Cloud Computing

Prof. Soumya Kanti Ghosh
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 35
Use Case: Geospatial Cloud

Hello. So, we will continue our discussion on cloud computing and today we will talk about a typical use case of spatial cloud or more specifically geospatial cloud, right. So, what we mean by this geospatial data and information and type of things. So, what we see these days that this spatial information about earth surface and both with coordinated system, right with location based information are becoming extremely popular we are we all of us are more or less used to or using day in day out these location based services or navigation services for finding trajectories from one location to other and type of things, right.

So, all these are possible because there is a there are spatial data which are being maintained by various organizations, right as data is becoming more heavy with and multiple organization a storing this data. So, there is a need that how to integrate this data for decision makings purpose, right one navigation maybe one of the things there are several other decision making things specially different development program of government and government agencies or state agencies and type of things which need require some sort of a ground level information to be integrated, right.

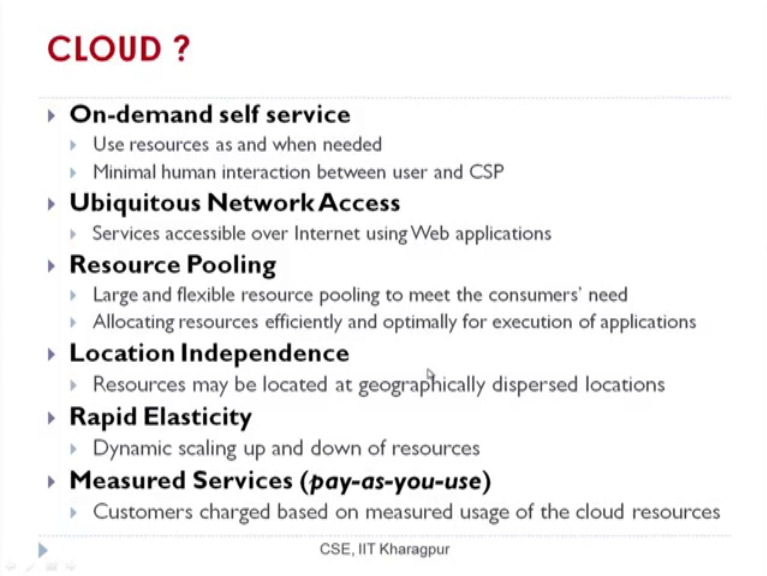
So, one way of looking at it that I store the data in a central place and then query on the look at the things, right, but if there are multiple stakeholders of the data or multiple providers of this data, then these making it is a central repository sometimes become a challenge, right.

First of all the data itself is cost intensive if there are expertise and there are manpower there are fund involving a collect in collection and maintenance of the data. So, once you put the data in some other place then the whole thing goes for a different type of management right. Secondly, that the volume of data is enormous. So, this sort of transferring of the data over things may be extremely difficult, right

So, what is what one of the solution what this spatial science or spatial management and things are looking at whether we can migrate to a cloud, right? It is some sort of a cloud sort of infrastructure which can be provided on the on this platform, right. So, this is a lot of work going on across the globe and we would like to see that what is the feasibility what sort of things are they in here and IIT, Kharagpur also we have a small group working of those; this sort of activity.

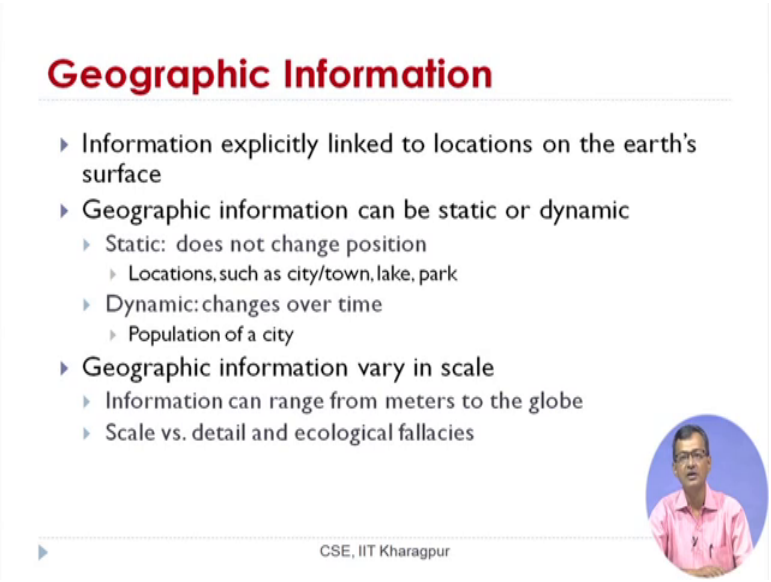
So, we will try to see that what is the feasibility of this sort of domains spatial domain going to the cloud, right.

(Refer Slide Time: 03:31)



So, if you look at. So, this is all we are discussing for long like what cloud keeps a on demand service ubiquitous access resource pooling location independence rapid elasticity measured services right and so, this is one part of the cloud side.

(Refer Slide Time: 03:47)



If you look at the geographic information side or the information spatial information or geospatial information side; so, information explicitly linked to the location on earth's surface, we are primarily looking at the spatial data which is looking specifically on the earth surface, right. So, it is a code dine driven thing. So, everything every data has a location incidentally if you see there was a report that more than ninety percent of our data are somewhere other related to some location, right.

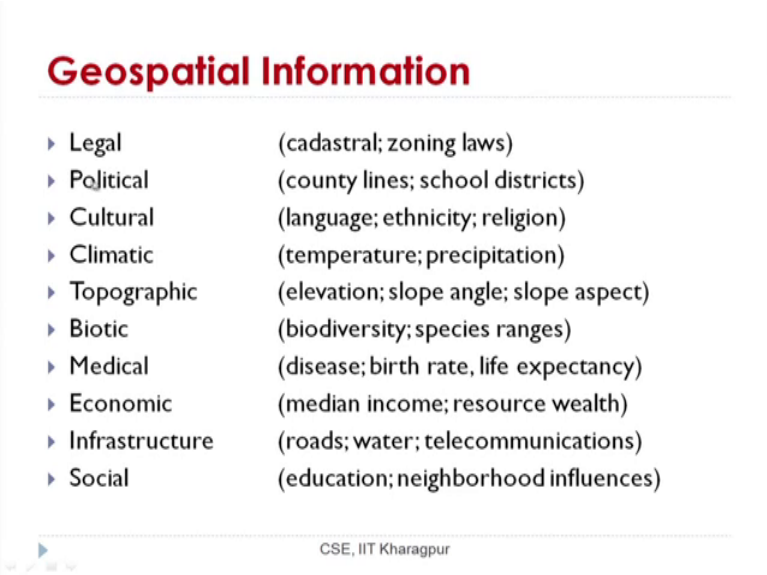
Suppose if I see our student data here. So, along with lot of other things it says that which hall he or she is staying which department he or she is studying right where he is or his or her hometown. So, all those things has a PoI a location on the earth surface. So, somewhere other with this spatial information is inherent or it is there in all our day to day life not only that in our most of our databases also right we may not be using it, but it is there if there is a pin code it says that which region it is referring to; right.

So, it is explicitly earth surface geographic information can be static or dynamic like static we does not change with time or changes over a long period of time like it may not be like a boundary of a state or a boundary of a district or a or sometimes in park lake etcetera they are not changing je day in the out are there are can be dynamic like some of the say traffic movement or population of a particular city which varies over time right there are lot of influx during data in people are going out during night time and type of things if I look at this particular building where we are sitting here at now and so, it has

the if you look at the building wise it is static right or we can say pseudo static it is not changing day in day out, but if you look at that food fall on the building that is changing right or even electricity need of the things may be changing.

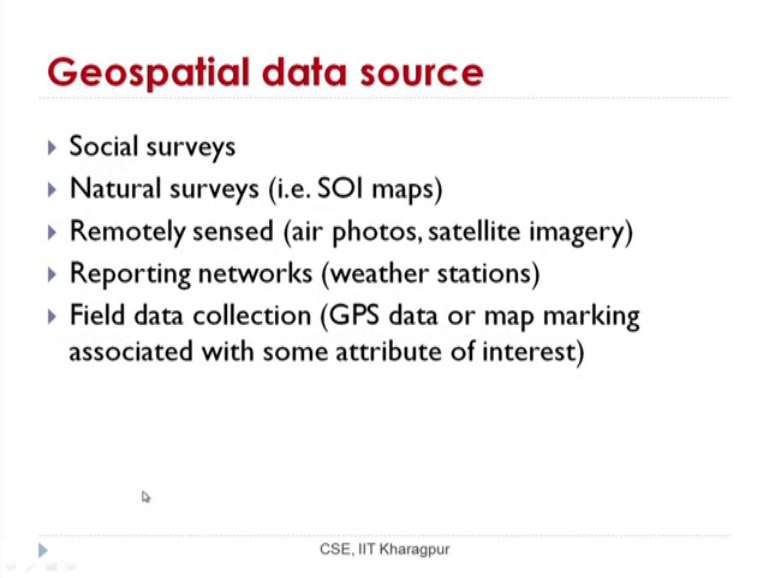
So, some of the data can be static some of the things can be dynamic and there is a typical thing this geographic data varies over in scale right. So, information can range from meters to globe right some of the data if I want to point, it needs say up to a meter scale right find out that what is the location of a lamppost right or, but it is can be higher the basically the map of a country or map of a district or map of a state may be much larger scale as; so, and scale versus detail and ecological fallacies and those things are there. So, with the scale things goes on changing right like if I look at a building from a higher from the sky it is a point, but if I close down it becomes a polygon and the with the scale the overall nature may change. So, these are some of the characteristics of the data, right.

