Cloud Computing
Prof. Soumya Kanti Ghosh
Department of computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 11
Service Level Agreement (SLA)

Hello, welcome to this lecture on cloud computing. Today we will be discussing one of the important topic of cloud computing or one of the major aspect of making this cloud computing to realize this is the service level agreement, right. Whenever a consumer and a producer or a service provider and a service consumer want to exchange any services or what we say meter services which involves some costing, there is a there should be some agreement on the things, right. So, the agreement involve the pricing factor agreement involve the service availability factor the agreement may involve say for as other quality factors.

Now, this is very tricky issue and because a organization or even individual is switching for his own personal or own proprietary computing paradigm to some cloud taking the service from some service provider, thinking that you will get same type of reliability and other reliability same type of performance same type of security what he was having in his own premises, right. The system said it been in own own control.

Now, in order to that he need to have some agreement on the things. So, as such there is a still today till today they we do not have any very standard format of doing that. What you usually face whenever we are purchasing any services or any taking any VM or any cloud provider what we are usually do we basically sign off something right. We say that these are the terms and condition we agree. So, some sort of agreement will be there. So, we want to see that what are the different parameters to be taken into consideration and how this service level agreement typically look like. What are it is different components and at the end some of the popular cloud service provider. What sort of SLA or what sort of SLA parameters they are providing that we will we will see in this particular lecture today.

(Refer Slide Time: 02:42)

So, if we look at what is what is SLA or service level agreement a formal contract between the service provider and the service consumer. So, that is the thing SLA is the foundation sometimes known as the foundation or sometimes refer the foundation of the consumer trust in the provider, right. Whenever I purchase some service the first thing come as how much I can trust to the provider. So, this may gives me a formal way of trusting the thing that if the SLA is there. Not only that this may help me this parameters to compare between one provider with the other provider whom to I will take the service.

So, purpose is to define a formal basis for performing and availability of service provider guarantees to the to delivery, I need to a guarantee to the delivery. If I require a uptime of say 95 percent. So, there should be guarantee of delivery of 95 percent, right. So, I want those service provider which guarantees that I am I am more than 95 percent uptime it provides, right.

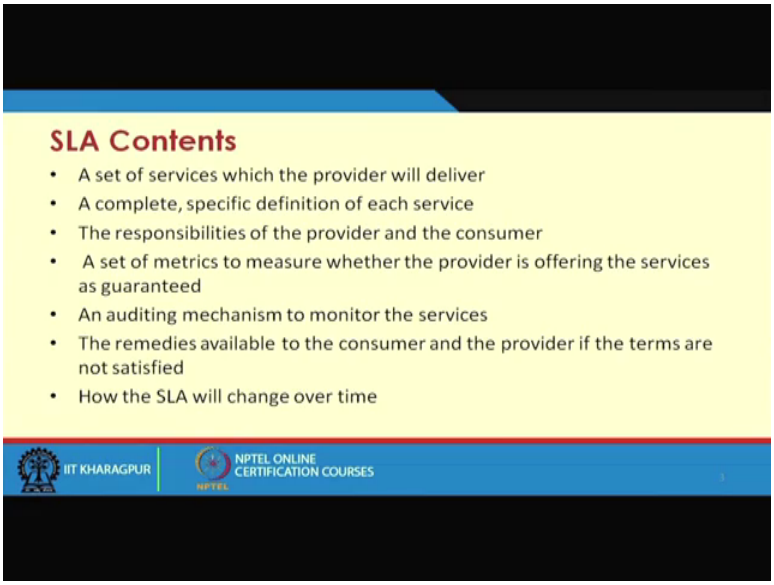
So, SLA now if you see SLA is a broad term right. So, it has to have different objectives, right. So, that what we say service level objectives. So, what this SLOs does? Objectively measurable condition for services right. SLA is the agreement has different components and this these are objectively measureable, I say that performance should be 95 percent. So, it is it is cannot be. So, somewhere objectively I should measure the thing. So, this So, SLA constitute of number of objective basis and these later on we will

see these objectives can be calculate from different parameters right. Or system level parameters which I can calculate. So, these are these are objectives which are there.

Now, interestingly the objective may vary from the consumer to consumer. Like even a consumer if a if a service consumer like a academic institution like us, may have different type of objective or may have not totally different I say may have different type of focus on the objectives then a consumer like may be a financial organization. Or a software company so, they have a different type of strategy level objective. Like I can say that my uptime maybe I can have 95 percent; however, I may require that data persistence to be much higher, right.

Some other organization say that my uptime is typically more than 99 percent it cannot be compromised less than 99 percent; however, my I always take a backup of the data. So, I may not always require a backup facility and objective etcetera right. So, that can be other things I can have my objectives varies or varying over time I say during peak hours my availability should be more than 99 percent; whereas, during of peak hours my availability requirement of availability may be more than 90 percent, right. So, these things are somewhere objectively measurable thing. So, this components should come into play.

(Refer Slide Time: 05:59)



So, what are the different SLA typical contents we are looking for? So, it is a set of services which provider will delivery. So, it should specify the set of services the

provider will delivery. I would like to again reiterate. So, these are these SLA's are valid for different type of service provisioning, may I would like it can be IaaS type of services where the SLA he want to have you can have pas type of services you can have SaaS type of services or any type of other type of things like may be data services etcetera. So, this SLA's correspond to that type of with that respect right. So, SLA for particular which is a software as a service type of module and so and so forth.

