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

Lecture - 39
SDS / Spatial Cloud / GeoViz – IV

Hello. So, we will let us continue our discussion on the Spatial Informatics. So, last couple of classes we were discussing about spatial data science and several application areas, where spatial informatics vis a vis spatial data science can be meaningful or helps in doing. So, we have seen that how spatial interpolation techniques can be deployed we have seen that how predictive models using some Bayesian networks techniques can be deployed.

So, what we tried to look at and also we have seen that how GPS footprint on the other size GPS footprint or means location with temporal spatio temporal data sets can be can help in different type of activities, right. So, the major objective is to look at different type of case studies or different type of scenario, different type of works with spatial informatics which shows that how the spatial informatics helps in different activities.

Like starting from climatological prediction to missing value interpolation to some urban planning or traffic management,or even analyzing human intent and type of things. Like last class we discussed in detail about that different how these GPS footprints or this sort of trajectories. So, to say can help in finding that different movement patterns or predicting different traffic conditions or not only that it may also help in city planning, right.

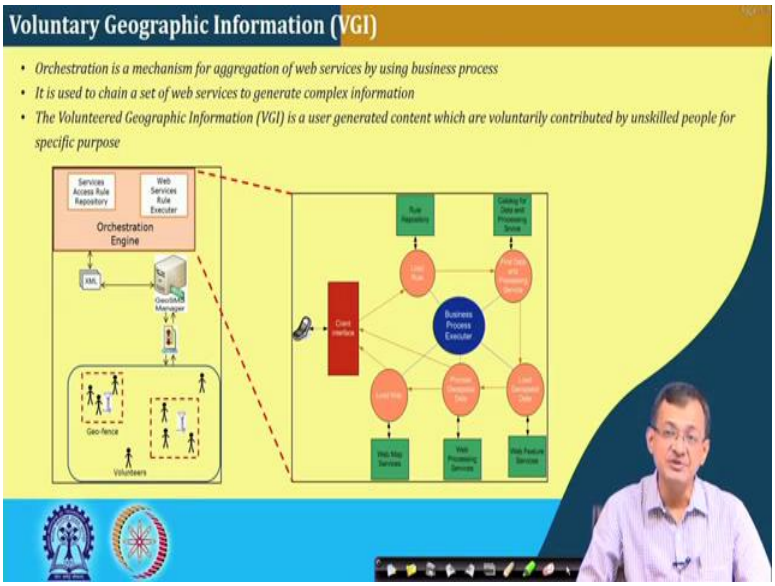
So, these are several activities which has a direct application or opens of different avenues of applying spatial informatics or geospatial informatics in different development planning operations, different case of disasters management and type of things. So, we will see one or two more application area quickly. And then what we like to see as another aspects of these what we want to discuss is that geo visualization and the geospatial cloud.

For that matter we are looking at the geo visualization now and then or what we have written as a geo viz that how to visualize data in different perspective, right. So, it can be

a visualization which a with a real time perspective right you get. And display data and going on or in some sort of a reporting mechanism giving that particular visualization for a particular category of users or particular domain where the users is querying about so along with definitely textual data etcetera.

So, let us first continue with some one or two more interesting scenarios, I should not say detailed case study. It is scenario which will help us in understanding how things help.

(Refer Slide Time: 03:41)



So, one very popular things which is coming up it is a VGI that is voluntary geographic information's right, with huge purification of our GPS enabled mobile devices mobile phones. So, then that it has opened up a scenario where I can have a some sort of a crowd sourcing right or volunteers, which can help me in different managing different scenarios right different situation or events etcetera.

Like if you say if there is an accident, if there are if I can if we can make a platform by which people who are near the accident can share. Basically some information about the accident places like whether it is a fatal or it is a non-fatal, whether it is a block the things how many people are involved and what sort of overall vehicles etcetera.

So, that may help in other administration like police, ambulance in some case fabricate to have their may enable their things in a better way right well equipped thing obtained.

