passing through the state Bihar and through the Sundarban forest right. So, name of the rivers which is passing through a state of Bihar of India and through some through the Sundarban there is a forest so, Sundarban forest right. So, select unnamed from river r state s forest f is name equal to Bihar and f name equal to Sundarban and cross r dot shape r dot shape into s dot shape equal to 1. That means, particular state and r dot shape and f dot shape equal to 1 right. So, that is that both are true and type of things right. So, only thing is this cross is again a spatial operations.

(Refer Slide Time: 24:39)



Similarly, list the name of the roads in descending order of their length which are within 10 kilometer of the river Ganges. So, if this sort of a things are there, we are doing here also if you see that there is a overlap operation right, overlap r d dot shape and buffer r r dot shape 10. So, that these overlap operations of that creating a particular buffer and seeing that river is overlapping or not right. So, river name equal to Ganga order by road length equal to descending right. So, this is the thing what we are trying to do.

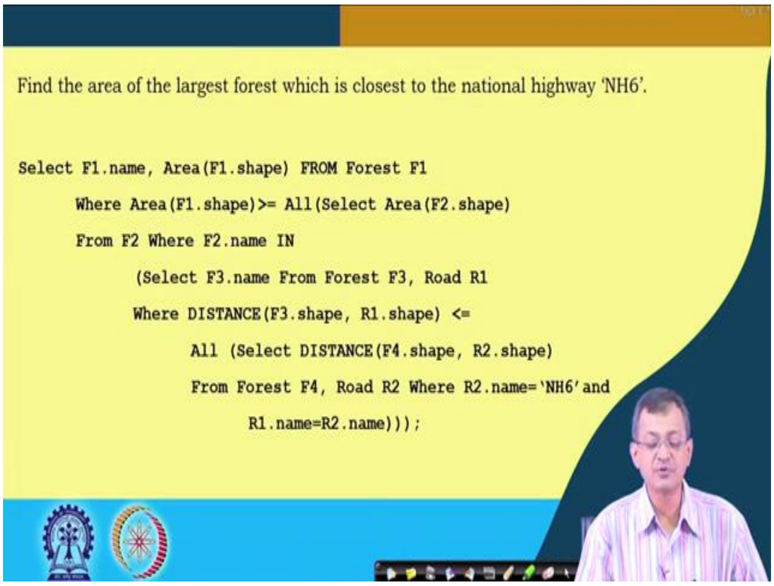
(Refer Slide Time: 25:14)



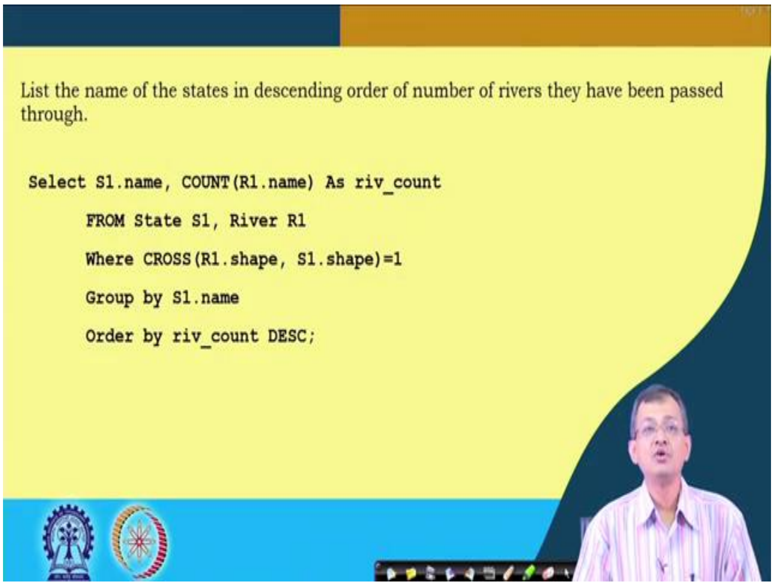
Similarly, for each of river find the name of the state through which it is traverses maximum of its course and also mention the length of that particular river segment. So, same type of things so, little complex queries to find out that for each river find the name of the state through which it traverses. So, taking each river find all the states which is traverse and it traverses maximum of its course and also mention the length of the particular river segment.