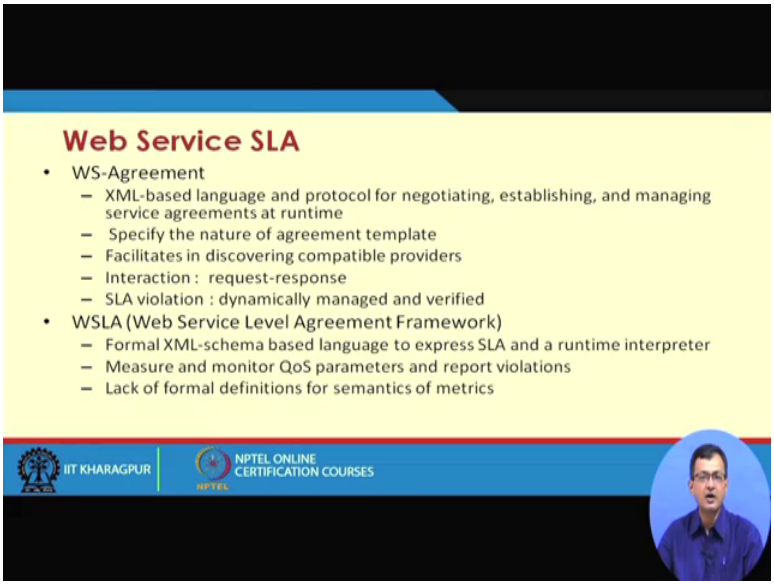
So, it should be complete specific definition of ease service right. So, it is been complete specific definition responsibilities of provider and the consumer to respect these SLA. A set of metric to measure whether the provider is offering the services as guaranteed like how do I know that whether the thing. So, some metrics should be there that which we shows that that it is providing the whatever it is guaranteed.

A auditing mechanisms right. Interesting a auditing mechanism to monitor the services. So, I need to have a some auditing mechanism to monitor the services right. So, it may be a third party auditing like I purchase as a service consumer purchased from service provider, but a third party auditor basically sees that where that there is a violation of the services etcetera right. So, these are the things which are to be to be monitored so, that is auditing is required. The remedy is available to the consumer and the provider these are terms are not satisfied that is another thing is there what should be that some sort of a payback mechanism or what should be the remedies if this is not. Like if I say if it is if I require a uptime of 95 percent and if the provider fails to give me 95 percent for sometime. So, whether he should pay me back or provide some other incentive to do that right. So, that should be some mechanisms where things are valid right. So, there should be.

And if you if we see that for a our day to day activity organizational activity there are this type of agreements mous etcetera, where there are things for defining that objectively there are remedies for violating those things what will happen. And so and so forth there are penalty what we say penalty of not providing the guaranteed service. If I if I say that I have give a guaranteed service and it is not provided then what I should pay give some penalty or I need to give some other incentive to do to compensate that that should be also specified, right.

What if there is not guaranteed etcetera and how whether the SLA will change over time right. It may change over time during the particular day of the things it may change over time over different days of the time and etcetera, etcetera. So, there is a whether the SLA will change over time that is another. So, whether it is a temporally changing phenomena, right.

(Refer Slide Time: 09:04)



Now, there is a concept of web services I believe most of you are having a thing service oriented architecture and web services, which are one of the what we say prime over for coming up the of this cloud services for benefit of all. I will take one short lecture on later on in subsequent togs one web services and service oriented architecture those who are not a custom, but I believe that most of you are used to it and type of things right. So, there is a in case of web services where service provider and consumer there is a concept of they communicate with each other.

There is also a agreement thing right. Or what we say web service SLA. So, web service SLA there are component like one is that web service agreement. The xml based language and protocol negotiating establishing and managing the service agreement at the run time. So, it is a de facto xml based things actually the whole web services things and so on basic foundation of XML specify, specify the nature of agreement template. So, what should be the agreed upon format of the template of the things. Facilitates in

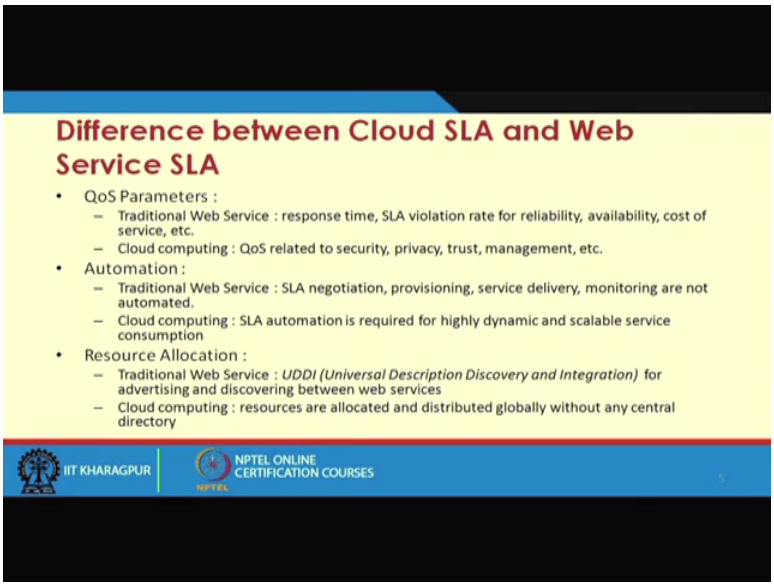
discovering compatible providers right. So, I there can be more than one provider. So, it should be able to see that where the compatible providers.

Interaction usually request response right. And SLA violation dynamically managed and verified. So, if there is a SLA violation this need to be dynamically managed and verified, and necessary action to be initiated of there is a violation of the SLA. There is a web service level agreement framework there is a concept called WA WSLA web service, level agreement framework. So, there is a framework for ws services formal xml schema based languages to express SLA and runtime interpreter.

Measure and monitor QoS parameter quality of service parameters and report violations if any lack of formal definitions for semantics of the metrics right. So, there are there is what we are talking about is more of a syntactic way of looking at there are few issues of semantics of looking at, but there are lays standardization of what the semantics of the whole thing means like whether the whole this parameter still something like I can say that whether this parameter tells that the system is performing well is going down and type of things on I am expecting a failure something or this type of things may happen when this type of things. So, there are different underlining semantics which is which is did to be formally defined or standardized.

So, if you look at this WSLA this as the cloud evolve from this web service and service oriented architecture keeping those the framework in mind. So obviously, this SLA's are etcetera are also having a relationship with them with the things, right.