Even it can send a lot to nearby hospitals if there are casualties and type of things right so injured person and so things.

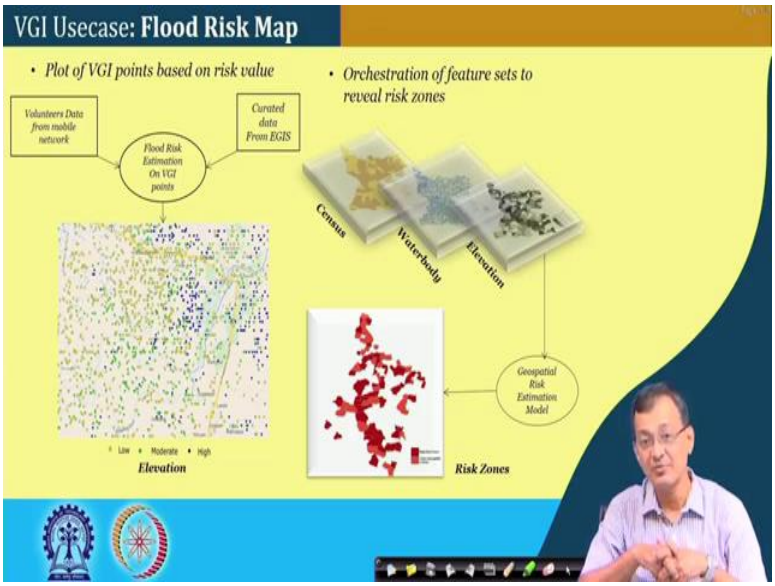
So, those are things even if you see that like some will show you a very small scenario where we did some experiment with a in a flood prone area, where the that people or residents of that region can reporting that how much the water level is there, whether it is dry land or knee deep water or how much water is there, right.

So, it helps in making a flood map and also that administration that which is inundated and not, right. We can do with high techs; high tech things like microwave or radars and microwave imagery etcetera. But this keeps a in-frame to when the person who is affected or who is nearby is reporting which is there. Definitely there are issues of outliers, there are issues of corrects data, how the data need to be corrected and things like that, but there are things.

So, there are so the volunteers who can be registered like I have deployed volunteers like this and who can who some of the volunteers who may not be public in general who are unregistered volunteer, right. So, voluntary voluntarily sharing geographic information is the basic core of vision and found success in different parts of the world. So, orchestration is a mechanism so for aggregating web services by using business processor.

So, one when we have a different type of web services right, like when I get this receive this information from the volunteers in order to integrate to my SDI. So, to say I need to have some services enabled, right. Now different type of services I require some sort of a orchestration between the services right. So, it is used for a chain of web services to generate complex information. VI, VGI is a user generated content which is voluntarily distributed by unskilled people for specific purposes, right. So, it can be unskilled, it can be skilled people also for a particular specific purposes it can be generated, right.

(Refer Slide Time: 07:23)



What we can see that we did a study of a generating a flood risk map, right. So, plot of VGI based on the risk values. So, there are different volunteers, some are registered and some are non registered or citizen in general and we get some information volunteer data from mobile network and curated data from some UGI, EGIS or enterprise GIS platform which able to handle or a SDI.

So, to say which are able to handle layered information and this flat risk estimation on VGI points are there so maybe I estimate. So, the EGIS layers are there along with that from the volunteers that what are the water level etcetera are being there. So, I can instead of giving that 20 centimeter water level, say 20 millimeter or 50 millimeter water level. I can if we can make it like this that water level of the of say need or a dry or something that 3, 4 level.

So, it is easier for the user to mark it and the location is automatically captured by the mobile device. There may be, we agree that there may be little bit of error in the location etcetera due to the GPS accuracy. But nevertheless is gives a broader picture of the things, based on that we can basically create a geospatial risk estimation model we can run.