(Refer Slide Time: 06:55)



And if you look at that common geographic data what we which we are accustomed with there can be some legal data that is location at these boundaries etcetera political maps cultural climatic some economic, etcetera.

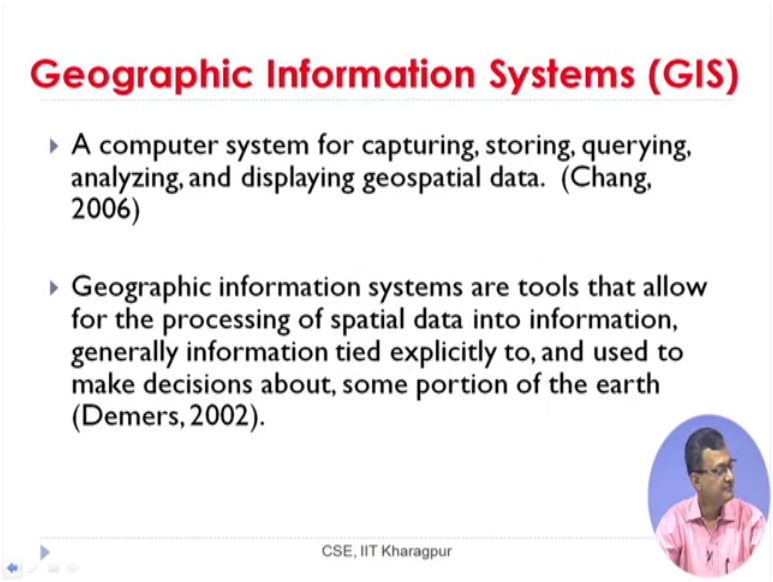
(Refer Slide Time: 07:14)



So, there can be n number of such structure and there can be different agencies who are collecting the data or who are maintaining the data that can be some sort of social survey there are national survey like we have survey of India maps, etcetera there can be some remotely sense sensor bits things or sensor data like it can I can have remotely sensed data like air photos photographs taken from the air or satellite images even reporting by weather stations collected by GPS and map marking associated with some attribute interest there can be pollution sensors across the things which gives data with locations where things are there; there can be air quality control pollution sensor temperature sensor and other type of atmospheric sensors switch gives data.

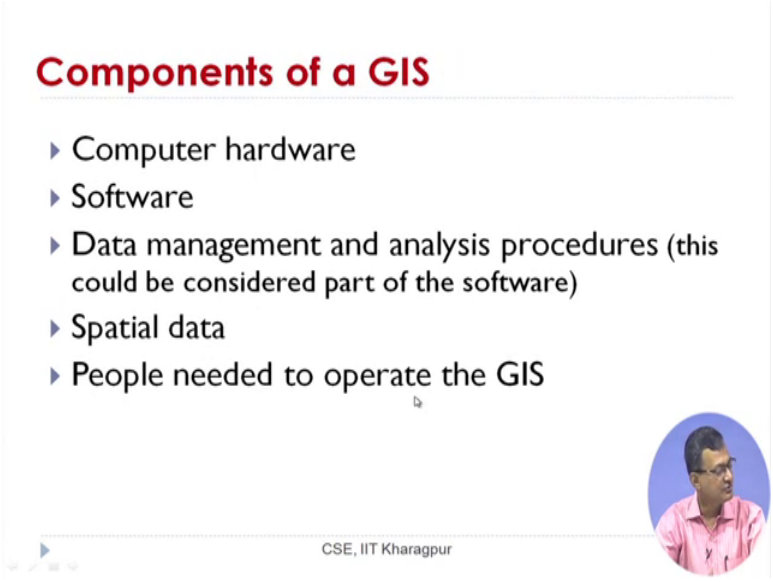
So, these are different sources of data incident with these data are maintained by different organization by their proprietary nature and when I want to query of one more than once such data that I may need to see our or interoperate between the data, right where this we need to see that where this cloud will be make some sense where using it.

(Refer Slide Time: 08:26)



So, if we look at some of the formal definitions like for capturing, it is a capturing storing querying analyzing and displaying geospatial data geographic information system or tools or platform which allows you to spatial data information related explicitly or implicitly for that surface.

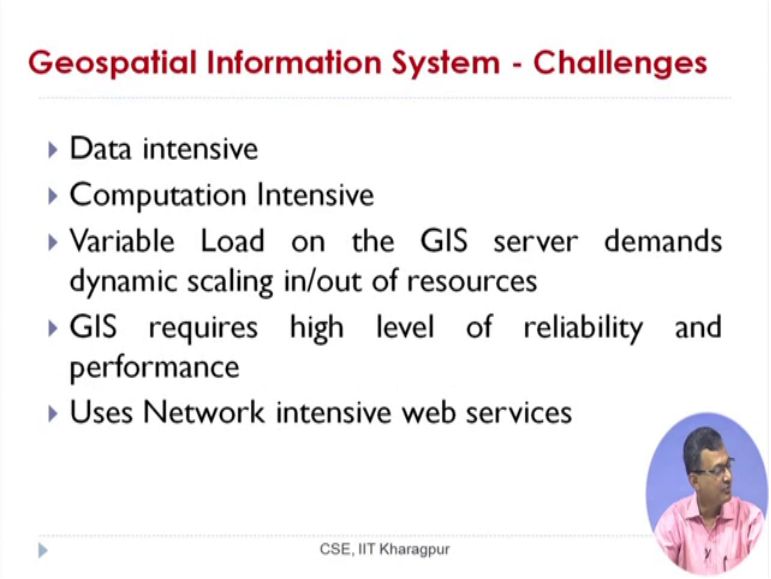
(Refer Slide Time: 08:48)



So, we are not going into deeper things just to show that why what sort of data just to have a idea of the and it can have hardware required hardware software data management is a major part spatial data which is has a different type of characteristics

both volume wise meaning wise and type of things and you require a special category of people who can operate on this data, right.

(Refer Slide Time: 09:16)

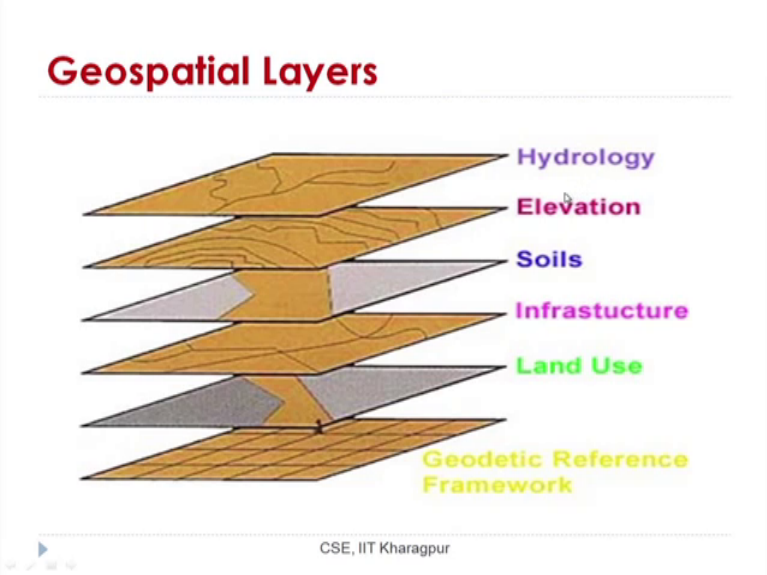


So, that is what we say people who can work on those type of things. So, these are the different component and what are the major challenges which are faced by this sort of huge spatial data it is data intensive; it is computation intensive, if I want to find a shortest path from a to b where the number of roads etcetera are like if I look at the representative as a road network as a graph. So, huge volume of the things of this graph then it is extremely computationally intensive not only that if I want to find out a some sort of a overlaying type of things then also it is computationally intensive. So, variable loads on the GIS server demands dynamic scaling in and scaling out of the resources and another interesting thing; it is not like that always you are doing what. So, data is there whenever the query comes you are doing and that it may also varies depending on the things.