(Refer Slide Time: 12:12)



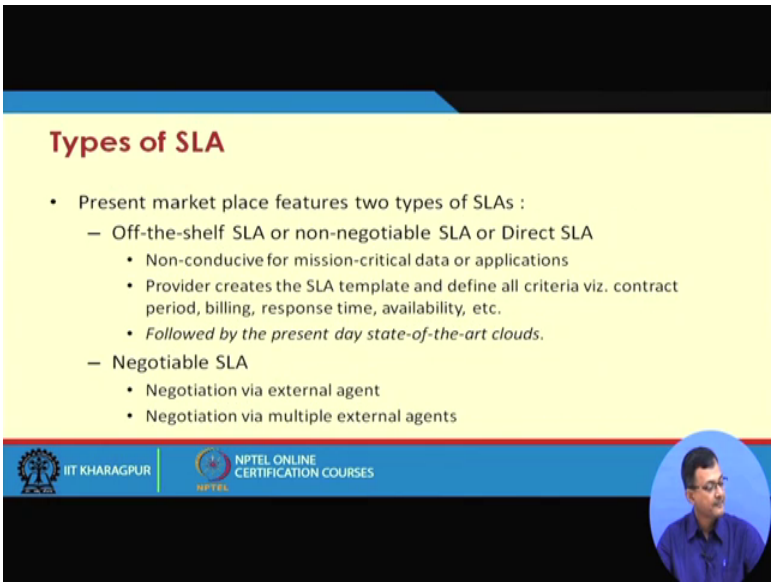
So, but if you if we look and try to see that little litigate of the what are the differences between the things mainly divide mainly trying to look at into 3 major components, one is QoS parameters that quality of service parameters, one is automation and another is resource allocation. So, these are the 3 thing what we are looking there.

So, in case of QoS parameter in case of a traditional web services. So, what we see response time SLA violation for SLA violation rate for reliability, availability, cost of services etcetera. So, in traditional things there is a response time plays a vital role and there are SLA violation rates for reliability, availability and cost of services etcetera. So, if there are violation then we with there are what should be that there is rate that what we things will be there.

In case of a cloud QoS is related to security privacy trans management etcetera, have more importance right. So, the basic way of handling may be there, but in case of a cloud we are more concerned about security more concerned about privacy trust overall management and so and so forth. If we look at the automation point of view traditional web services SLA negotiation provisioning service delivery monitoring are not automated right. In case of a cloud SLA automation is required for high dynamic and scalable services. Like I can scale up scale down and thing. So, I dynamic and scalable services may require this monitoring.

Resource allocation traditional web services UDDI, that is universal description discovery and integration that is a UDDI is one of the protocol which is very prominent in web services. It provides a (Refer Time: 14:07) services and we do use those type of things in cloud also for advertising and discovering between the web services. So, this UDDI is there in case of a cloud resources are allocated and distributed globally without any central directory perceive right.

(Refer Slide Time: 14:34)

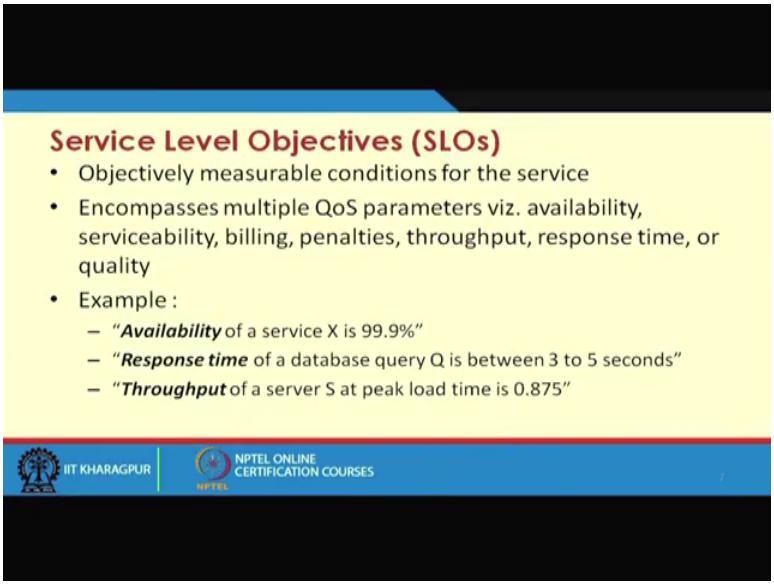


So, ideally it is a distributed things and there is a different mechanism of knowing that who is having what. So, if we look at the different types of SLA's right. What are the types of SLA's, one is the observe cell or non negotiable SLA or sometimes known as direct SLA right. So, one type of things of the cell. So, non conductive not conducive for mission critical data or applications. So, it may not be if n would be suitable if you have a if you have a very mission critical application you want to do this thing and so and so forth.

Provider creates a SLA template and define all criteria that is contract period billing responsible response time availability etcetera provides, say template and all whatever the things is there. It could followed by the present day state of art clouds. So, the this specially this public cloud follow this type of thing. So, whenever you want to buy some services from the public cloud it will it will go you this. So, with this form with other terms and condition. So, you need to agree on the things to work on the things right.

Whereas negotiable SLA; that means, you negotiate and find out that what things will be there negotiable via external agent. So, I may have there may be external agents which negotiates between the provider and consumer. And there can be negotiable via multiple external agent if there you are buying multiple services and then you amalgamate those services to achieve one particular thing then it can be multiple layers.

(Refer Slide Time: 16:11)