This requires a domain expert where the this elevation things are there, water bodies or water level layer is coming and there may be a population map over there right. Based on that I can run a functional model we say that which are the risk zone right. Like there are

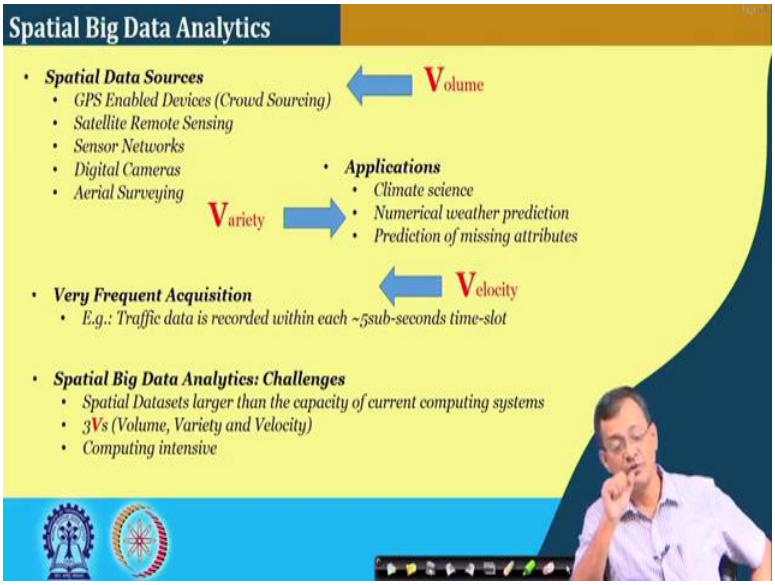
this deep red is the high risk zone, little lighter is the lower is called less risk zone and you can create a map like that, right.

So, orchestration of the feature set to reveal risk zone. So, how to this multi layer how to merge them or output of aggregation of the two layer may be input to the next functional model those needs orchestration, right. So, it may not be the thing like I say I want to increase as we have discussed earlier, I want to increase the road from 4 lane to 6 lane want to calculate. How much agricultural field will be covered by or so to say how much because they are failed field will be affected, right.

So, first of all what I need to know, I need to before the 4 lane things I need to create a buffer of a decision buffer and make out the polygon, right. Once that polygon is there, then I have to do with the overlay with the agricultural, layer right. If there is no agricultural layer as such, but generally vegetation cover layer then I have to find out that which are the agricultural field and type of things, right.

So, once the buffer is there then only I can layer. So, that is a there is a hierarchy, if you recollect our SQL discussions specially SQL discussion. So, I need to have a it is query tree right. So, unless something is calculated I cannot do the next level of things, right. So, that requires a orchestration of the services because these services are independently coming into the things. So, unless they are appropriately orchestration is enabled we may not be able to extract appropriate information.

(Refer Slide Time: 10:54)



So, this is one scenario where this volunteer data allows me to do that right. It can be different things like it can be I can say cyclonic weather or if there is a storm how much intensity of the storm. Even there is during peak summer what are the temperature profile, if there are rainy season, if how much rain accumulation is there.

From there I can calculate that how much drainage flow is there that may not be inundation by the flood. But, the other will be huge amount of outflow from say IIT Kharagpur, it is as of now we have not faced that it is inundated on the for water level. But, there is during heavy rain there is a huge amount of water flow over the this nallas and canals and type of things, right.

So, what happens those discharges goes to some places, right. So, it is again getting so this get accumulated and some point of time it may overflow, right. So, I can have a pre calculation of this, if this is my rain condition. So, this is like to be the amount of water in the larger channels after time t, right. So, this type of functional model I can do, but require orchestration and getting this information can be a volunteered geographic impose. So, there are several cases like what we did as a flood risk map there can be different type of things which you can do.

