

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

**Lecture - 21**  
**SLA - Tutorial**

Hi. Today we will have a small tutorial on SLA it is in continuation with whatever we discussed on SLA. So, SLA as we understand for so service level agreement, and it is important for any cloud service provider and cloud service consumer to have an agreement to execute this either consume or provide this service. This as we understand that when we are using cloud computing we are basically leveraging on a third party services which are the service providers, and you the consumer things are hosted. So, it is both way to we should sign up or we should have a SLA.

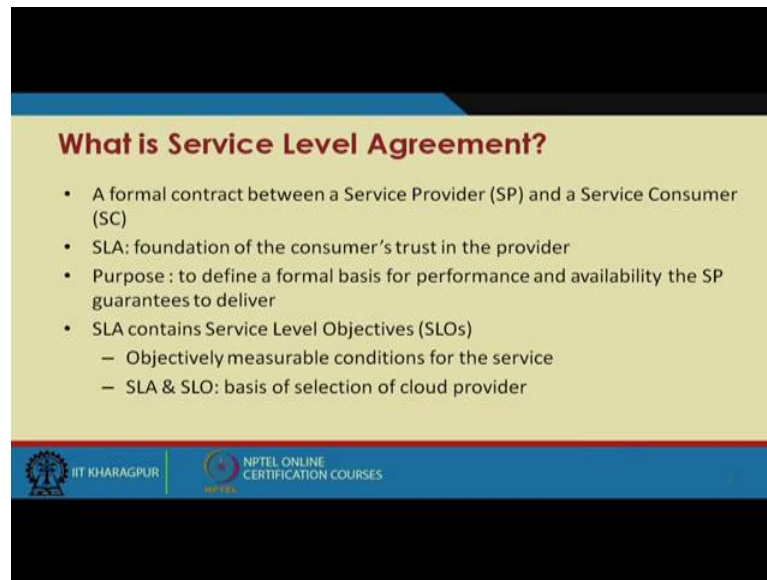
So, unfortunately there is as such there is no standard or there is a rather I should I should say that standard state evolving to have a standard SLA across the different services that, one of the major reason, maybe there are different type of services which are provided by different providers they are at different category of consumer we who require services at different scale.

So, never the less there are some broad guideline by which these SLA's are governed. So, whenever you take a service from any service provider let it be commercial any of the commercial cloud like say microsirisio or Google cloud, or IBM cloud or Amazon cloud or any cloud. So, you need to agree on some of the some agreement. So, what we have seen in our SLA talk that how this agreements are means how this agreements can be formed or what will be the basic underlining infrastructure for that.

So, for that my matter we have talked about SLO, KPI's and so on so forth, which allows us to build up this SLA. So, in some cases this is some of the metrics which will be there in some cases it is policy driven right. Like where your data through reside what should be the backup policy and so on and so forth, are more policy driven where as some of the agreements are more on parameter based, like what is the see up time or CPU uses or disc uses these are some of the things which are metrics.

So, what we will do to we will we will try to look at one or 2 problem, before that we will see that how different parameters are considered in different type of SLA's right. In different type of commercial life, we have taken this from again from internet resources primarily from commercial providers like Azure, IBM, Google and Amazon and others.

(Refer Slide Time: 03:37)