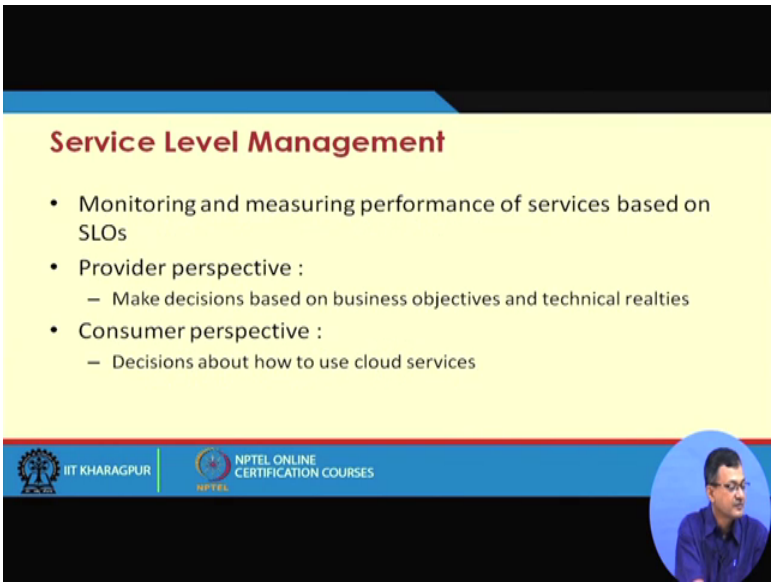


So, it can be. So, it can be either of the sell that is a standard or whatever is provided or negotiable. Usually what happened whenever we are having small requirement want to do a something which is there it in our in our own means some very small scale operations then we go for this type of the of self staff. Whenever we have a large requirement like he want to have whole organization process into the thing and etcetera then we want to look at the more have a negotiable SLA. So, want a special rate it is as good as you are booking a particular hotel or transport for one or 2 percent then you go for that whatever is available, but if you are buying the whole booking the whole hotel or the booking a transport bus or something then you have a negotiation to do that that what, we have also negotiate to have some fine tune the agreement poses.

So now what you are discussing about. So, we have SLA's right. Service level agreement and this SLOs are contributing to from this SLA's right. So, service level objectives. So, objectively measurable condition for service. So, that is one point encompasses multiple QoS parameters. So, SLOs can have a different QoS parameter. Like availability,

serviceability, billing, throughput, response time, quality etcetera. So, these are somewhere other need to be measured. So, for example, may be availability of service is 99.9 percent. So, that is a thing I can say that response time of a database query should be between 3 to 5 seconds right. So, that is a response time I am looking for throughput of a server particular server at peak time should be something 0.875, right. So, that may be another point of looking at the thing. So, these are this can be different type of SLOs.

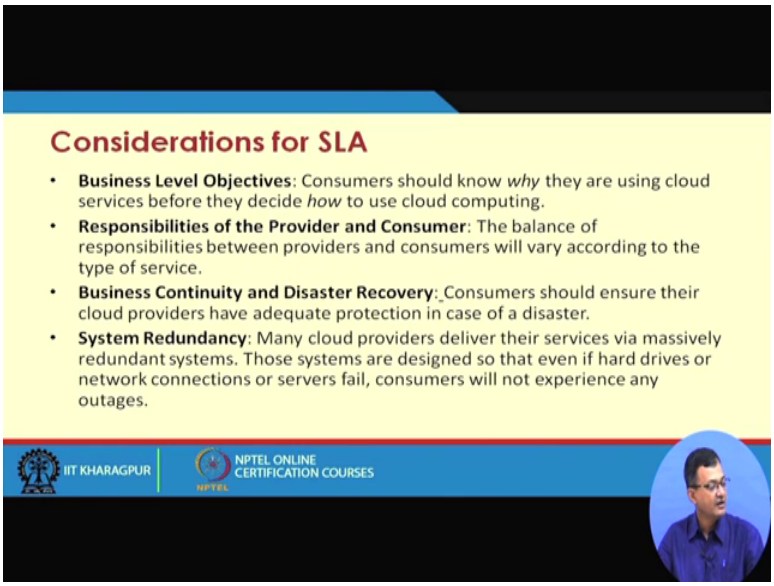
(Refer Slide Time: 18:15)



So, service level management. So, I have agreement I have SLOs now I have to manage this (Refer Time: 18:23). So, monitoring and managing the performance services based on the SLOs. So, SLOs are reporting that what are the different values etcetera. Provider perspective make decision based on the business objective and technical realities right. So, it has a business objective it is has a business goal and technical realities how much things he the provider is having at the back end that is also important. From the consumer perspective decision about how to use cloud services, right.

So, it is more that whether it is suitable for my organization where I suitable for my personal dream need, etcetera, then that way I measure.

(Refer Slide Time: 19:00)

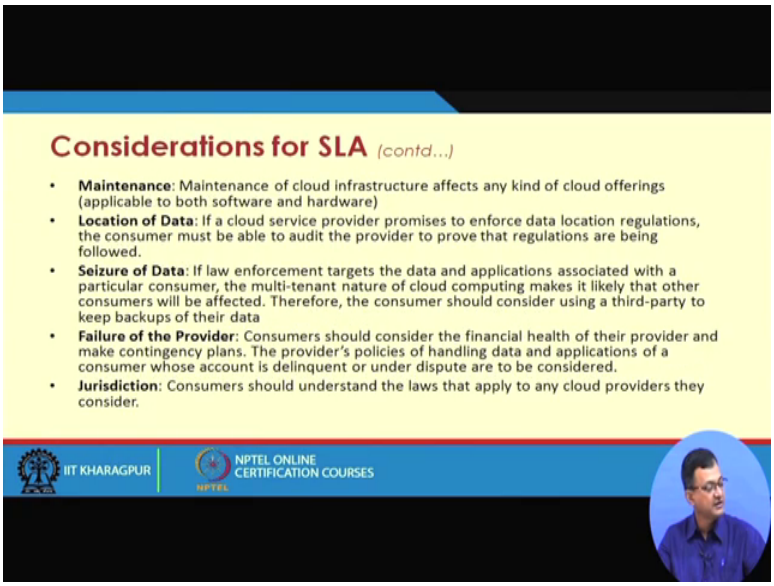


So, there are several consideration for SLA's few of them is one as we have our business level objectives. Responsibilities of the provider and consumer is important; that means, the balance of responsibility between the provider and consumer will vary according to the type of services, right.

So, there are some services where the provider's responsibility is much, much higher, whether there is a services which are which also require the consumer responsibility right. Like I put some data and get output of it I may have a things that I can the consumer the your provider can basically accept data as the rate of say x megabits per second. And then if I send some data as I x dashed where x x dash is more than the x then there may be problem of things problem of overflow of data and type of things then, which may intern violate other SLA's of the thing SLA of the provider.