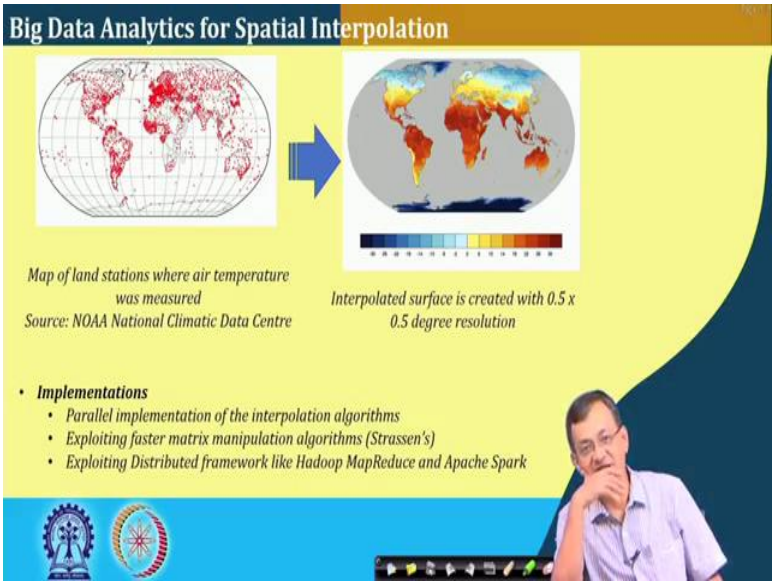
Next, we already discussed is that what we see that a big data challenges in a spatial. So, truly because spatial data is GPS enabled device it is crowd sourcing or satellite remote imaginary or sensor network, digital camera, aerial surveying these are all highly voluminous data, right. Even if you see our mobile data set or CDR data sets, if it is accessibility the huge volume of data, right.

So, it is a huge volume of data, application, climate science, numerical, weather prediction, prediction of missing attributes and these are variety of application. There are several variety of data and that it data comes in a highly in high frequency right very frequent acquisition, traffic data recorded within 5 milliseconds sort or something, right. So, there is a good amount of velocity, so all these three Vs are satisfied by the this sort of data a spatial data.

So, it is a ideal category of spatial data or big data analytics type of things. There are several challenges, spatial data larger than the capacity of the current computing system, right. So, you need to have lot of IO fetches and this data have properties of both

compute intensive and IO intensive right, so it has a both type of things. So, unlike our traditional data, so it is a huge challenge out here, right.

(Refer Slide Time: 14:15)



So, here in big data analysis like one say if we have a map of land stations, where air temperature was measured source is NOAA National Climatic Data Center. So, that can be interpreted in a grid of 0.5 x; 0.5 cross 0.5 degree hub degrees so, to say resolution. So, generate map of a hub degree resolution from this only some of the centers of the whole earth or a large space.

So, there are several implementation challenges like parallel implementation of interpolation algorithm is a major challenge right. So, otherwise it will take a huge amount not only that as most of as we are discussed that most of our algorithms are say different are programmed or things that thinking that the data can be brought into the main memory and this, but this is not true here. So, there are lot of fetch and you cannot get the whole data at a time like even say our school days things like as we had discussed that matrix multiplication.