So, we may require scaling in and scaling out of resources, right. So, that will be advantageous that when I have huge computation I scale bring in resources when I do not require; I release those resources type of things. So, there is a chance of having this sort of utility based computing or cloud sort of computing here and uses number of time uses network intensive web services as we have discussed web services are services which are based on this service oriented architecture and instead of data driven we have a service

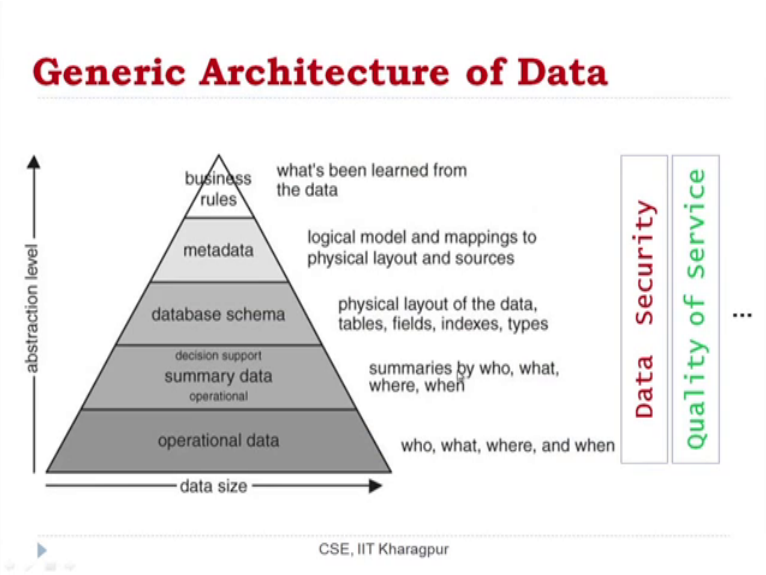
driven. So, we are talking about spatial web services which take care of spatial data type of things. So, it is extremely network intensive because the data at very place maintained by different organization and you need network intensive thing.

(Refer Slide Time: 11:02)



So, if we look at a typical batch of land. So, if you see the spatial data there can be hydrological data there can be elevation data soil data infrastructure same place and we require a referencing system, right, the means based or the well known referencing system what we understand from our school days is the lat long, it can be other referencing system which basically try to pin the data to a particular location.

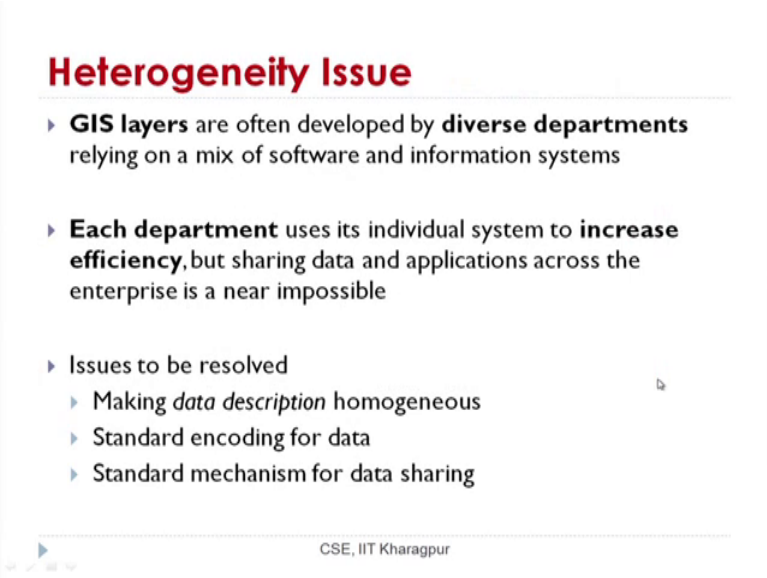
(Refer Slide Time: 11:28)



And if this is not only true for spatial data is true for any sort of data repository like we have operational data then summarizes making schema then metadata information of the schema which contains not only this schema information; lot of other things and there are business rules which are which the finally, the end user wants right that I want to know from location Kharagpur; IIT, Kharagpur to going to say IIT, Chennai what should be the root or IIT, Kharagpur. So, some place in Kolkata. So, I am not bothered about that how where data is stored how they are maintained I am I am more bothered about that what should be my path and what are the PoIs or the point of interest across the path and what is; how I should move and what may be the expected time to reach that things what are the congestion level, etcetera.

So, I am more interested in the service rather than the actual data itself right. So, these are required. So, that is my that is my business rule that I want to find the shortest path I may want to find a particular corridor of the things like if I make a say if one man makes a canal. So, what should be its catchment area how much it will resolve? So, it says that if there is a canal is. So, much flowing water is there then I want to see the means if the if the rule is that it says that both side 2 kilometer buffer will be the catchment area where the that water can be served. So, out of this 2 kilometer how many population has been served how much agree means land agricultural land has been served can be the way or means can be my calculation.

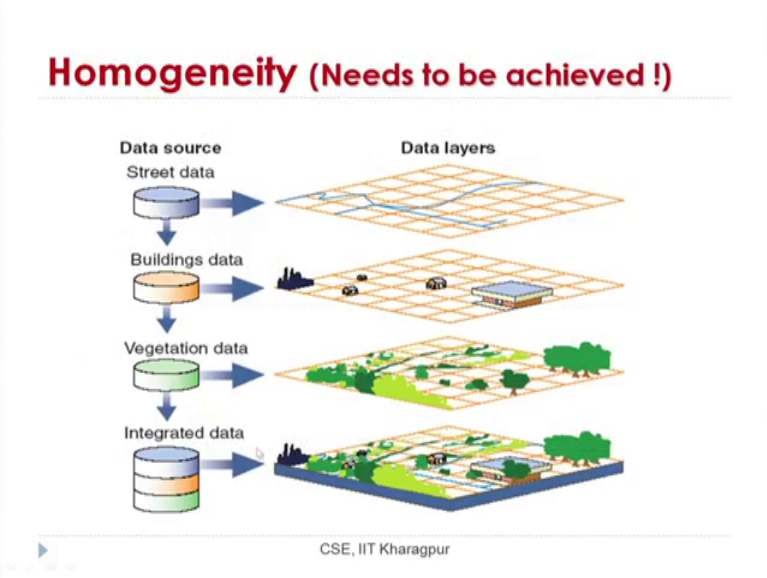
(Refer Slide Time: 13:32)



So, that is my business. So, the end user is more at the top of the thing and this is true for all sort of data and we have some of the verticals like data security which is a into end there are quality of service how reliable there can be management services and different other verticals which are which goes from the different aspects and what we try to see in case of specially in case of this sort of spatial data of geospatial data, we want to have some sort of a homogeneous way of looking at it like that different organization keeping that data when I query; I require them in the same type of format same type of means both syntactic and semantic wise some sort of a homogeneity otherwise I cannot query on those data, right.

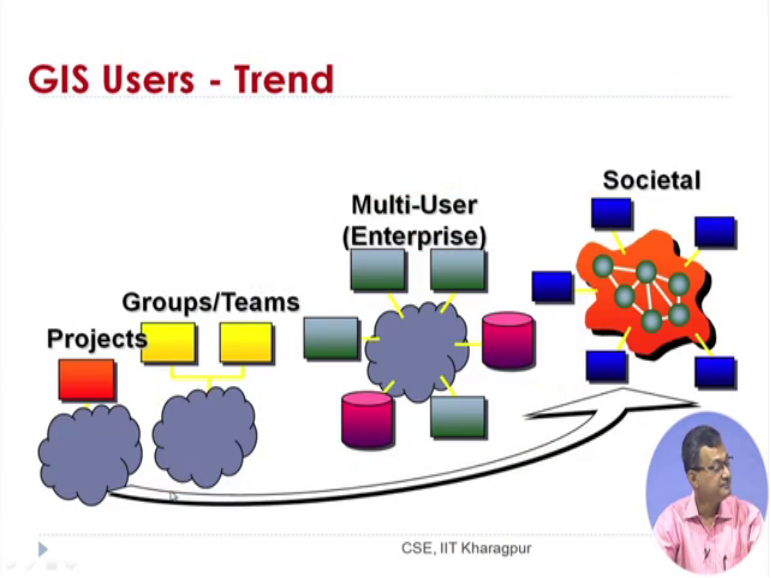
So, some references system is different, then I cannot quarry on the data if the scales are different then and. So, I cannot query data, right. So, that we need to resolve issues like data description that were homogeneous I should understand standard encoding of the data standard mechanism of the data sharing. So, all this what it says that some sort of a computing centralized computing facility which are like cloud type of things which can be helpful for the things, right.

(Refer Slide Time: 14:30)



So, though I can have these separate type of data like; it may be some data of the street building vegetation, but some sort of a homogeneous integrated data I am looking for because I need to query on this data.

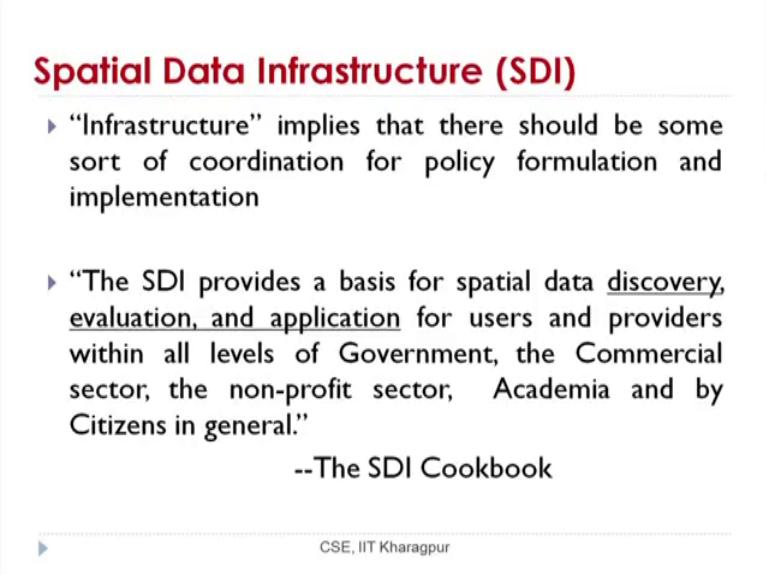
(Refer Slide Time: 14:44)



And we have seen that the trend is now more of a starting from the project mode; now of a more societal mode; that means, data is available to everybody based on because it end of the date is taxpayers money.