So, want to find out that how much length of this particular river segment there. So, this is the way we want to find out the length and then for all those things find the things. So, we want to find the river name s dot name and length intersection r 1 dot shape 2 s1 dot shape. This is another finding the largest forest which is the closest to the national highway of NH 6 you can look at this query.

(Refer Slide Time: 26:08)



(Refer Slide Time: 26:18)



So, find the name of the state in descending order of number of rivers they have been passed through right so, that the maximum count will be that the top and the minimum count will be at the bottom. So, this way I want to save the river right. Select s dot name count r 1 dot name riv underscore count. So, from state s 1 river r 1 where cross r 1 shape and s 1 shape is equal to 1 group by s 1 name and order by riv count descending.

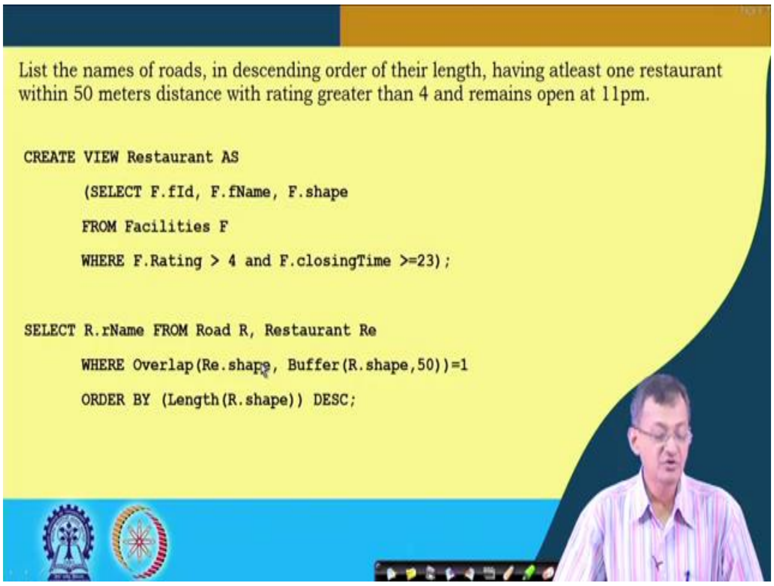
(Refer Slide Time: 26:55)

So, and so, these are different type of example which we see that the spatial operations are there right. These are some where other we did different example we thought of giving few discuss few more example to make that how these all are spatial networks type of things. This is another schema where we have a road, road ID road name, road length, speed limit and shape is polyline facility so, facility ID, facility type, facility name, opening time particular facility closing time, rating, connecting roads.

And shape is polygon a particular facility right eye camp etcetera. What we have seen and academic buildings, building ID, f name; name of the buildings shape polyline etcetera things are there, residential building similarly and there can be a gate, gate ID where we are type of vehicle opening time, closing time traffic direction and its a gate is a point and its look at the thing.

So, facility may includes bank, ATM, hospital market swimming pool etcetera, r type can be r b type can be r b type can be staff u g, p g where we have different type of types of a quarters and that like we can have a type quarter, b type quarter, l b s hall and this different type of things are there like here residential building what we say like r b type it can be staff or u g building and type of things and r b name can be different type of a quarter right. So, this these are the different consideration.