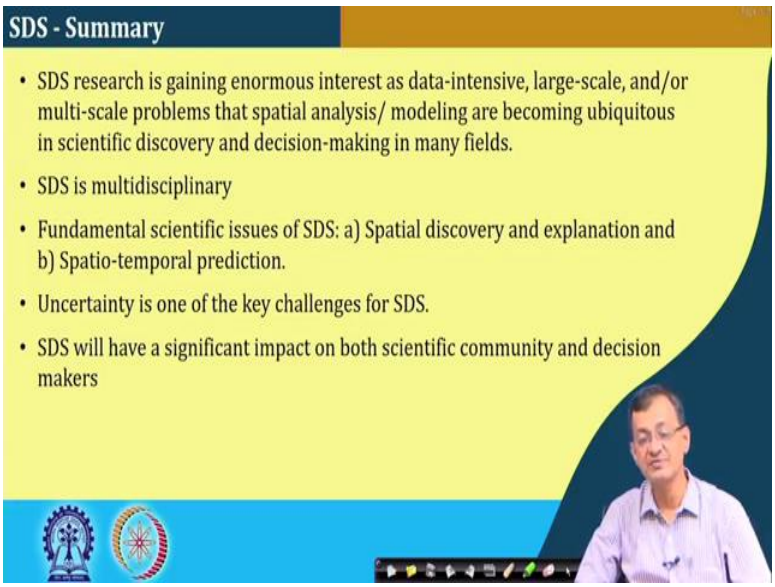
So, if it is a 10 cross 10 a say matrix inversion 10 cross 10 is fine, but if it is 100 cross 100 is fine; 10000 10000 there are your problem, whether your memory supports that if it is 1 million cross 1 million it is a serious problem. Now, if you partition that how to handle that inversion problem. So, exploiting faster matrix manipulation algorithm; so,

one is a Strassen's algorithm and exploiting distributed framework like Hadoop MapReduce and apache spark type of things.

So, how to engage those things entire things is a, these are the challenges with the big data type of operations. So, these are other application areas which are or I to I should say some of the things are still a open research or a means very recent research problems which can be handled by when we use those type of technique.

So, if you if we see that starting from image analysis, manipulation interpretation interpolation or prediction forecasting to sort of a big data type of analytics, there are several research area which constitutes that research goals of our spatial data science.

(Refer Slide Time: 16:52)



So, if we what we have discussed on spatial data science, there are lot of other things to discussed if we try to summarize them. It is a is gaining enormous interest as data incentive, large scale intensive large scale and or multi scale problems that spatial analysis modeling are becoming ubiquitous in scientific discoveries and decision making in many fields.

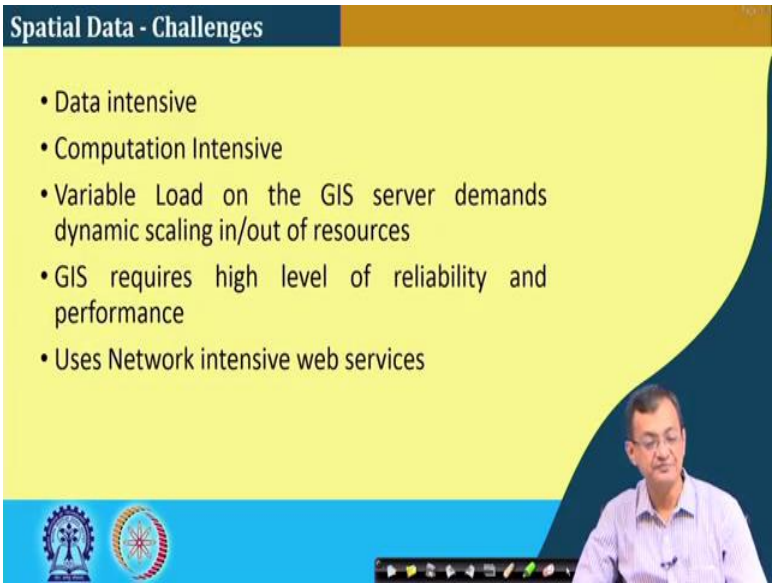
So, what is happening? This spatial data spatio temporal data set analysis becoming a ubiquitous or it is a de facto required for several applications. So, this SDS try to address or try to it is a what you say umbrella structure which take data how to handle this data intensive large scale multi or multi scale problems for data analysis modeling prediction

and type of things. SDS is typically or in general multidisciplinary. So, it is those for techniques etcetera required, but there is multidisciplinary.

Fundamental scientific issues of SDS like spatial discovery explanation and spatio temporal predictions are challenges. Uncertainties is one of the key challenge for SDS. SDS will have significant impact for both scientific community and decision makers right, or I should say is making lot of impact in the both the communities lot of interest in the research and of course, for the decision makers. So, that there are things which has need to be looked into.