So, there is a responsibility for the both the try to check and do and there are auditing to maintain that they all are for you. Business continuity and disaster recovery another important aspect for the consumer should ensure that cloud providers have adequate protection in case of a disaster right. It may be disaster natural manmade system failure etcetera. So, it should have been business there are system redundancy, so many cloud provider deliver their services via massively redundant systems right. So, that you can get guarantee.

(Refer Slide Time: 20:34)

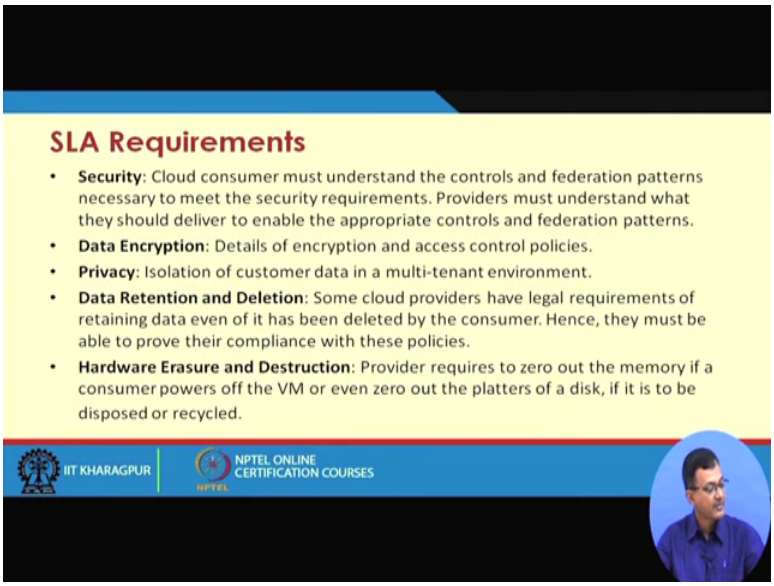


So, these are some of the things which are inter lead. There are other issues like maintains maintenance of the cloud infrastructure effects any kind of cloud offering applicable to both software and hardware. So, maintenance is a big (Refer Time: 20:43) factor location of the data if the cloud provider premises to enforce premises to enforce data location regulation the consumer must be able to audit the provider to prove the regulations are being followed.

Like and say that I can it may be So happen that IIT Kharagpur for example, say that all my data if in cloud should be residing within the jurisdiction of India right. I do not care how So, there should be a where to I verify that the data is not in some other country or some other land. Seizure of data, if low enforcement targets that data and application associated with the particular consumer the multi tenant nature of the cloud computing makes it likely that the other consumer will be effected also.

Like if suppose a consumer c one it is data the law enforcement or department want to seize then if it is residing multi tenant with some other consumer then the it may want to seize the whole at this then other things will be there. So, there is a issue which need to be there failure of the provider if in case of a failure of the provider.

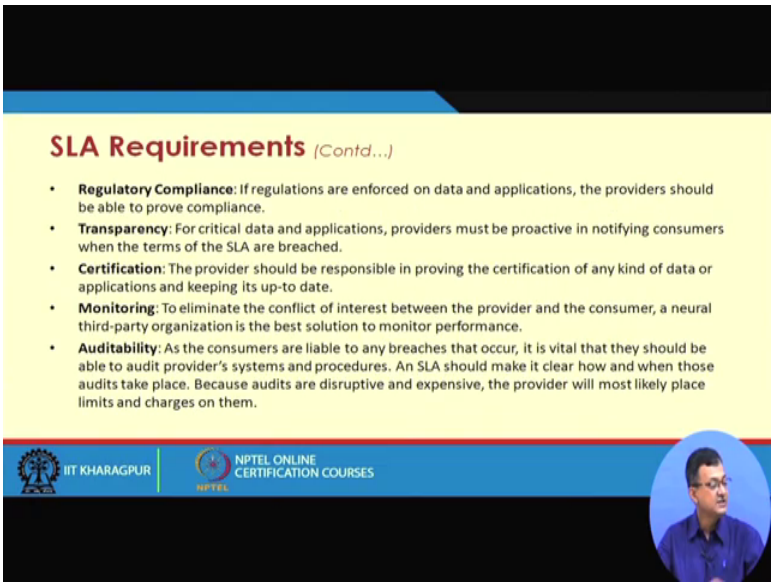
(Refer Slide Time: 21:56)



What should be there and jurisdiction that any type of litigation were where it will be addressed. There are other SLA requirements like one definitely security there is a big issue of data encryption if I encrypt the data how this key management will be there where the encrypted data will be residing where the key will be residing how the key will be communicate and etcetera.

Privacy issue isolation of the consumer data in a multi tenant environment how to isolate a consumer data and services in a multi tenant environment data retention and deletion right. Say if there is a if the how long the cloud will retain the data and where it will delete and type of things hardware erasure or destruction. So, provider requires to 0 out the memory if consumer powers of the vm or even 0 out the platters of the disk if it is to be disposed or recycled. So, if there is a thing as that the hardware erasure or disposing of the hardware etcetera then what will be the other effect on the things. So, this things need to be there. There are several other requirements.

(Refer Slide Time: 22:57)



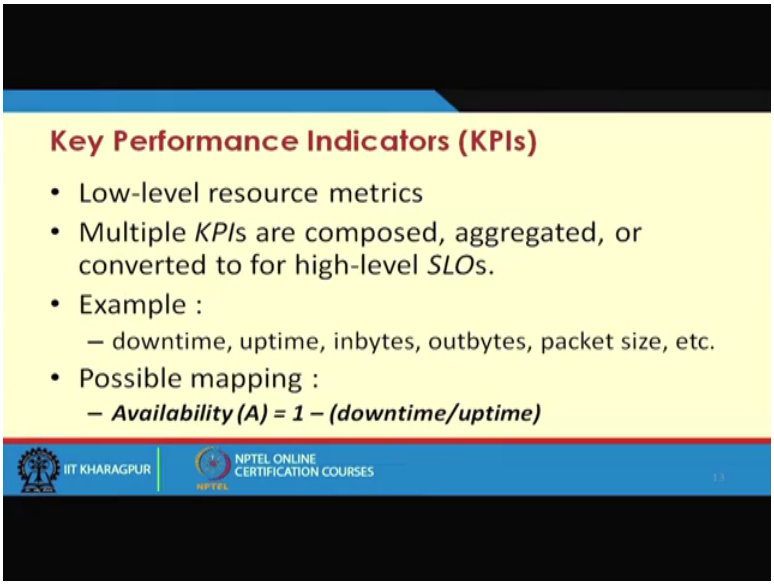
One may be one is regulatory compliance, transparency that what I am providing whether you are transparent to the consumer and the third party, certification whether provider has a certification process that it provide something which is which has a particular level of granted services and certification. Monitoring how to how to monitor the performance of the provider that is provider should be responsible for providing certification of any kind of sorry to eliminate the conflict of interest between the provider and consumer a neutral third party organization or what we say third party monitoring agency or third party auditor is the base solution for monitor performance.