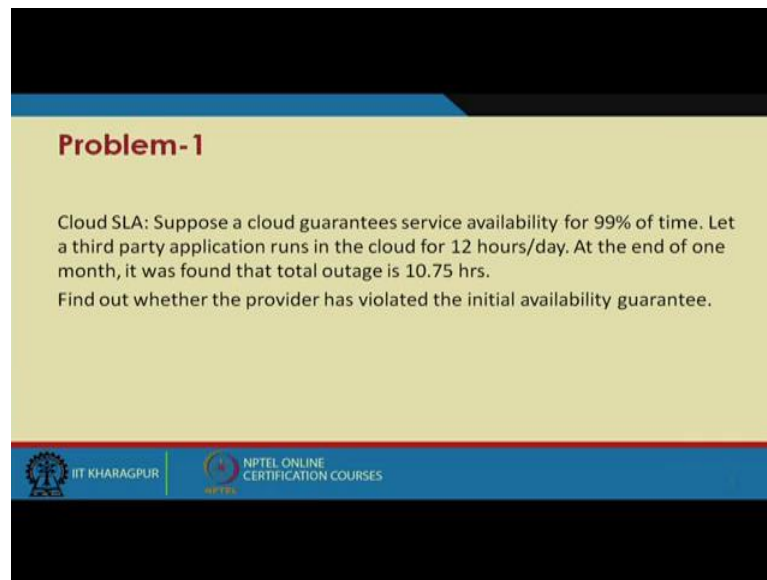
**What is Service Level Agreement?**

- A formal contract between a Service Provider (SP) and a Service Consumer (SC)
- SLA: foundation of the consumer's trust in the provider
- Purpose : to define a formal basis for performance and availability the SP guarantees to deliver
- SLA contains Service Level Objectives (SLOs)
  - Objectively measurable conditions for the service
  - SLA & SLO: basis of selection of cloud provider

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, it the idea is to say that what how they frame they are things in a in a way. So, before looking at those things those stuff we just see as we as we have discussed earlier this slide, like it is a formal agreement between contract between the provider and consumer. Foundation of consumer trust on provider, and sometimes visa-vise that how the providers wants to have this consumer on the things purpose to define a formal basis for performance and availability of service providing provider guarantee is to deliver. And as we have talked about SLO's like objectively miserable condition for services and SLA, SLO basic for the selection cloud provider this way seen just I kept the one slide. So that things will be there.

(Refer Slide Time: 04:21)



**Problem-1**

Cloud SLA: Suppose a cloud guarantees service availability for 99% of time. Let a third party application runs in the cloud for 12 hours/day. At the end of one month, it was found that total outage is 10.75 hrs.

Find out whether the provider has violated the initial availability guarantee.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, we will discuss this two problem and then try to look at some of the aspects of how somehow this commercials cloud another things and clouds look at, the how they define. So, like a let us have a simple problem like suppose a cloud guarantee service availability of 99 percent of up time right. Late third party application runs in the cloud for 12 hours a day at the end of one month it was found that there is a outage often 0.75 hours. Find out whether the provider has violated the initial availability guarantee right; so very straight forward.

So, it gives in the SLA as 99 percent of time, late third party run say cloud for 12 hours a day end of a month it is 10.75 and you want to find out whether the provider has violated the initial availability guarantee right. So, if we look at.

(Refer Slide Time: 05:33)

Total time for which the application to run (in a month)  
 $= 12 \times 30 = 360 \text{ hrs.}$

Outage time = 10.75 hrs.

Therefore, Service duration =  $(360 - 10.75) = 349.25 \text{ hrs.}$

% availability =  $(1 - 10.75/349.25) \times 100 = \underline{96.92\%}$

Initial Service guarantee = 99%

Final service availability < Initial Service guarantee  
 $\Rightarrow$  CSP has violated the SLA.

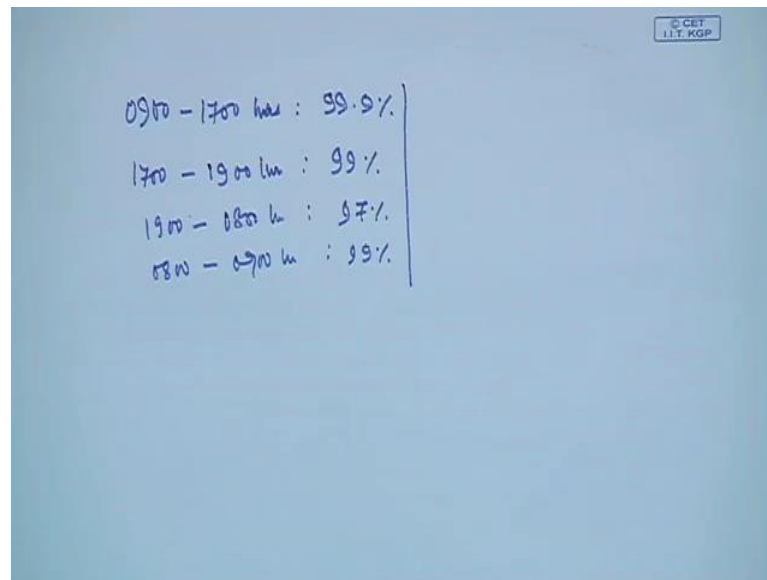
So, that is a problem. So, the total time for which the application to run in a month is equal to 12 cross 30, 360 hours right.