So, based on the considering policy securities, etcetera, but it should be available somewhere ubiquitously.

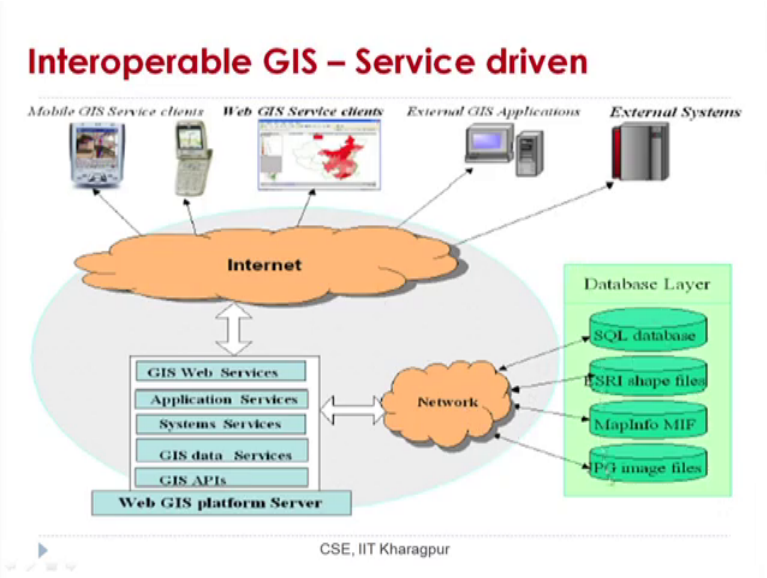
(Refer Slide Time: 15:10)



So, that may be a cloud may be a solution another concept came up or another very popular in advance country or also coming becoming popular or becoming a need in our country also; it is a making a spatial data infrastructure. So, looking the data as an infrastructure; so, whenever I want to query on the data I hook into this infrastructure right. So, it is a infrastructure. So, some sort of analogous I can we can look at that our telecom infrastructure, right, you have a there are different-different stakeholders and there have different type of say policies or plans etcetera, but I have a overall communication infrastructure somewhere other if I hook into the things like if I have a Sim card with particular plan then I can access voice data and multimedia type of things right based on again IP as I go.

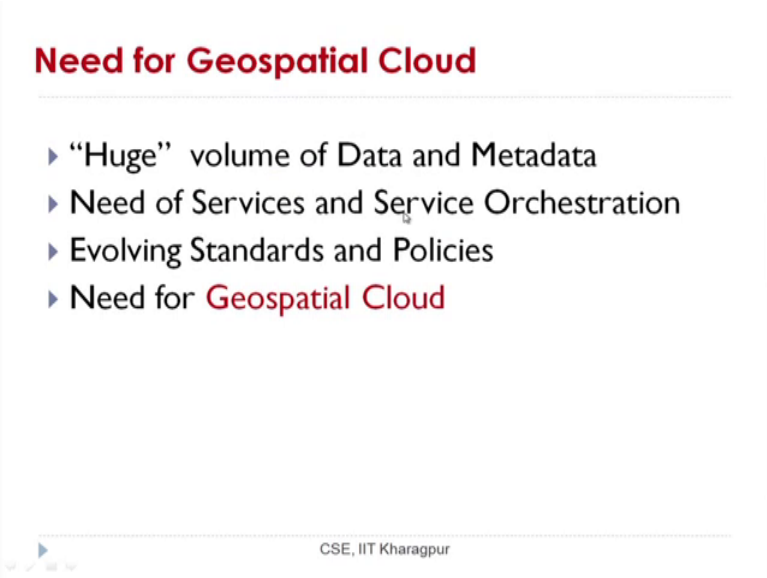
So, but as such that telecommunication is infrastructure, right. So, whether I can have a spatial data energy infrastructure if it is there, then I can do lot of things like I can do say planning for like what it; what comment does for say development planning thing for developing a particular area sitting on this school to some disasters management also right otherwise you need to always collect data put somewhere and do this.

(Refer Slide Time: 16:38)



So, this sort of infrastructure is the need not only it is coming; becoming true for our country also slowly. So, in other sense what we need there can be different way of accessing the data there can be different instances of repositories. So, what we harvest here is more of a web services or spatial web services which can take requests from here and talk to the back in databases and serve that it, right.

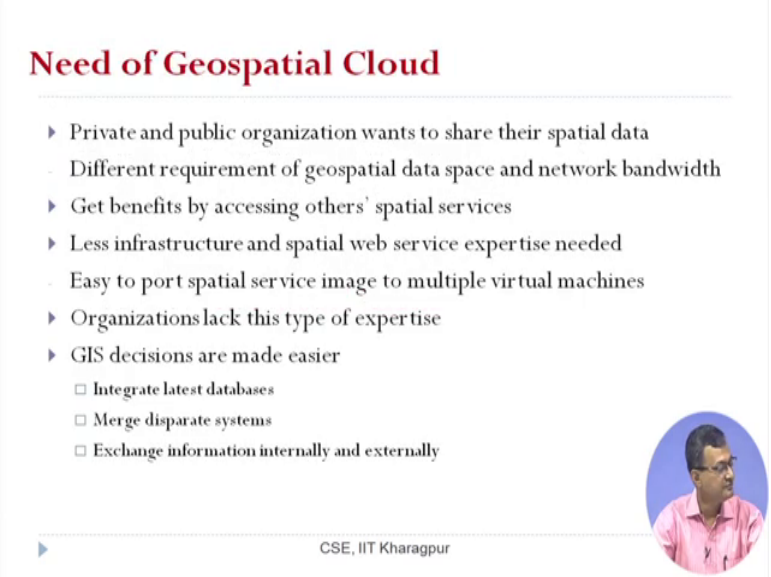
(Refer Slide Time: 17:05)



So, what we see that huge volume of data and metadata need of services and service orchestration because the if the if I my business rule may not be a just find a shortest

path I can have a things which has a series of services or the processing service to be initiated evolving standard and policies need for geo this may this basically lead to a need for a spatial cloud, right or geospatial cloud.

(Refer Slide Time: 17:32)

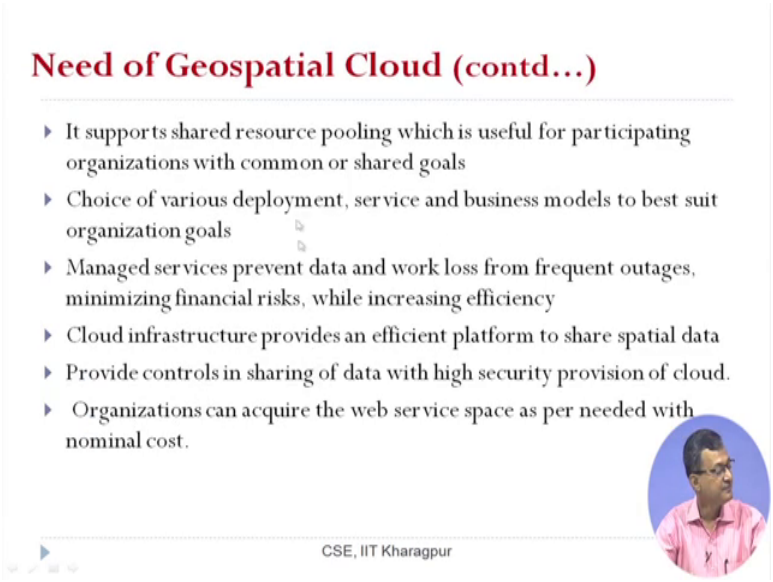


So, there are as already we have discussed this point must as to look at. So, private public organization wants to share their spatial data. So, different requirements for spatial data space network bandwidth, etcetera based on the data volume get benefits from other things sharing the things, less infrastructure and spatial service expertise needed, right, there is another important thing, right.

So, if I want to store; if I am a organization of collecting data and storing. So, I require a infrastructure to store there the data are volume us data are requires a different type of bandwidth etcetera and a I need a processing thing and not only that for that I require separate expertise on manpower do that. So, if I have and as we understand that every infrastructure goes off or in the sense has need to be renewed every 3 to 5 years. So, it is sustainability of these things become say extremely difficult. So, instead if I hire infrastructure for my purpose that can be a more beneficial thing, right.

So, sometimes our organization lacks who are more specialized in spatial data lacks in maintaining the infrastructure and GIS decisions are used ubiquitously in various government same government and private things and if we look at this spatial cloud geospatial cloud.

