

PROBLEMS ON SLA

Courtesy:NPTEL Online Course on Cloud Computing by Prof. Soumya K. Ghosh, IIT Kharagpur

SLA

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

SLA Contents

- A set of services which the provider will deliver
- A complete, specific definition of each service
- The responsibilities of the provider and the consumer
- A set of metrics to measure whether the provider is offering the services as guaranteed
- An auditing mechanism to monitor the services
- The remedies available to the consumer and the provider if the terms are not satisfied
- How the SLA will change over time

Service Level Objectives (SLOs)

- Objectively measurable conditions for the service
- Encompasses multiple QoS parameters viz. availability, serviceability, billing, penalties, throughput, response time, or quality

Example :

- “Availability of a service X is 99.9%”
- “Response time of a database query Q is between 3 to 5 seconds”
- “Throughput of a server S at peak load time is 0.875”

Key Performance Indicators (KPIs)

- Low-level resource metrics
- Multiple KPIs are composed, aggregated, or converted to high-level SLOs.

Example :

- downtime, uptime, inbytes, outbytes, packet size, etc.
- Possible mapping :
 - Availability (A) = $1 - (\text{downtime}/\text{uptime})$

Monitoring:

- Natural questions:
 - “who should monitor the performance of the provider?”
 - “does the consumer meet its responsibilities?”
- Solution: neutral third-party organization to perform monitoring
- Eliminates conflicts of interest if:
 - Provider reports outage at its sole discretion
 - Consumer is responsible for an outage

Auditability

- Consumer requirement:
 - Is the provider adhering to legal regulations or industry-standard
 - SLA should make it clear how and when to conduct audits

– Metrics for Monitoring and Auditing

- Throughput – How quickly the service responds
- Availability – Represented as a percentage of uptime for a service in a given observation period.
- Reliability – How often the service is available
- Load balancing – When elasticity kicks in (new VMs are booted or terminated, for example)
- Durability – How likely the data is to be lost
- Elasticity – The ability for a given resource to grow infinitely, with limits (the maximum amount of storage or bandwidth, for example) clearly stated
- Linearity – How a system performs as the load increases

– Metrics for Monitoring and Auditing

- Agility – How quickly the provider responds as the consumer's resource load scales up and down
- Automation – What percentage of requests to the provider are handled without any human interaction
- Customer service response times – How quickly the provider responds to a service request. This refers to the human interactions required when something goes wrong with the ondemand, self-service aspects of the cloud.
- Service-level violation rate – Expressed as the mean rate of SLA violation due to infringements of the agreed warranty levels.
- Transaction time – Time that has elapsed from when a service is invoked till the completion of the transaction, including the delays.
- Resolution time – Time period between detection of a service problem and its resolution.

Example –Cloud SLAs

Cloud Provider	Service	Type of Delivery Model	Service Level Agreement Guarantees
Amazon	EC2	IaaS	Availability (99.95%) with the following definitions : Service Year : 365 days of the year, Annual Percentage Uptime, Region Unavailability : no external connectivity during a five minute period, Eligible Credit Period, Service Credit
	S3	Storage-as-a-Service	Availability (99.9%) with the following definitions: Error Rate, Monthly Uptime Percentage, Service Credit
	SimpleDB	Database-as-a-Service	No specific SLA is defined and the agreement does not guarantee availability
Salesforce	CRM	PaaS	No SLA guarantees for the service provided
Google	Google App Engine	PaaS	Availability (99.9%) with the following definitions : Error Rate, Error Request, Monthly Uptime Percentage, Scheduled Maintenance, Service Credits, and SLA exclusions

Example –Cloud SLAs

Cloud Provider	Service	Type of Delivery Model	Service Level Agreement Guarantees
Microsoft	Microsoft Azure Compute	IaaS/PaaS	Availability (99.95%) with the following definitions : Monthly Connectivity Uptime Service Level, Monthly Role Instance Uptime Service Level, Service Credits, and SLA exclusions
	Microsoft Azure Storage	Storage-as-a-Service	Availability (99.9%) with the following definitions: Error Rate, Monthly Uptime Percentage, Total Storage Transactions, Failed Storage Transactions, Service Credit, and SLA exclusions
Zoho suite	Zoho mail, Zoho CRM, Zoho books	SaaS	Allows the user to customize the service level agreement guarantees based on : Resolution Time, Business Hours & Support Plans, and Escalation