Now, what we say outage time is 10.75 hours, this has been declared this has been given. So, therefore, service duration equal to 360 minus 10.75 hours. So, percentage availability equal to 1 minus 10.75 by 349.25 into 100 which is 96.92, right. So, this much percentage availability will be there. This you can you can straight forward calculate. So, what it as it was their, initial service guarantee was 99 percent? So, has hence as final service guarantee. So, final service availability is less than initial service guarantee. So, what we can conclude that the service cloud service provider CSP if we say has violated the SLA ok.

So, it is a very straight forward simple arithmetic right, but what we see that if we can somehow measure this type of things, I can I can as per as the availability is concerned, we can basically calculate weather this SLA violations are there. If there is a set of SLA's which need to be looked into for everything every component, we can have this sort of simple calculation and or in some cases it may be little complex, when you want to do some statistical analysis to find out something. And then you can say that this is weather the SLA violation is there or SLA has been honored or not right. So, this way we can calculate this whether this any SLA is satisfied or not right ok.

So, this is this is pretty straight forward, but in reality it may not be that straight forward, what I can have different type of availability at different point of time, for that for that matter I can say that say if I considered a commercial say for example, a banking organization.

(Refer Slide Time: 09:59)



Time Interval	Availability Percentage
0900 - 1700 hrs	99.9%
1700 - 1900 hrs	99%
1900 - 0800 hrs	97%
0800 - 0900 hrs	99%

So, the a SLA's I can say that during my peak hours like I say 9 to 1700 hours, I require a availability of 99.9 percent. Whereas, in a off peak hours like 700 to say I can divide them into different scale, I can say that 700 to 1700 to this hours I can have 99 percent whereas, 1900 to next day 0800 hours I can still bring down to say 97 percent, and 0800 to 0900 hours I can say it is something again 99 percent. Now what I mean to say this availability requirement me also vary over time right, based on your business requirements right.

So, based on your requirement things will be like a institute like us I can I say, that if my lab if our labs are running between 2 to 5 or say morning in the morning say 2 to 6 and morning since an 8 to 12. So, during those lab hour I require a high percentage of availability. However, during the off peak hours or evening hours I may I may require much reuse thing. Because more you guarantee the services more you pay for it right. So, that is require so, there are there may be more complex calculation to look at, right.

(Refer Slide Time: 11:46)

**Problem-2**

Consider a scenario where a company X wants to use a cloud service from a provider P. The service level agreement (SLA) guarantees negotiated between the two parties prior to initiating business are as follows:

- Availability guarantee: 99.95% time over the service period
- Service period: 30 days
- Maximum service hours per day: 12 hours
- Cost: \$50 per day

Service credits are awarded to customers if availability guarantees are not satisfied. Monthly connectivity uptime service level are given as:

Monthly Uptime Percentage	Service Credit
<99.95%	10%
<99%	25%

However, in reality it was found that over the service period, the cloud service suffered five outages of durations: 5 hrs, 30 mins, 1 hr 30 mins, 15 mins, and 2 hrs 25 mins, each on different days, due to which normal service guarantees were violated. If SLA negotiations are honored, compute the effective cost payable towards buying the cloud service.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, a similarly a we can we look at another problem, which is a little bit extension of the other of the previous one. So, we consider a scenario where a company x, a service provider x sorry a company x, want to use cloud service from a provider p. So, there is a company x which wants to use a provider p like say IIT Kharagpur have to correct once more cloud service from a external one or any anything. The service level agreement guarantees negotiation negotiated between the 2 parties plat to initiating the business are as follows.

So, before that the service level guarantees are like this like availability guarantee is 99.595 percent time over the service period. So, the service period it should be 99.95 percent time. Availability period is 30 days maximum service hour per day is 12 hours, and cost say 50 dollar a day right. So, this is this is the type of agreement or type of requirement and that formal requirement which has been agreed upon with within the service provider and the service provider. So, availability 99.95 percent, service period 30 day maximum service hours per day is 12 hours, and cost is 50 dollar a day right. So, serve with credits are awarded to the consumer if availability guarantees are not satisfied right.