So, these are easy to tell about, but while implementing it is very difficult, because suddenly in the provider things they may not allow this third party to work on their systems etcetera right. Then auditability as the consumer are liable to breach liable to any breaches that occur it is vital they should be able to audit provider systems and procedure right. As it will affect the consumers own business if with the providers with. And SLA should make it clear how and when those audit takes place. Because audits and disrupts audits are disruptive and expensive providers will most likely place limits on when charges on them right

So, if you want to do frequent auditing then it may basically it is a expensive both on not only on minute items it is on resource term also right. There may be downtime there may

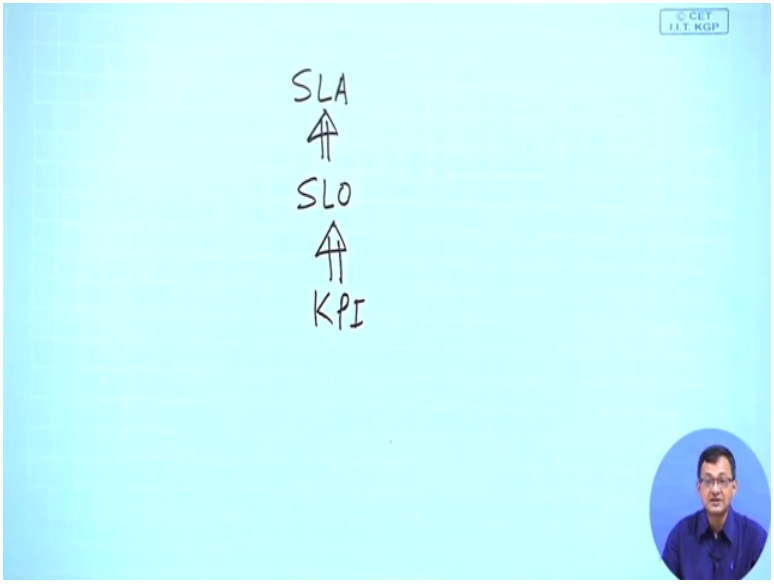
be other resource requirement, etcetera. So, provider may not be interested in that very frequently so, that to be need to be properly provision.

(Refer Slide Time: 24:51)



There is another factor call what we component called key performance indicator right. So, it is low level resource metric right. So, multiple KPI's are composed aggregated or converted to high level SLOs. Multiple SLOs are integrated to have that SLA. So, service if we look at I have SLA, then SLO and then KPI, right.

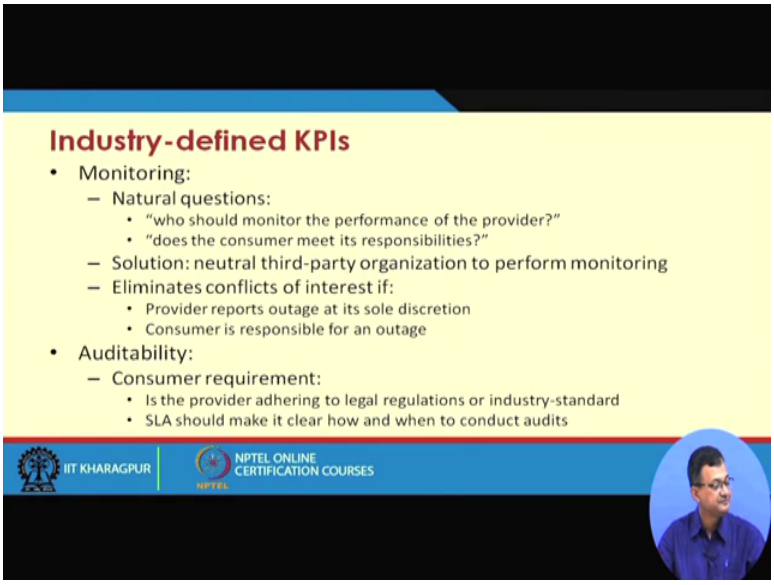
(Refer Slide Time: 25:19)



So, a KPIs is the lowest level metric which we have. So that means, here is a multiple KPI composed aggregated and converted to high level SLOs, like KPIs like typically like may be downtime right. Uptime in bytes out bytes packet size etcetera right.

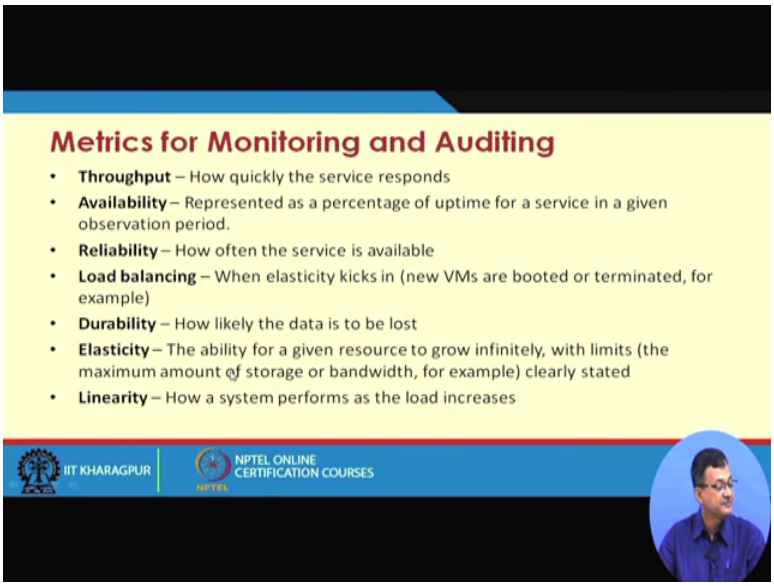
So, these are the different KPI there can be there. So, possible mapping like if availability is a one of the objective. Then availability is 1 minus downtime by uptime right. So, these are the components this KPIs which are very low level and directly measure measured from the system parameters. And SLOs are measured based on this KPIs AMD this KPI this SLO aggregation of SLOs are these SLA.