So, this is a brief overview I believe that you will find lot of interest, those who are interested in looking at the research aspects of the things you will find out of interest it is.

(Refer Slide Time: 18:36)



Now, let us try to look at some few more aspects like one is spatial cloud another is the geo visualization, though we knowingly or unknowingly we are going through looking at the visualization problem now and then whenever we are discussing. So, that is there inherently, but this spatial cloud is a serious aspects of the things and one which is becoming inevitable.

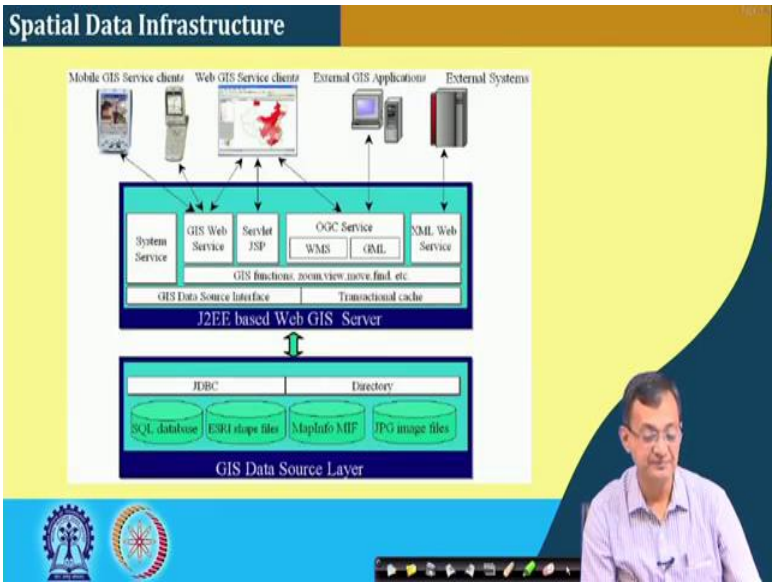
So, what we have seen the spatial data is data intensive. So, huge volume of data; volume whereas, variety and velocity are there it is both IO and compute intensive. Variable

owned load on GIS server demands dynamic scaling in and out of the resources right, say the load on the GIS server are pretty variable, right. So, it is sometimes it is you have a huge amount of computation when you require a huge computing power starting from memory and type of things. But, in some times there are hardly any large requirement, right say routine requirement.

So, what is happening usually what happened that we try to have a peak load type of things right like say disasters management. So, it is not every day or every moment disaster happening. So, there is a, but you cannot ignore that when it happens it happens like anything, right. So, either your infrastructure will be at a peak type of infrastructure, where things will be there or you have to make other mechanisms it.

Now GIS require high level of reliability and performance one other sense like you require that system software and means overall working of the thing; people working around it actually to be very efficient to make a high level of reliability and performance. Uses network intensive web services also the spatial analytics or even spatial data science which has network intensive web services. Because, this data repository has distributed and need to be hooked or need to be fetch data as and when required at a with high reliability or high performance.

(Refer Slide Time: 21:01)



So, this picture we have seen earlier also. So, I have spatial layer at the thing and there are this is that our interoperable layer which constitute a backbone for our SDI also, and

these are different stakeholder or consumer of the things. So, based on their request and type of things it request to the underlining spatial layer or spatial for that matter spatial databases and type of things. So, there are several type of databases like SQL database, ESRI shape database, MapInfo MIF database, JPEG database and type of thing.

So, different type of data bases which accessed by this middle layer and it catered to the higher layer. So, this may be distributed so all the services over the network. So, there is a reliability of the network come into play other than things like that right. So, this is the typical spatial data infrastructure per se.