So, there is another part like if you if the provider fails to provide fails to provide service at the guaranteed level for which it has been agreed upon and the consumer is charging, then thus there has to pay the penalty right. So, penalty can be in terms of money detents,

or the penalty can be in terms of giving some extra compute hour or data or whatever way like

So, in this case availability can set monthly connectivity up time service level given as like a monthly up time percentage is less than 99.95 percent, but more than 99 percent greater than equal to 99 percent then the service get ride is 10 percent right. Whereas, of it is less than 99 percent then the service credit should be 25 percent right; however, in reality it was found that over the service period the cloud server support 5 outages right. During for these following durations, like one is 5 hours 30 minutes, one is 1 hour 30 minutes, one is 15 minutes, 2 hours 25 minutes each on different day. So, this due to which normal service guarantees where violated right; if unless if SLA negotiation are honored we need to compute the effective cost payable towards buying this cloud services right. So, this is this we need to check that how much, effectively need to be paid by the consumer to for this cloud services right.

So, that is fine. So, again just to quickly repeat. So, there are some of the guarantees are there availability 99.95 percent service period 30 days, maximum things 12 hours 50 dollar, and there are some penalty for not providing the services less than 99.95 percent, but greater than equal to 99 percent is 10 percent and less than 99 percent is to 25 percent. And there are 5 outages 5 hours 30 minutes, 1 hour 30 minutes, 15 minutes and 2 hour 25 minutes. I mean to find out that the effective cost payable towards the buying the cloud services. So, this we have to work on again not a difficult problem, but it gives us a idea the how things works.

(Refer Slide Time: 15:56)

Service period duration = 30 days  
Service duration per day = 12 hrs.  
Total service uptime =  $(12 \times 30)$  hrs = 360 hrs.  
Unit: \$ 50 per day.  
Total cost (at the time of service negotiation) =  $\$ (50 \times 30) =$   
\$ 1500  
Total downtime =  $(5 \text{ hr} + 30 \text{ min} + 1 \text{ hr } 30 \text{ min} + 15 \text{ min} + 2 \text{ hr } 25 \text{ min})$   
= 9 hrs 25 min.  
Service Availability =  $1 - \frac{\text{downtime}}{\text{uptime}} = \left(1 - \frac{9 \text{ hrs } 25 \text{ min}}{360 \text{ hrs}}\right) \times 100$   
= 97.385%

So, service period duration is 30 days right, 12 hours. So, total so, there for we have total so much hours, or 360 hours cost what we have seen 50 US dollar per day. So, total cost so this is at the time of at the time of service negotiation is dollar 50 cross 30 or this fine. So, these are the facts what we have given 30 days service duration per day is 12 hours as we are using, total service up time is expected this one 50 dollar this 50 per day is the cost and total cost at the time of service negotiation is, 15 to 30, 50, 1500 dollar, that is the thing.

Now, total service total service down time is 5 hours plus 30 minutes plus, 1 hour 30 minute plus 15 minute, plus 2 hours 25 minute right. So, if you add up it is 9 hours 25 minutes. So, this is the total outage time or the total down time for the things right. So, we can we can say service availability equal to 1 minus, this we have seen previously also and this is the standard thing 1 minus downtime by uptime equal to what we can say 1 minus 9 hours 25 minutes by 360 hours 100 so much percentage, so this fine. So, this was our total expected out time and this is the outage or the downtime, so 1 minus down time by so and so forth. And so, what we have the 97.385 percent. So, this is fine.

We calculate the service availability as 97.385 percent, so as per this data available.



(Refer Slide Time: 20:19)

So, what we see monthly up time percentage is 97.385 as we have calculated which is less than 99 percent right. Not only 99.5 percent, but like that 99 percent. So, service credit available, due to that whatever we whatever during that service negotiation or SLA things are there 25 percent of total cost. So, it is total cost as we have calculated 1500. So, it is dollar 375; so effective cost payable towards by buying the service equal to dollar 1500, minus dollar 375, equal to dollar 1125. So, this is the effective cost.