(Refer Slide Time: 28:41)

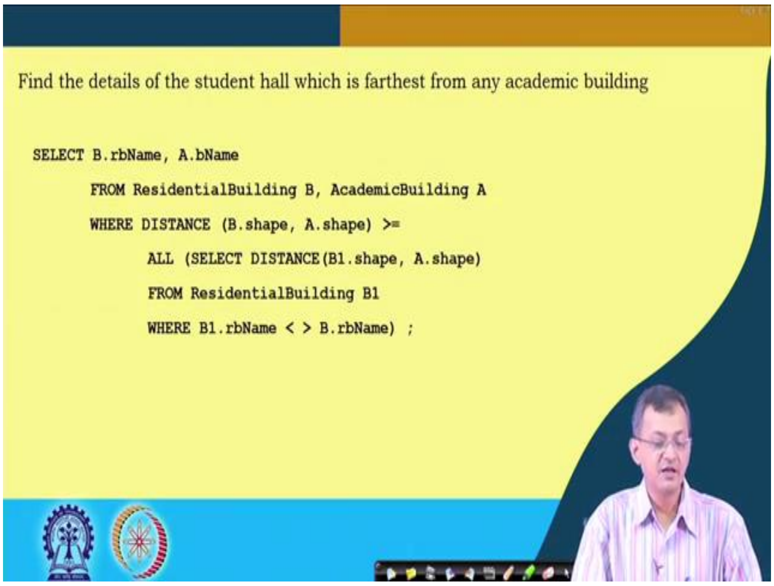


So, if we tried queries like list the names of roads in descending order of their length having at least one restaurant within 50 meters distance with rating greater than 4 and remain opens up to 11 p m. So, I want to find out a restaurant with rating greater than 4 and remains open at 11 p m with within 50 meter distance from the at least 1 restaurant within 50 meter distance with rating etcetera.

So, given a name of the which are the roads etcetera. So, create restaurant a so, I create a view we have create a view of restaurant as select f dot f i d f dot name and f dot shape from facility f and what we see that f dot rating is greater than four f dot closing time is greater than 20 open till. So, it should be closing time should be less than equal to 23 hours.

So, there is a small typo, so, it should be less than equal to 23 open till remains open at 11 p m oh remains open I mean at 11 also it is open. So, at least eleven it is open or 11 pass it is open right. So, I do not know that we cannot say that it is up to 11 p m or 11 plus p m. If it is open till 11 p m that can be less than equal to right it depends on that what type of queries I have. Now, select r name from road name etcetera where overlap r dot shape buffer r shape 50 equal to one that is within the 50 meters and type of things order by length r shape descending. So, order by the length of the road in a descending fashion.

(Refer Slide Time: 30:25)

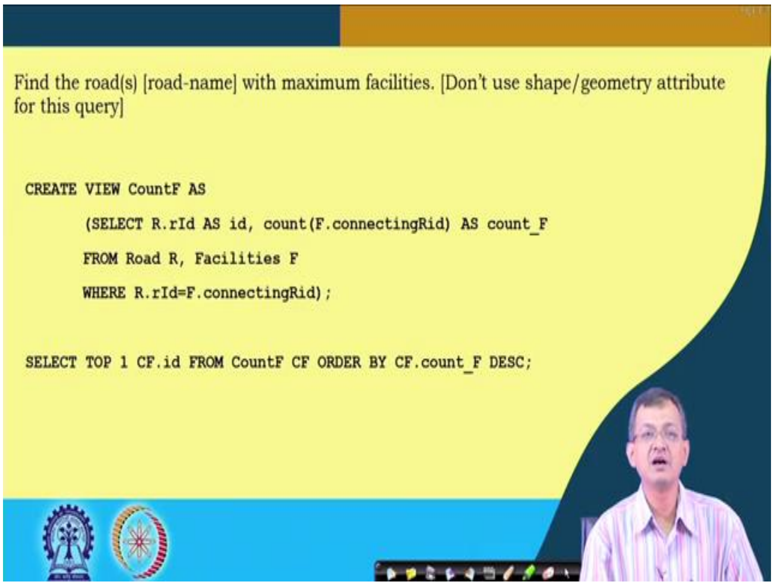


So, find details of the student hall which is farthest from the academic building, want to given a student hall in a typical academic institute, I want to find out that which of the which halls or which are the halls with farthest from the academic building right. So, we academic building we considered as a point feature and also as a point feature that may be the centroid of that particular thing all the type.