Example –Cloud SLAs

Cloud Provider	Service	Type of Cloud Delivery Model	Service Level Agreement Guarantees
Rackspace	Cloud Server	IaaS	Availability regarding the following: Internal Network (100%), Data Center Infrastructure (100%), Load balancers (99.9%) Performance related to service degradation: Server migration, notified 24 hours in advance, and is completed in 3 hours (maximum) Recovery Time: In case of failure, guarantee of restoration/recovery in 1 hour after the problem is identified.
Terremark	vCloud Express	IaaS	Monthly Uptime Percentage (100%) with the following definitions: Service Credit, Credit Request and Payment Procedure, and SLA exclusions

Example –Cloud SLAs

Cloud Provider	Service	Type of Cloud Delivery Model	Service Level Agreement Guarantees
Nirvanix	Public, Private, Hybrid Cloud Storage	Storage-as-a-Service	Monthly Availability Percentage (99.9%) with the following definitions: Service Availability, Service Credits, Data Replication Policy, Credit Request Procedure, and SLA Exclusions

$$- \textit{Availability (A)} = 1 - (\textit{downtime/up\textit{time}})$$

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.

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

Uptime 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.

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.

Service period duration = 30 days

Service duration per day = 12 hrs.

Total service uptime = $(12 * 30)$ hrs = 360 hrs.

Unit: \$ 50 per day.

Total cost (at the time of service negotiation) = $\$ (50 * 30) =$
\$1500

Uptime=360hrs-9hrs and 40 mins=350 hrs 20 mins

$$\begin{aligned}\text{Total downtime} &= (5\text{ hrs} + 30\text{ mins} + 1\text{ hr } 30\text{ mins} + 15\text{ mins} + 2\text{ hr } 25\text{ mins}) \\ &= \underline{9\text{ hrs } 25\text{ mins.}} + 15\text{ mins} = 9\text{ hrs } 40\text{ mins}\end{aligned}$$

$$\text{Service Availability} = 1 - \frac{\text{downtime}}{\text{uptime}} = \left(1 - \frac{9\text{ hrs. } 25\text{ mins.}}{360\text{ hrs}}\right) \times 100$$

$$= (1 - 9\text{ hrs } 40\text{ mins} / 350\text{ hrs } 20\text{ mins}) \times 100$$

$$= (1 - 9.66\text{ hrs} / 350.33\text{ hrs}) \times 100$$

$$= 0.97242 \times 100$$

$$= 97.24\%$$

$$= 97.385\%$$

Service period duration = 30 days

Service duration per day = 12 hrs.

Total service uptime = $(12 \times 30) \text{ hrs} = \underline{360 \text{ hrs.}}$

Unit: \$ 50 per day.

Total cost (at the time of service negotiation) = $\$ (50 \times 30) = \underline{\$1500}$

Uptime = 360 hrs - 9 hrs and 40 mins = 350 hrs 20 mins

Total downtime = $(5 \text{ hrs} + 30 \text{ mins} + 1 \text{ hr } 30 \text{ mins} + 15 \text{ mins} + 2 \text{ hrs } 25 \text{ mins})$
 $= \underline{9 \text{ hrs } 25 \text{ mins.}} + 15 \text{ mins} = \mathbf{9 \text{ hrs } 40 \text{ mins}}$

Service Availability = $1 - \frac{\text{downtime}}{\text{uptime}} = \left(1 - \frac{9 \text{ hrs. } 25 \text{ mins.}}{360 \text{ hrs}}\right) \times 100$

= $(1 - 9 \text{ hrs } 40 \text{ mins} / 350 \text{ hrs } 20 \text{ mins}) \times 100$

= $(1 - 9.66 \text{ hrs} / 350.33 \text{ hrs}) \times 100$

= 0.97242×100

= 97.24%

= 97.385%

Monthly uptime % = 97.385% < 99%.

$$\begin{aligned}\text{Service Credit available} &= 25\% \text{ of total cost} \\ &= \left(\frac{25}{100} * \$1500\right) = \$375\end{aligned}$$

$$\begin{aligned}\text{Effective cost payable towards buying the cloud service} \\ &= (\$1500 - \$375) \\ &= \underline{\underline{\$1125}}\end{aligned}$$