(Refer Slide Time: 22:09)

Monthly uptime % = 97.385% < 99%  
Service Credit available = 25% of total cost  
 $= \left(\frac{25}{100} \times \$1500\right) = \$375$   
Effective cost payable towards buying the cloud service  
 $= (\$1500 - \$375)$   
 $= \underline{\underline{\$1125}}$

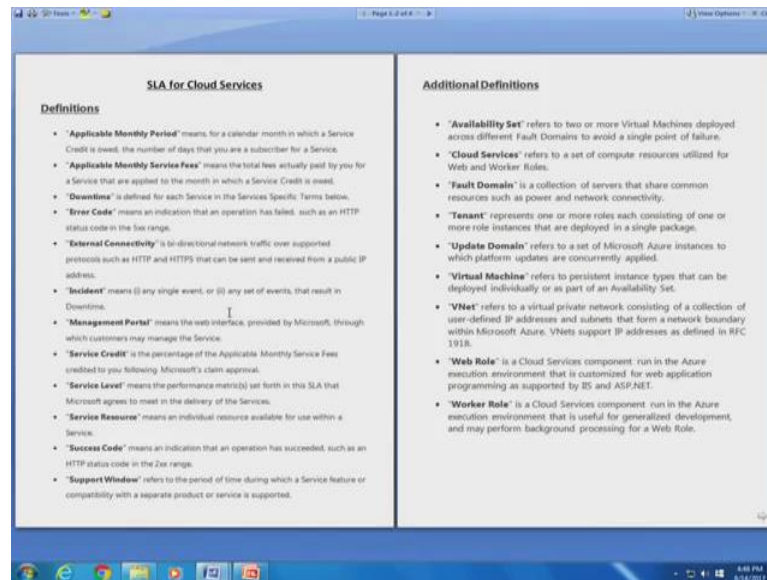
© CET  
IIT KGP

So, what we see that based on the outage. So, it is the if you look at the problem 2 it is the extension of the means little bit extension of the first problem. But this are in reality things happens right. So, we need to measure this things log this things, and calculate the things accordingly so right. This is with respect to the up time and there can be with respect to means it may not be total down, but you are availability band width may be may availability in the network may be slow and type of things.

It depends some lot of other aspects it is not that always straight forward, but there are lot of other complex consideration in doing so. So what we tried in this 2 problem. So, show that the how a SLA guarantees can be calculated or looked into an type of a means how it can be calculated and see that whether the violation of SLA or not right. So, this I believe this will give you a broad means board idea or a things of the.

So now what will see that in like what are the different base practices or what are the different components like one what we have calculated is the up time. So, what are the different components of a cell which are considered by primarily the commercial cloud, right?

(Refer Slide Time: 23:59)



So, that just couple of things will see right. So, so SLA for cloud service provider already we have seen, but just to. So some of the aspects which we would like to highlight this like in case of commercial cloud what things that there is applicable monthly period right, which means for a calendar month in which the service credit is odd the number of days you are a subscriber of a service right.

So, it is applicable service period similarly, applicable monthly service fees right. So, the concept of downtime like as we have seen services in the service specific terms below. Error code means the indication that the operation has failed such that http status code is 5xx or something right. So, that is services should have a error code otherwise you will it is difficult to try pinpoint we services has failed this type of thing.

External connectivity is a bidirectional network traffic support like protocol for http or https can received a public IP and so on and so forth, where they are external connectivity incident means any single or a set of incident that result in down time. Management portal means that the wave interface provided by this is basically meant for Microsoft azure. So, through which the consumer may manage the services like that

management portal. Service credit if there is a failure that how much credit will be given that we have seen in this problem. Service level means the performance mistakes says set forth in the SLA, and in case of myself azure, it agrees to meet the delivery of services service resource.

Success code like as we have seen failure code we have a success code like in a http we know that 2xx is the success code. Support windows refer to the period of time which during which the service features on compatibility with the separate product services is supported. So, there is a support window where the where the things will be there. Along with that there are some additional definitions like availability set refers to 2 or more virtual machine deployed across different fault domains. So, that it will not go for down time at the same time at the same period to avoid single point of failure. Cloud services refers to the compute resources utilized for web and web role and worker role. Fault domain is a collection of servers that share common resources which are power and network connectivity.

Tenant, represent one or more roles that is one or more role instances that it deployed in a single packages this we have seen like it can be a worker role it can be a web role type of thing. Update domain, refers to a set of in this case Microsoft azure in instances which platform update are concurrently applied, virtual machine this we know. VNet this is a virtual private network and this also is known that web role and worker role. So, these are some additional definitions which will be utilized for service level for SLA calculations, right.