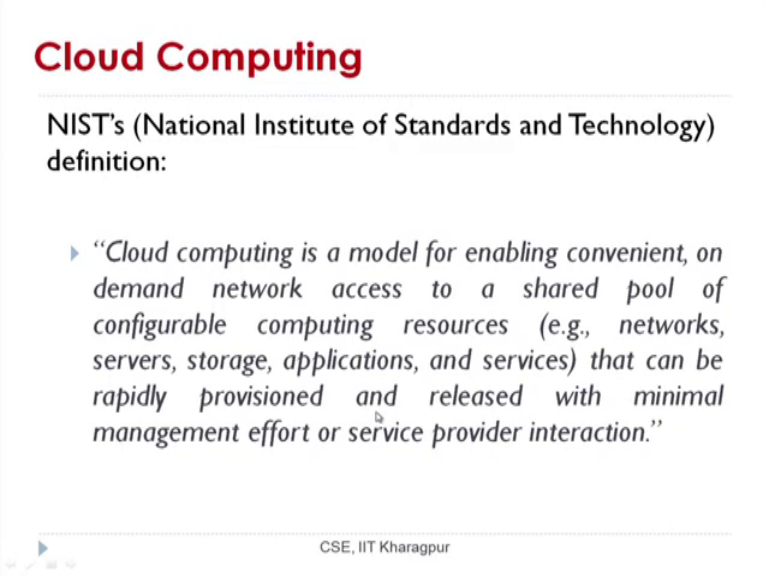
(Refer Slide Time: 18:58)



So, it support shared resource pooling which is useful for participating organization commons or shared goals as we are discussing a couple of slide back that it may be bumpy the requirement; that means, I need to scale up scale down on my resources. So, it is a say shared resource pooling will be there and choice of various deployment service business model to base suit organization requirement managed service prevent data and work loss from the frequent outage etcetera like.

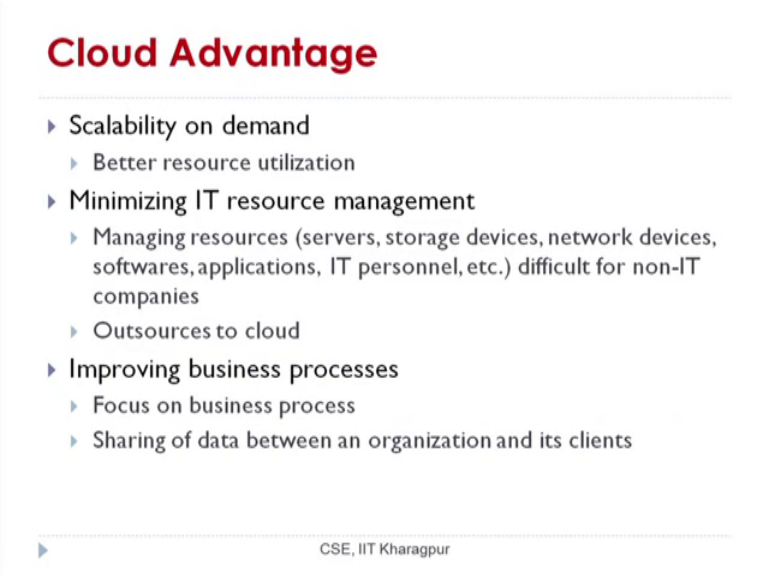
So, if it is having more; many service and which may minimize my financial constraint etcetera provide controls in sharing data with high security provision of the cloud as cloud also looks at the security features and looking aggressively on the security features. So, and security always come with a lot of cost. So, all we can leverage on the things and it can be hum much what we say economical to use those type of things and organization can acquire web service space as per their need. So, if I need only this sort of spatial services, I only harvest those services, right, I or if I need this type of data I only look for this data.

(Refer Slide Time: 20:18)



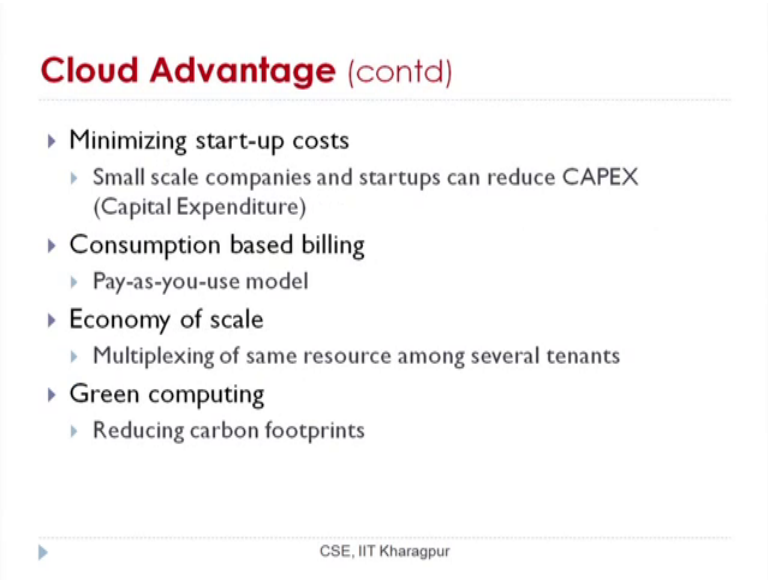
So, it is customized based on the need from the infrastructure right this is the NIST definition I just repeated.

(Refer Slide Time: 20:23)



So, already we know and cloud advantages also are known to us that scalability on demand minimizes IT resources improving business processes.

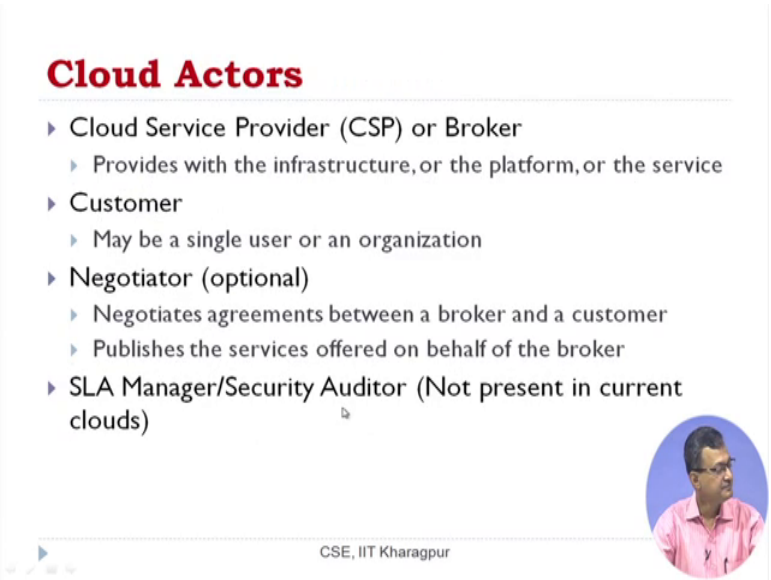
(Refer Slide Time: 20:33)



And all these advantages minimize startup cost consumption base building or pay as you go billing pay as you go model economy of scale green computing these are all is need for these spatial data infrastructure or spatial data management and type of thing. So, geospatial data infrastructure.

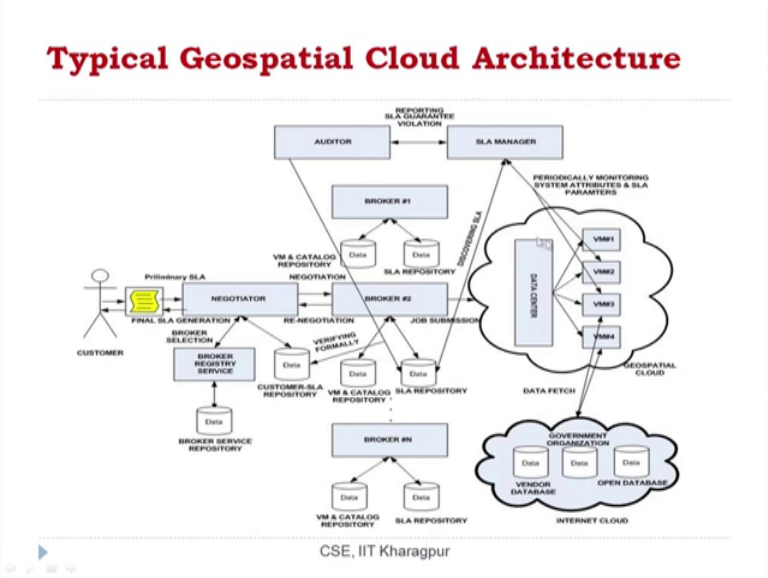
So; that means, it is it is somewhere it is suited for this sort of things, right.

(Refer Slide Time: 21:00)



And this also we have seen that there are who are the cloud actors like one we have cloud service CSP or broker customer is the other end negotiator is can be optional that negotiate between cloud and SLA manager of security auditor.

(Refer Slide Time: 21:18)



So, if you look considering all those things and if we look at typical cloud architecture for geospatial data this architecture may be true for any type of other services, but if we look at the geospatial data where the geospatial or spatial services are put into place. So, these overall things come into play. So, there is a once a customer or a user or a service consumer comes into play.

So, there is a negotiator which there can be number of brokers where which broke where which basically broker acting as agent for different data or different sets of data there can be the same data may be handled in 2 repositories may so happen and we have different data centers which gives VMs and there are different organizations which launch their data and services on this VM and these are being negotiated and connected with the end user they are other than these we require SLA manager to see that that this is SLAs or survey level agreements are not violated or things and there can be a auditor or auditor to see that whether this sort of policies in terms of SLAs, there can be security auditors security policies etcetera are being maintained in the this type of situation.