(Refer Slide Time: 22:12)

So, when you talk about national spatial data infrastructure. So, there is a basic objective of the things to ensure the spatial data from multiple sources. It can be national, state, local government, academia, private sector anything are widely available and easily integrated to enhance knowledge and understanding of our physical and cultural aspects. So, it tells everything, right.

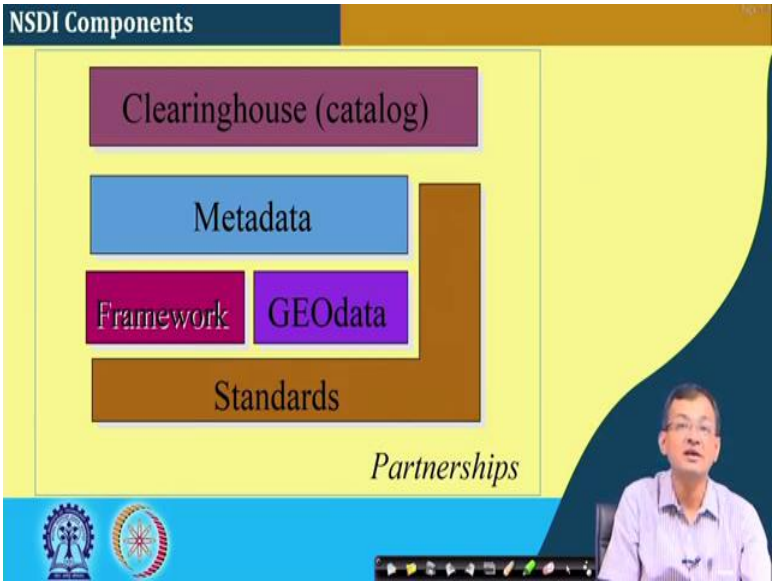
So, you have huge volume of distributed resources or which can be easily accessed and quickly integrable and able to enhance knowledge etcetera. How do I integrate different type of repositories or databases or type of things? I need to know that what is the content of the things right; what is the so to say how what is the model of that data sets right, starting from the schema and type of relationships etcetera.

So, there is a need to model this data set and when we try to integrate we actually create a integrated model right or in our CS parlance we say that it is a integrated schema need to be created which able to answer the query for that particular domain of interest, right. So, SDI is able to house these different repositories schemas or model and type of things.

So, the one part is there in harvesting this schema actually not giving out the data. So, as such if you say that IIT, Kharagpur to say are they are student data schema that may not be a major problem, right. So, it is not sharing that data it is sharing that my student is kept as this is my the structure of the things a roll number name, hall number, fathers name, mothers name and etcetera date of birth and etcetera etcetera, right.

So, that helps the other end to do if the data set at any time required to do that particular develop that appropriate interface. So, when it required it sends a request and based on some agreed upon things, it may be selling the things or government to government there may be only agreement. So, data can be shared easily so that encompasses this NSDI.

(Refer Slide Time: 24:30)



So, if we look at it has a metadata framework, geo data standards and clearinghouse we have seen that, right. So, if you look at that is geo data is a part of it, there are lot of other things which encompasses that. And of course, it require we require partnership or collaboration with other things to look into the how the data set a how to interoperate.

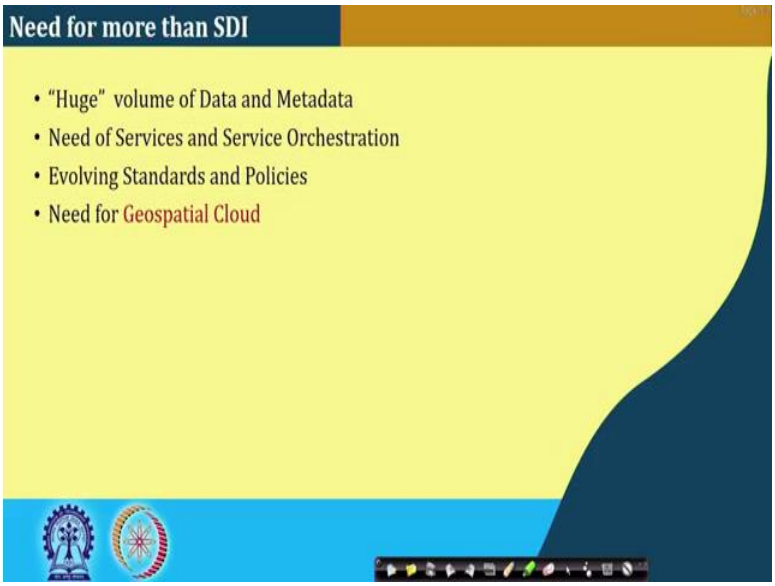
(Refer Slide Time: 24:55)