(Refer Slide Time: 27:36)

**Monthly Uptime Calculation and Service Levels for Cloud Services**

- **"Maximum Available Minutes"** is the total accumulated minutes during a billing month for all Internet facing roles that have two or more instances deployed in different Update Domains.
- **"Downtime"** is the total accumulated minutes that are part of Maximum Available Minutes that have no External Connectivity.
- **"Monthly Uptime Percentage"** for Cloud Services is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month
  - $\text{Monthly Uptime \%} = \frac{\text{Maximum Available Minutes} - \text{Downtime}}{\text{Maximum Available Minutes}}$
- The following Service Levels and Service Credits are applicable to Customer's use of Cloud Services:

MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
< 99.95%	10%
< 99%	25%

So, similarly as we have seen here, in our example here if you can see the monthly up time calculation and service level for cloud services using those definitions say monthly available minutes is the total accumulated minutes during a billing period for all interfacing roles and 2 or more instances deployed in different update domain similarly down time is that how much time and up time percentage is, the maximum available minutes minus down time by maximum available minutes.

(Refer Slide Time: 28:16)

instances deployed in different Update Domains.

- **"Downtime"** is the total accumulated minutes that are part of Maximum Available Minutes that have no External Connectivity.
- **"Monthly Uptime Percentage"** for Cloud Services is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month
  - $\text{Monthly Uptime \%} = \frac{\text{Maximum Available Minutes} - \text{Downtime}}{\text{Maximum Available Minutes}}$
- The following Service Levels and Service Credits are applicable to Customer's use of Cloud Services:

MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
< 99.95%	10%
< 99%	25%

That we have calculated here right. So, that is the thing and there are can be service script credit rules as we have seen right, 99.95 percent in these exactly the same type of valves we have used.

(Refer Slide Time: 28:30)

**Monthly Uptime Calculation and Service Levels for Virtual Machines**

- **"Maximum Available Minutes"** is the total accumulated minutes during a billing month for all Internet facing Virtual Machines that have two or more instances deployed in the same Availability Set
- **"Downtime"** is the total accumulated minutes that are part of Maximum Available Minutes that have no External Connectivity.
- **"Monthly Uptime Percentage"** for Virtual Machines is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Microsoft Azure subscription.
  - $\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$
- The following Service Levels and Service Credits are applicable to Customer's use of Virtual Machines:

MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
< 99.95%	10%

Similarly, this is for calculation and service develop cloud services. Similarly we can have for the VMS right. So, VMS like I want to have infrastructure service and the virtual machines are allocated similarly. So, maximum available minutes is the total accumulated time is billing period and so on and so forth.

(Refer Slide Time: 28:47)

Available Minutes in a billing month for a given Microsoft Azure subscription.

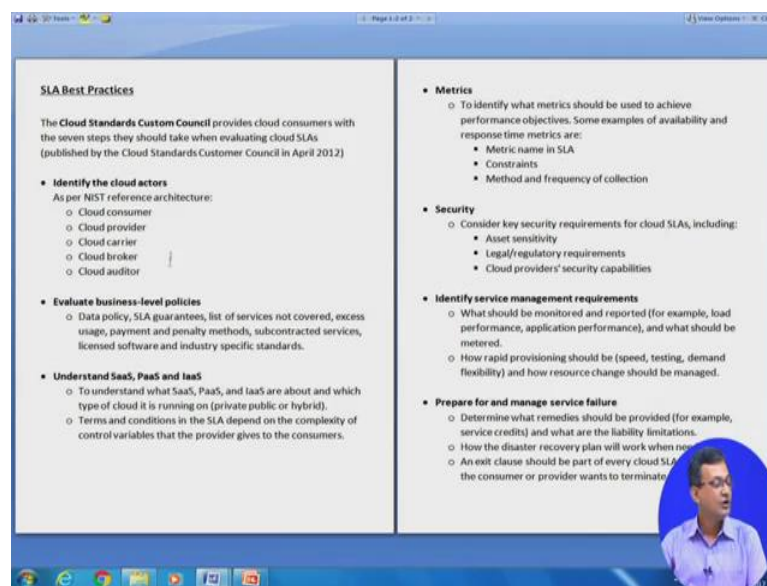
- $\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$
- The following Service Levels and Service Credits are applicable to Customer's use of Virtual Machines:

MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
< 99.95%	10%
< 99%	25%

Down time is similarly we can calculated. And we can have several separate same type of service credit. So, it can be at what we mean to say it can be a different type of level it can be at a IaaS level, it can be a PaaS levels any SaaS table. I can have as a storage level of the if there is storage downtime or access ability problem so on and so forth.