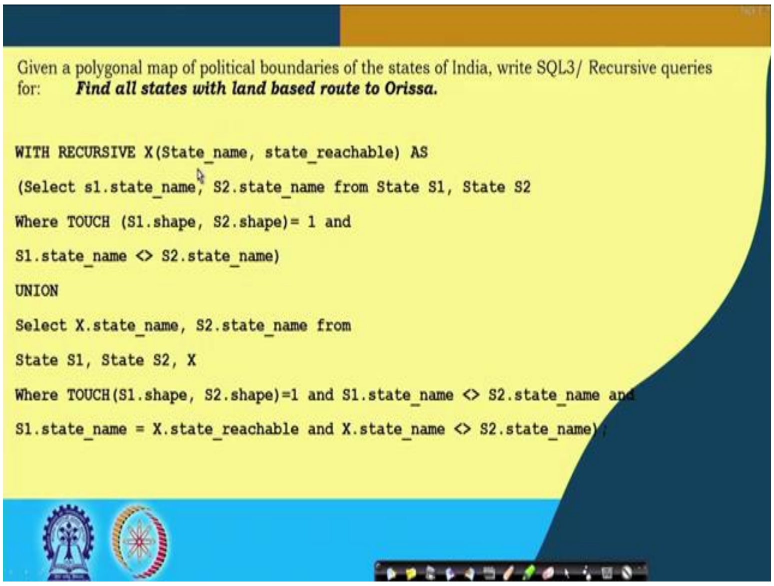
So, b dot r name a dot b name and then from residential building b academic building a. So, student will be hall will be residential building distance B dot shape and a dot shape greater than equal to, so these two distance between them two is greater than equal to select all distances B1 shape a shape from residential building B 1; where B1 r name is not there B dot r b name right.

So, we want to find out all those buildings which are which is not all those building which are the farthest by considering this particular query. So, similarly find the details of the largest p g hall and distance between within the swimming pool, but different type of queries are there.

(Refer Slide Time: 31:40)



(Refer Slide Time: 31:49)



Maximum facility find the roads rode name with maximum facilities this another type of queries there and there is a another things what we have seen is recursive query. Find a given a poly poly map of political boundaries of the states of West Bengal write recursive queries find all states with land based route to Orissa right. So, all states which has the land based route to Orissa.

While recursive; while recursive x state name and state reachable as select we want to do s dot state, name s dot state name from the state s 1 and s 2 which touches this right. So, I can if it is touching then it is directly it can there is a possibility that if the road is there on the if it is touching then it is going through the it is connected right. And the s dot state name is not equal to s 2 dot state name right. So, that those state we considered which has a touching the thing.

And select x dot state name s dot form state s 1 and s 2 and x where touch r dot shape and equal to 1 and it is not that s dot state name not equal to s dot s 2 dot state name and s two state name equal to x state reachable, like x state reachable we have their x dot reachability state and x dot state name is not only. So, what we are trying to see that we if I want to have a land based road either it should be touching, or it will be touching someone or recursively touching someone which has recursively touches and goes to Orissa right. So, this type of things will be there.

Like typically if i see in for our country other than from Andaman Nicobar or some other islands so, all other all other things I can have this will return those things are which are touching right. So, this is a recursive queries we are having few more will share with you. So, what we try to see in spatial networks is that how it differ from the things where the proximity mostly dictates over sorry connectivity dictates over proximity.

So, they requires a separate dealing and also these as a void spread application is spatial data basis or special database related applications right and the other advantage is there or other between is there it is as it is a we can deal with as a graph theoretic approach so, all those graph theory or graph algorithms can be put into play have to find out the things right. So, with this let us conclude our discussion on spatial networks, will be continuing with our discussion on other aspects of spatial informatics in our subsequent lecture ok.

Thank you.