(Refer Slide Time: 26:27)



So, industry defined KPIs. So, monitoring so, natural question who should monitor the performance of the provider, does the consumer meets the responsibility solution neutral third party organization to perform this monitoring, eliminate conflicts of interest then there is issues of auditability as we have discussed.

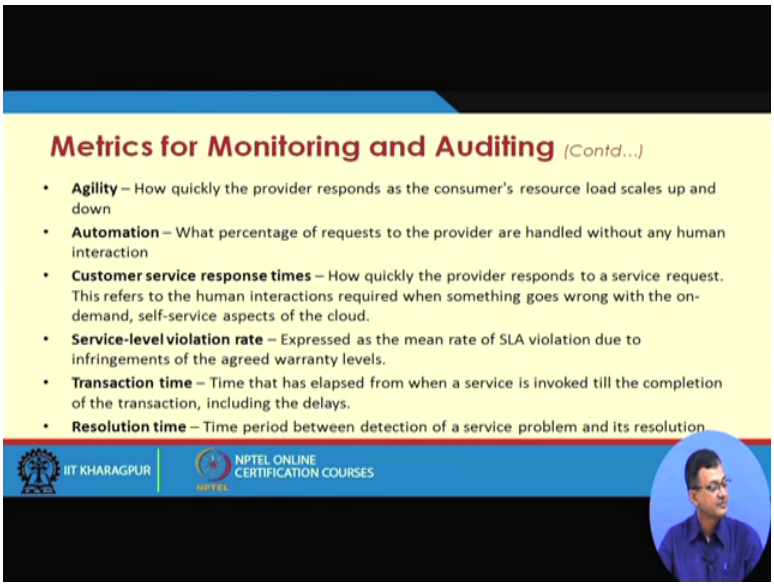
(Refer Slide Time: 26:48)



So, the metric for monitoring and auditing these are the typical widely used metrics. Throughput, availability, reliability, load balancing So, when elastics elasticity kicks in, So new vms are booted or terminated for example, then the load balancing of the systems is a important factor, right.

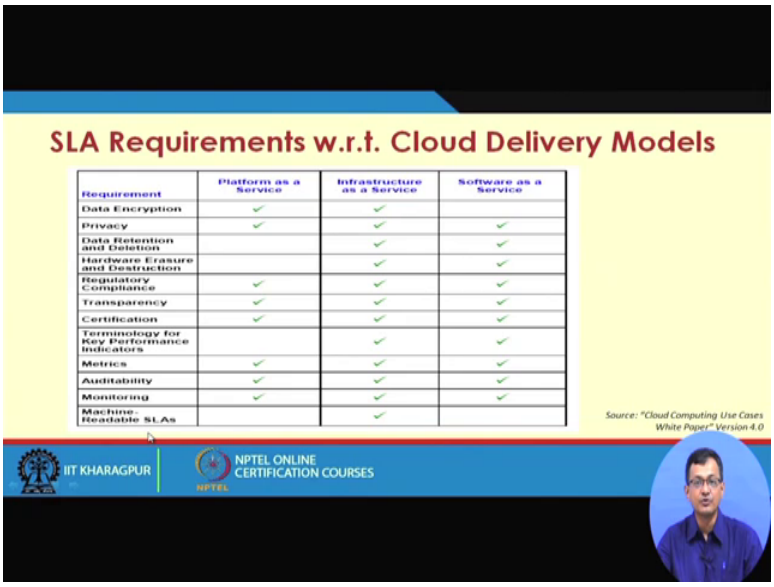
Durability how likely the data to be lost how much the, how much durable this data or services are there elasticity then linearity, how the system performs at the load increases if it is a linear increases like load increases and the system also increases the system provisioning also increases whether it is a linear craft if it is a linear craft, then it is easy to scale up like easy to chase this higher demand. If it is a non-linear specially exponential etcetera then it is a difficult process we will see later on.

(Refer Slide Time: 27:44)



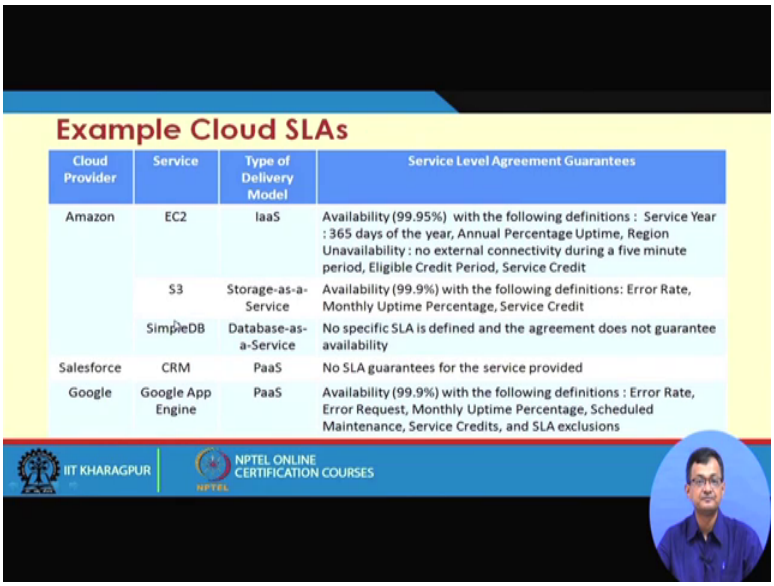
So, there are few more metrics like agility, automation, customer service, response time, service level violation transaction time resolution time these are different other components.