So, this need to be clearly specified, now when we want to do this type of thing; so what are the different best practices or rules we need to follow right. So, that way let us see some of the best practices what are follow.

(Refer Slide Time: 29:31)



So, the cloud standard custom council right, provides cloud consumer with 7 steps they are they should take when they evaluating the SLA's like it is also provided in their document of April 20, 12 right. So, identify the cloud actors. Who are the actor according to NIST architecture? So, these are the actors, consumer, producer or the provider, carrier, broker and auditor. So, these are the 5 actor which are there as far as NIST. So, we need to evaluate the business level policies right. This is important what will be the data policies SLA guarantee least of services, not covered under this excess uses payment penalty sub contract services, license software, industry specific standards and these are different aspects of the things.

So, what while we are discussing about simple SLA's in actually the things are more complex like what should be the data policy is how which are covered which are not covered, if there is a sub contract of services what should be that policies, whether you

are using license software licensing mechanisms and so on and so forth, right. Because most of the cases when we try to use these we may be using different license software, and those licensing cost etcetera come into play not only that licensing period, and so and other things come into play.

Then we need to understand that which level operations we are looking at, SaaS PaaS or IaaS, because the different type of things are the different type of services have different type of requirement. In some cases SaaS is much easier to control maybe or measure, but we need to look at that which type of services we are leveraging on, whether we are having multiple this sort of services. So, to we need to understand what SaaS, PaaS, IaaS are about and which type of cloud it is running whether is a public private or hybrid. Terms and condition in a SLA depends on the complexity of control variables that are provide that the provider gives to the consumer or the service consumers. So now consumer need to calculate the availability etcetera. So, for that the controlled control variables are provided by the like I say it gives me the CPU up time etcetera or different uses time or hard that disc uses parameters. So, these are the different control parameters provided.

Now, more the complexity of this parameter depends on which level of operations you are doing and where you are running the things, like is a IaaS, space or SaaS or whether it is a hybrid or your public or private. So, the other things are one is that metric what we are discussing about, to identify what metrics should be used to achieve performance objectives right. Some examples availability at availability response metrics are like metric name in the in SLA like availability and other type of things, there may be other constraints whether and frequency of collection of these data is also important. The aspect the next aspect is the security like, consider key security parameters for cloud including asset sensitivity, legal regulatory requirements like I a I may say that the that datas would reside within these particular geographically boundary or within this type of things. Cloud provider security capabilities what is the capability of the cloud provider to provide that.

Then we have service management requirement, to need to identify the service management requirement. So, what should be monitored and reported for example, load performance application performance or what should be metered right. What you are need to be bill meter. How rapid provisioning should be like speed, testing, demand

flexibility and how resource changed should be managed right. So, how is the provisioning, what need to be monitor and reported and meter type of things need to be looked into.

And then prepare for and manage for the failure right. There is another important exit. Determined what remedy should provided like for example, service credits and what are the liability limitation. So, how much service credits to provide and what are my liabilities on the provider the provider point of view, and in order to that what the consumer are signing of. How the disaster recovery plan will work when needed. So, how the disaster recovery plan will work when it is needed, and exit clause should be a part of every cloud SLA right. In either the consumer or the provider wants to terminate the relationship. So, SLA what it is there it is a agreement. So, what should be the exit clause suppose the consumer at some point of time to exit or the providers says that I am not able to provide that thing. So, that should be in the thing.

So, what we says that this are some of the essential best practices, or some best practices we should keep in mind when formulating the SLA etcetera. Like identify the cloud actors, evaluate business level policies, understand this type of services, what are the different matrixes security capability of the; and requirement security requirement of the consumer, and the capability of the provider. Service management requirements, and how to manage failure or what should be the remedies for failure.

So, what we tried this is a what we send a extension of the SLA already we have discussed. So, what we try to give that the there are different aspect to the things and try do in this thing which we have also seen to simple SLA related problem, how it can be how this type of SLA's are calculated though the problems are very simple and straight forward, but it gives us a idea that how you can apprise the approach this type of things. So, will let us conclude here for this SLA tutorial.

Thank you.