Now, I can have a sort of a spatial cloud type of things which is connected with several government organization. The spatial data resources, crowd source data, some external agencies data and type of things right, and it different application development, decision analysis, big data framework, visualization location based services and so on and so forth, spatial data analysis and so and so forth.

Now, what this particular so called cloud is giving. So, GIS web service, application web service, system services, GIS data services, GIS, APIs and type of things. So, it gives us a different way to handle things in a thing. So, what we are looking at; so, what if you see what we are looking at. So, it is as if there is a cloud if I can able to hook in I can get GIS geospatial services or I can even contribute to that services. If I am a registered for those type of things right, like I being a district level spatial data administrator.

So, one thing I want to know do that whatever the data is collected I put it on some government portal. So, that the for any analysis etcetera it can be used. On the other hand I may require some datasets from the portal itself, right. So, the datasets which are which can be at a collected at a national level like census data. Like it has been collected at the national level and those datasets I may require to put it into for my analysis. Now, this sort of things is possible if I have this sort of a cloud right. And that exactly what we are trying to see that how this can be made feasible.

(Refer Slide Time: 26:52)



So, if we look at that there is a need for something more than SDI, not only the spatial data infrastructure something more than the SDI what we require. So, huge volume of data and metadata has being collected continuously need of services and service orchestration. Evolving standards and policies and for this we may need a geospatial cloud, right.

So, over encompassing cloud infrastructure which allows me to work on those type of things, work on with different type of data and services. So, what we are trying to look at. So, I like one of our successful cloud or successful infrastructure is our so called quote unquote telecom clouds so to say, right.

If you have a SIM card or a connectivity thing and if you able to connect to this telecom cloud, you can have different type of services; voice, video and different type of services. So, multimedia services and so that; so, what you require a connectivity with the things right. So, here also if I have a like this so if I able to connect it I will be able to different services or I can able to contribute to these services also right. So, there is a need for spatial cloud geospatial cloud.

So, if we see that in today's discussion what we tried to look at is that again few more applications of a spatial data science right. And also we try to see that how different aspects of the things, like if you look at volunteered geographic system it is a different

ball game altogether. You, we are trying to bring people citizen at large to contribute to these information base, right.

So, in this lecture so we try to look at that and also we looked at that how a spatial cloud or geospatial cloud may help in shaping a may give a geospatial services in ubiquitous way. Of course, there are issues of data protection, there are issues of even service protection what services will be enabled etcetera, but there are, they are these are challenges for normal cloud systems also. So, if we can integrate things are like that.

Another major problem what we have major challenge what we have seen is the big data challenge right. Spatial data as pretty large extremely large data sets and handling these things we require some mechanism of big data analytics or thing. It may be some sort of MapReduce, Hadoop framework to handle parallel algorithms are run those things in the data set or some even the spark technologies what way people are using.

So, with this thing what we try to looked in this particular session is that to how this spatial cloud and things can help in a bigger way to make this geospatial technology. So, just doing for a spatial informatics and spatial data science to make more usable to the, for decision makers, users and scientific community. So, let us continue, I will continue our discussion with the spatial cloud in our subsequent lecture and to see that what are the effects into this.

Thank you.