So, I require a brokering service or who broke for the things there are different data centers which hosts this spatial data of various organizations and there can be SLA there

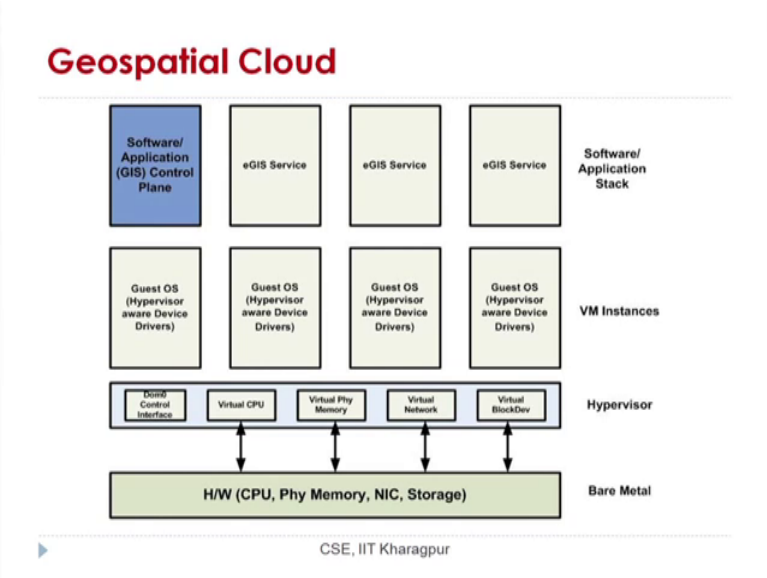
will be SLA manager and auditor which takes care of the overall management things, there can be other component of the thing other different components there can be separate auditing services, etcetera, but this is the overall broad infrastructure.

(Refer Slide Time: 23:19)

So, if I look at some sort of interface GIS or what we say some sort of a realization of this SDI which gives which gives spatial services what do you mean by this interface GIS is that it itself does not have any data per se, but it basically allows the consumer or the customer to connect with different service provider.

So, it is some sort of a cloud or some sort of a middle tire which connects those things, right. So, resource service resource allocation manipulation data services interface services all can be provided.

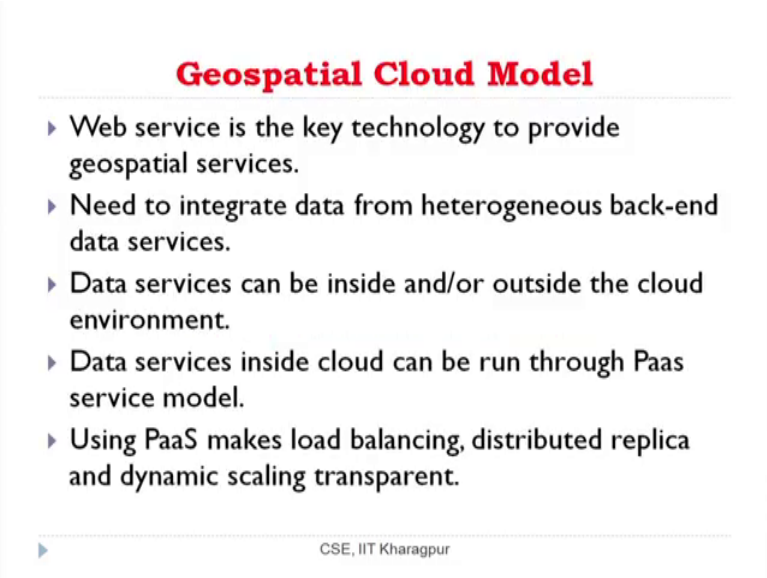
(Refer Slide Time: 24:02)



And we can see that a typical structure we have different instances of this there can be. So, this is my backbone hardware over there we have this hypervisor which emulate different virtual machines there can there in guest OS over that I can have different type of instances of this spatial cloud, right or instances of these spatial web services which are there. So, I can say that if it is say that spatial web services for a particular organization may particular state may be particular district can be instantiated on the things right and then I work on the thing like if you if we try to look at a some sort of analogy may not be very good analogy.

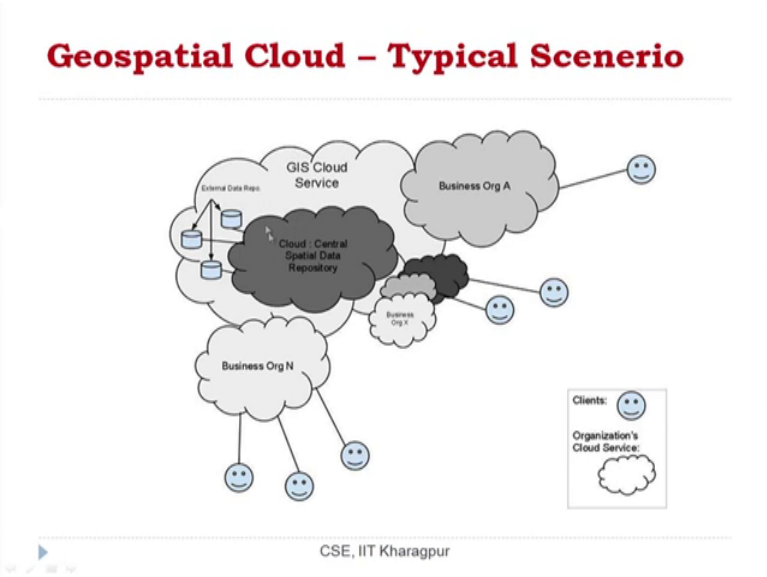
So, if I say that if I when I am using somewhat processing service on the cloud say Google docs. So, what we are doing we are basically instantiating my domain on that particular Google doc right somebody else is again instantiated things, but at the back end, it is the same it is the basically the Google infrastructure we are using where the hypervisor and it is basically software as a service. So, we are at a much higher level, but it is instantiating my thing. So, here whether I can have spatial service instantiated for my particular.

(Refer Slide Time: 25:28)



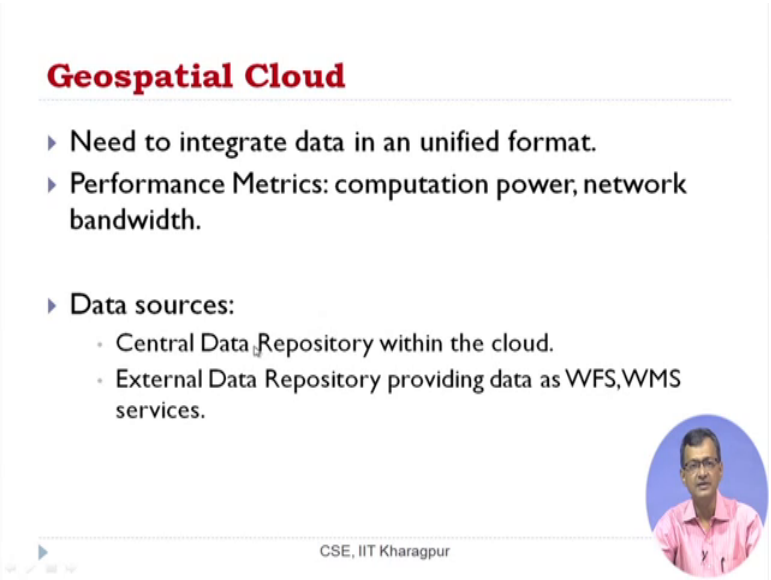
So, if we look at the geospatial cloud model. So, web services is here also key technology to provide spatial services need to integrate data and heterogeneous back end data services service can be set data service can be inside or outside the cloud there can be that some of the services can be outside can be run through PaaS service model using PaaS makes load balancing etcetera we can make scaling transparent.

(Refer Slide Time: 25:55)



So, what way of looking at it that a typical scenario where different organization are there and I have a one coordinating cloud which has the central repository which takes care that where the organization.

(Refer Slide Time: 26:10)



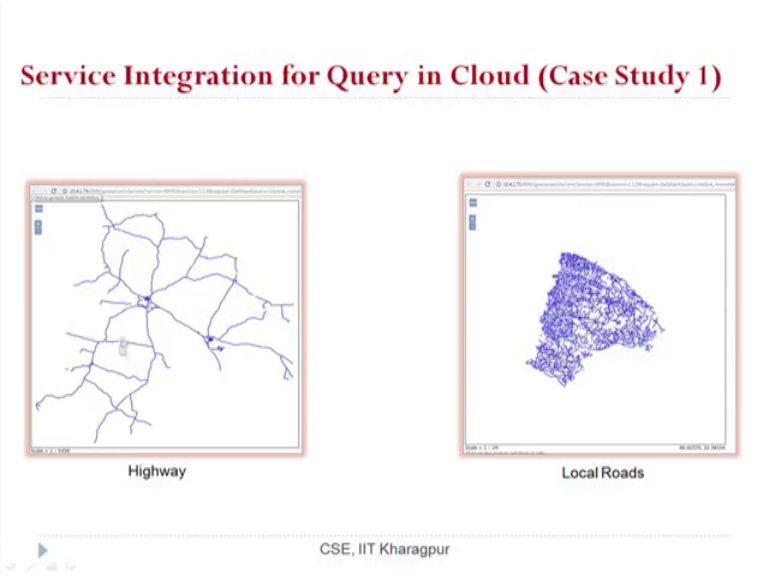
How they are working and type of things, right where the data is there. So, that we have already discussed need to integrate data in a unified format or homogeneity it needs to be instead or interoperated instead performance metric computation power network things etcetera need to be judged and there can be various data sources.