(Refer Slide Time: 27:58)



Now, if we have this chart there are different requirements at different levels right. Like IaaS may requiring something PaaS may requiring something and SaaS may be requiring, something on these different type of some a few of the requirement components.

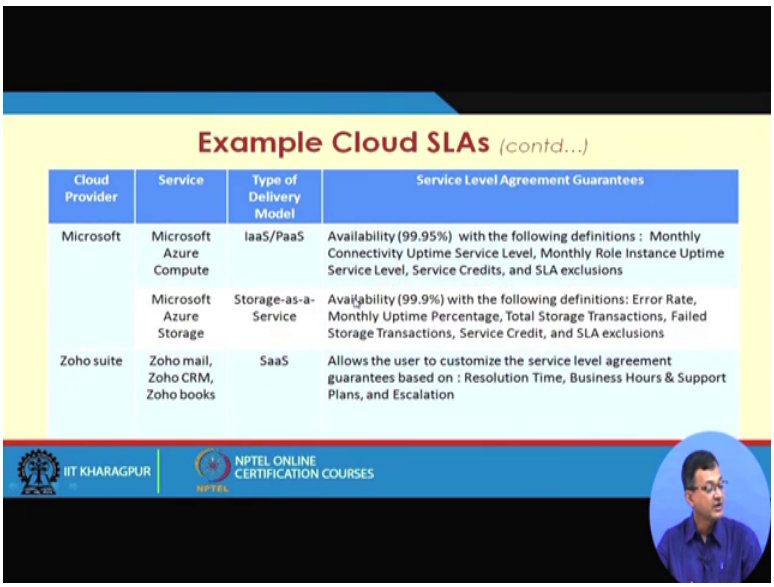
(Refer Slide Time: 28:16)



So, we will quickly look at some of these popular or example cloud providers SLA's or what they type like try to give the like Amazon ec 2 if the IaaS service provider it shows that the availabilities 99.5 percent. Like service year 365 days. Annual percentage uptime region and so and so forth.

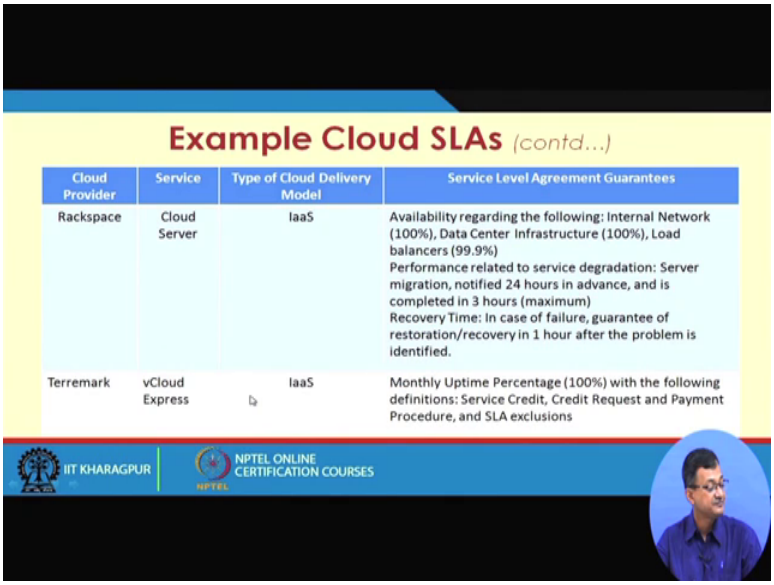
Whereas s 3 which provides Amazon is to provides storage as a service, availability 99 percent with the following definition error rate monthly uptime percentage etcetera. So, those components those parameter index we can find out.

(Refer Slide Time: 28:59)



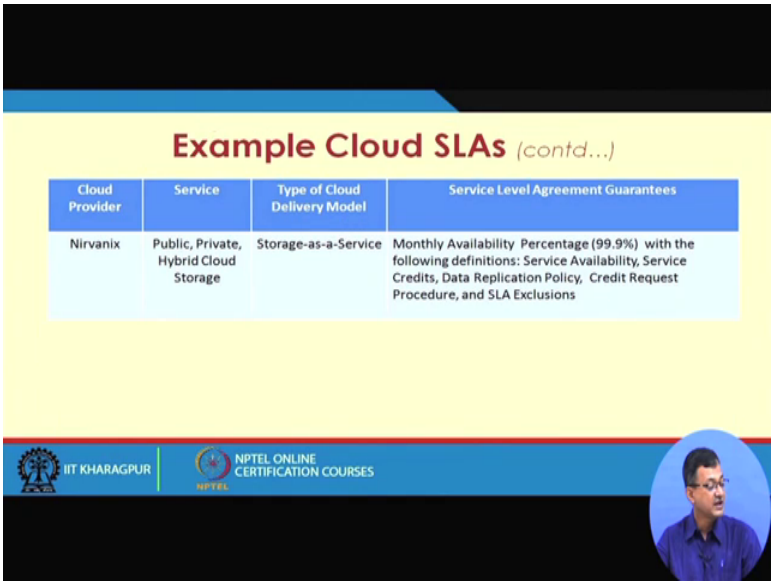
Similarly Google, sales force, Microsoft provides IaaS PaaS availability 99.9 percents. With following definition monthly connectivity uptime service level and etcetera.

(Refer Slide Time: 29:12)



Similarly, zoho which provides SaaS rack space terremark. So, there are some of the things which are listed here.

(Refer Slide Time: 29:16)



So, what we try to discuss in this particular talk is that this service level agreements in plays a vital role when we want to use want to make these cloud computing in our in practice right. I want to when I want to for my personal use, or organizational use, where

when we want to migrate from my conventional traditional system to this cloud computing thing. Then I need to look at around across these different SLA's right. What I need, what are the parameters I need and whether it is whether it is measurable by the at the provider ends and how I can guarantee that my work or my work process business process should not be affected adversely by the cloud service provider. So, with these we will stop here.

Thank you.