(Refer Slide Time: 26:29)

We just saw some snapshot that we made it experimental geospatial cloud I should say in our; at IIT, Kgp on our own experimental cloud that Meghamala where we have seen that which were and so on open stack open source cloud platform and it is a in house cloud used its a private cloud to IIT, Kharagpur and over that we have did some experiment.

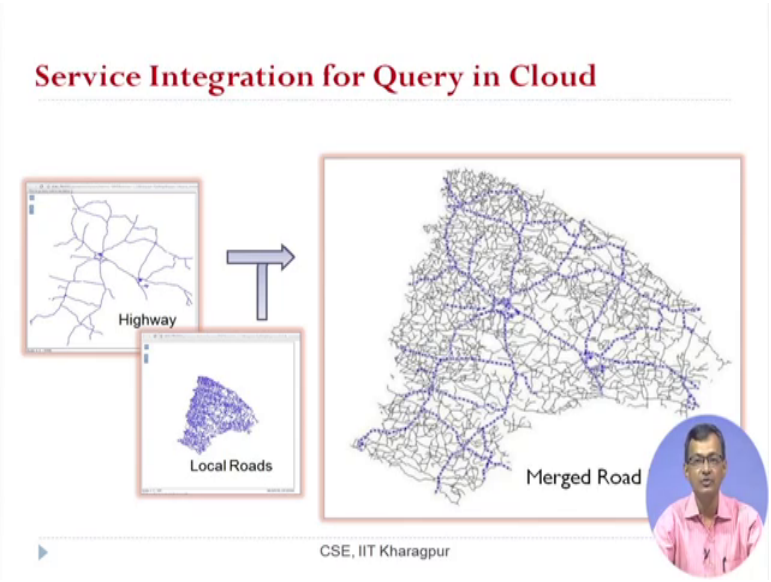
So, as such we are not showing that some performance, but showing that what sort of things may be there.

(Refer Slide Time: 27:01)



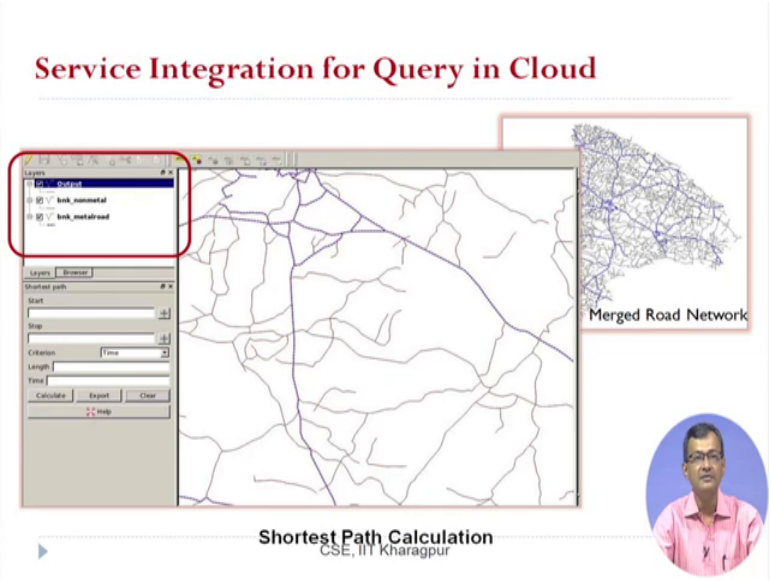
So, this is a particular area is a area of a Bankura districts which is a district in West Bengal. So, if I have a highway which is supplied by say highway national highway authority and on the other hand if I have a local road network which is a more handle by the state or the district authority then how to query something which involve this now if you the though the image may not be clear this are running into separate VM, it is something 1 dot 74 here, 1 dot 75 on the fly, I want to have a processing service which integrate, right.

(Refer Slide Time: 27:37)



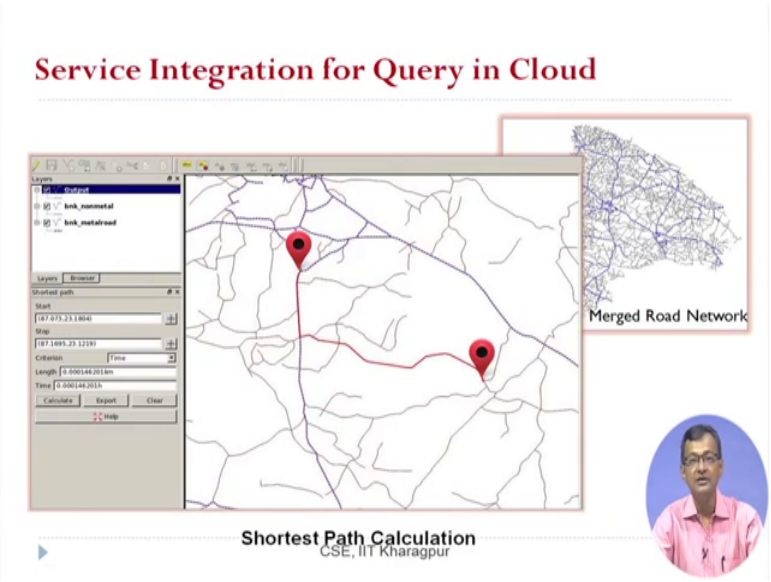
So, taking this 2, I need I basically have a merged road network right or in other sense I have a 2 services of the road, I am pulling and having a merged road or having the national highway what that and there we can basically.

(Refer Slide Time: 27:52)



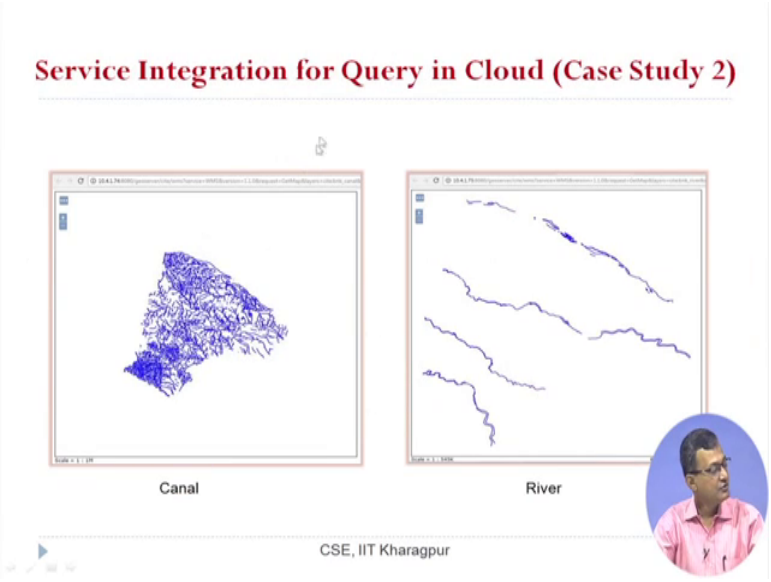
Now, query for a shortest path on these things, right which takes care both these national highway and the state highway like if it is my 2 location and then I want to find out the shortest path it does on the on this combine things.

(Refer Slide Time: 28:02)



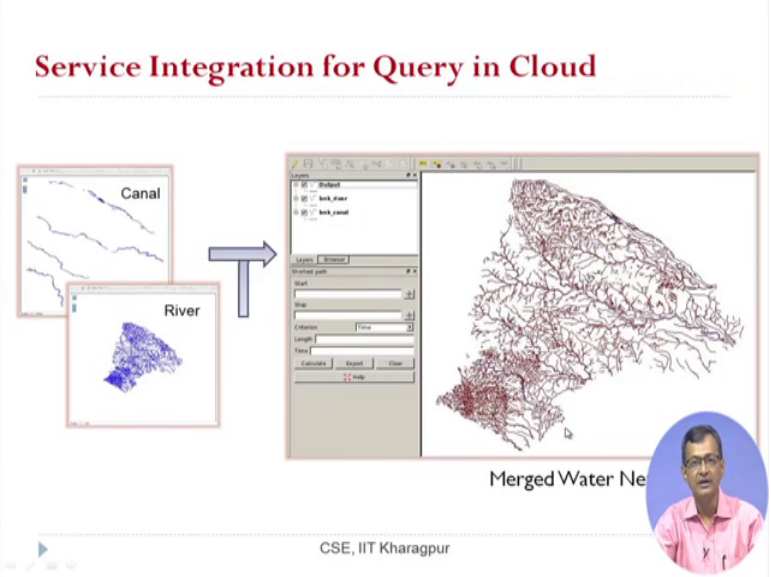
So, though the though the application may be trivial if not trivial it may begin well known application, but see 2 separate repositories we can pull on a service mode and which are running on 2 different VMs and pull on a service mode and basically my business rule is finding the shortest path, right, I am not neither interested in the whole data of West Bengal or whole data of India of road data, I am interested in that shortest path and which are the locations.

(Refer Slide Time: 28:44)



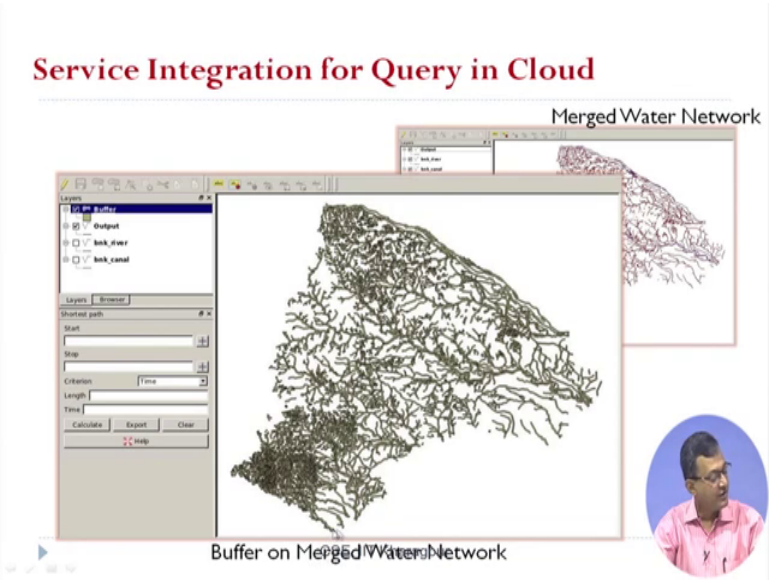
So, this is one sort of thing. So, other thing I can have again 2 sort of data where I have one a canal type of things on again same Bankura district and a river data which can be in a different thing.

(Refer Slide Time: 28:55)



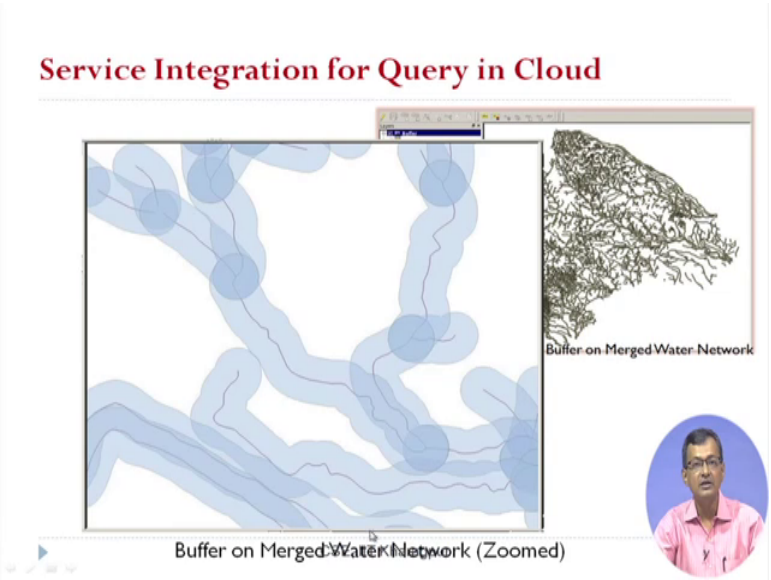
Now, I want to have a merged water network which integrate this canal along with the river again these 2 are in the separate VMs or separate cloud instances I have a merged data and I want to buffer on this merged data, right.

(Refer Slide Time: 29:09)



So, try to see that during heavy rain or flooding; what are the things.

(Refer Slide Time: 29:13)



Which if the business rule is that within 500 meter it is vulnerable, then; what is the vulnerable area of this particular buffer zone. So, again; I have 2 separate data instances which are integrated on the fly and gives this query, right. So, here we have experimented on a experimental geospatial cloud, we have installed over our Meghamala thing.

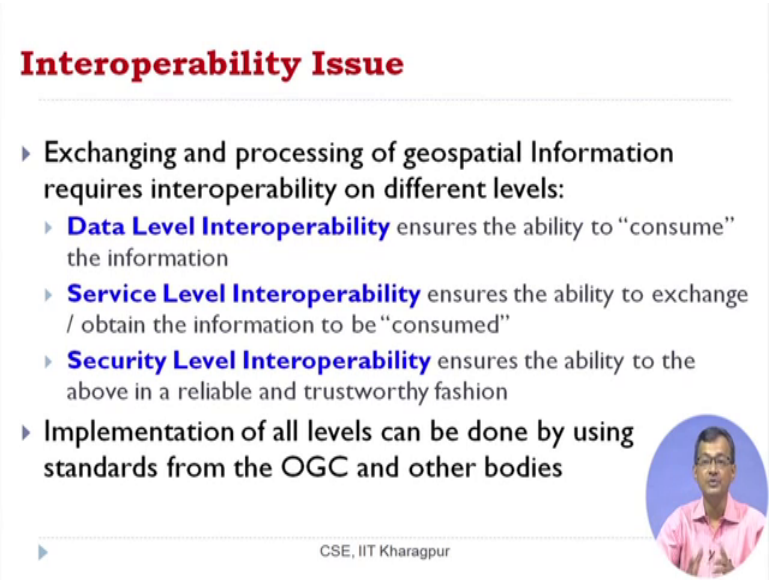
(Refer Slide Time: 29:39)

So, there are several challenges here other than which are typically something are typically for the spatial data which are not there in other type of what we say non spatial data type of things.

(Refer Slide Time: 29:55)

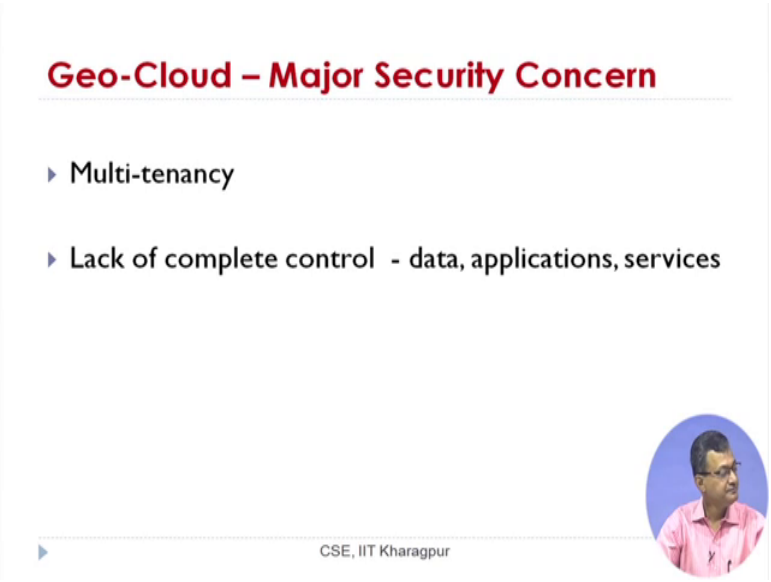
So, implementation of spatial database is a one of the challenge, it requires a lot of not only space need to understand the data semantics order scaling of the spatial databases right need for multi tenant thing policy management among these different tenants geographically situated backups if you want to keep data security some of the channels are also there for other type of things some are typically for this type of spatial.

(Refer Slide Time: 30:24)



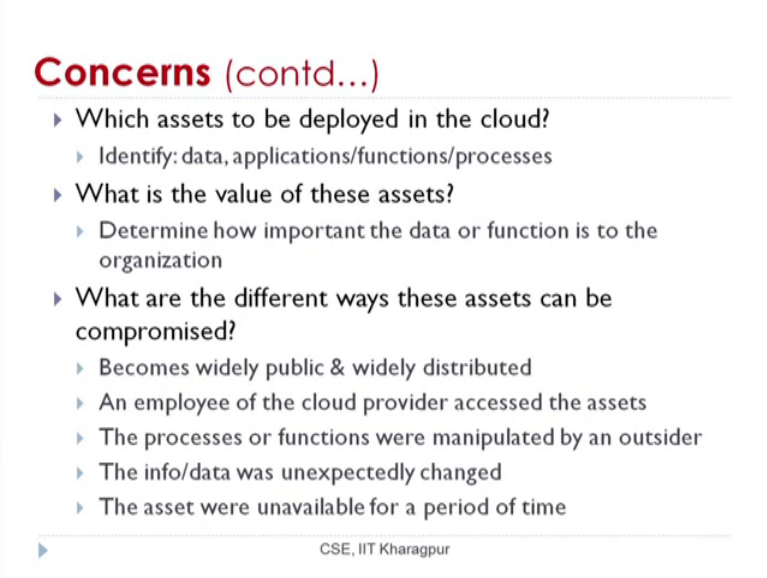
And also we have a major challenge of interoperability, right, data level interoperability like one data should match with other, while we are doing; say buffering type of things and things it can be a service level inerrability like 2 services can talk to each other security level in interoperability like; when I have different policies on different data sets then what should be the security level inter interoperability.

(Refer Slide Time: 30:49)



So, major security concern is the multi tenancy is already for other things and lack of complete control of data application services.

(Refer Slide Time: 30:59)



There are several other concerns which assets to be deployed on cloud what is the value of these assets what are the different ways this asset can be compromised etcetera. So, there are different types of other concerns which we need to look at in; if we look at the overall spectrum. So, this spatial or geospatial domain finds a immediate applicability or a huge applicability on this sort of a cloud thing. So, one of the good use cases which are which not only in our country in different organization in different countries how to make this heterogeneous data talk to each other for different type of citizen centric development planning and disastrous management is becoming a popular aspects of special cloud and this is a could use case.

Thank you.