Comparative Analysis of Present Day Clouds using Service Level Agreements

Preeti Gulia, Assistant Professor
Department of Computer science and applications
M.D. University, Rohtak, Haryana, India

Sumedha Sood
Research Scholar
Department of Computer science and applications
M.D. University, Rohtak, Haryana, India

ABSTRACT
Cloud computing is a novel computing paradigm in which resources are delivered to end users on the basis of pay as you use model. In cloud computing a customer risks his entire business on a cloud. Thus he needs to be fully aware and confident about the various services that a provider claims to provide. The solution to the above problem comes in form of a legal contract that is established between the two. This contract is formally referred to as Service level agreements. These are short documents that contain various technical performance promises made by the provider. They also includes the penalties that provider would have to pay for performance failures. Thus, SLA makes the provider legally liable for all his claims about various services. SLA further helps a cloud consumer to select the best cloud by matching his requirements with what the provider specifies in its SLA. This paper explores the concept of SLA in cloud computing by quoting examples from real world. It also compares SLA of various present day clouds like rackspace, google apps, etc and proposes a new approach that would automatically select the new cloud on the basis of user requirements and SLA.

General Terms
Service level agreements of present day clouds

Keywords
Cloud computing, service level agreement, service credit, SLA exclusions

1. INTRODUCTION
Cloud computing is a new computing paradigm that provides an easy to use, everywhere, on-demand network access to shared computing resources such as hardware (storage, servers) and software (applications). These resources can further be rapidly supplied and released with least service provider interaction. There are three main fundamental service models according to which service is offered in cloud computing. These are namely Software as a Service, Platform as a Service and Infrastructure as a Service. [1]

Software as a service (SaaS) allows the cloud client to use any software and associated data that is centrally hosted on

the cloud [2]. It removes the overhead of installation of software on client’s personal computer. The software can be conveniently accessed by client by means of a thin client interface or by means of any program interface that should be present on the client device. Here, the client has no control over any server or operating system or storage of cloud. The entire cloud infrastructure also remains transparent to cloud client. However he can control user specific application configuration settings since their control may be essential for use of software [3].

For example Gmail is a well known SaaS. Here, the user uses the mailing software provided by Google to send and retrieve mails. The user need not install it on his computer, and can easily access it by a browser anywhere in world. Other well known examples of software as a service are Google Apps, Salesforce, and Zoho Office Suite.

Platform as a Service (PaaS) provides an environment that enables cloud client to create applications with the help of various tools that are provided by cloud provider. Similar to SaaS, the cloud consumer in PaaS need not worry about cloud infrastructure as well as network, storage, servers and operating system. However, he is provided full control of deployed applications and configuration settings of application hosting environment. The control of above is essential for developing applications [3].

For example Google App engine is a PaaS provider that provides cloud clients to develop applications using Google API. Other examples are lunacode, Engineyard, etc.

Infrastructure As A Service (IaaS) provider offers to the client the ability to access resources which include processing, storage, network and other fundamental computing resources. In this model the provider is the owner of the hardware such as storage, servers, etc. These are then provided to consumer. In contrast to the above two models, the consumer here has direct control over operating system, storage and deployed applications. However just like in above two models, the consumer does not manage cloud infrastructure. [3] Amazon EC2, Rackspace Cloud are some examples of cloud IaaS providers.
In addition to the above given service models, cloud computing offers four deployment models. As the word deployment indicates, deployment models define the nature of cloud delivery. The four Deployment models are: Private cloud, Community cloud, Public cloud and Hybrid cloud.

Private cloud is implemented with help of a firewall and is build exclusively for a single organization; the community cloud on the other hand is designed for a community of users from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). Public cloud can be used by general public. Hybrid cloud is combination of two or more distinct clouds.

While Private cloud and community cloud may exist on or off a premises, Public cloud exists on the premises of cloud provider [3]. Fig 2 shows all the deployment models. In cloud computing the terms of service are comprised of two legal documents [3]

1. Service Agreement.
2. Service Level Agreement (SLA)

While a service agreement comprises all the rules that make up the legal contract between consumer and provider, the service level agreement is a comparatively shorter document that includes information about all the services that the provider promises to provide. Service Level Agreement (SLA) also contains penalties that the provider pays in case of any SLA violation. [3] The amount of penalty varies from one provider to another and is distinctly specified in service level agreement. This paper focuses on service level agreement. A need has been identified to automate the process of cloud provider selection using service level agreements in cloud computing. As the number of providers is increasing, the need for the same increases even more. This is so, because it becomes quite difficult for a cloud client to choose the best provider from large number of providers. Some research work the same purpose has been done before also. Efforts have been put to enable automatic selection of optimal service offerings for cloud users on the basis of user requirement [4] [5]. Recently some work also has been done that adjusts public service level agreement templates according to user requirement [6]. Then various efforts to automate SLA negotiation process in service deployment phase have been made. [7] However, still a large amount of work can be done in this field as the service level agreements are still being offered as documents in the real world.

1.1 Benefits of Service Level Agreement

An unambiguous Service Level Agreement has following main benefits for both cloud client and provider [8]

- Clear service offer
On the client’s side, SLA allows the cloud client to compare different providers and choose the best one according to his requirements while on the provider’s side, it allows the provider to promote its service by stating all the service offers in it.

- Clarity on legal liabilities of provider and consumer
It precisely defines the liability for legal compliance, consequential losses and penalties that the provider would pay in case of failure to meet the SLA. These are important for both consumer and provider in order to be aware of risk associated with the service contract.

- Clearly identify roles and responsibilities of cloud provider and consumer
The service level agreement distinctly lists all the performance promises made by the provider and the conditions under which provider would not be responsible for service level agreement violation. For the client, it defines the procedure by which he can claim for the credit and the duration in which he/she will be eligible to apply for claim.

The paper is organized as follows. In section 2 the concept Service Level Agreement is briefly explained. Section 3 explains the contents of a service level agreement with examples from real world cloud SLAs. Then, section 4 compares the service Level Agreements offered by various providers. In section 5 the proposed work is described and finally paper concludes in section 6.
2. SERVICE LEVEL AGREEMENTS

Service Level Agreements are negotiated agreements that identify expectations of customer, clarify responsibilities of provider and facilitate communication between a service provider and its customer [9]

2.1 Service level objective

One of the key elements of SLA is service level objective. They act as a criterion for measuring performance of a given service. These are measurable terms that vary from one SLA to another according to application needs. In a cloud application an SLO can be throughput, availability [9] [10].

For e.g. in an SaaS SLA like Google apps it is mentioned that Google Apps Covered Services web interface will be available at least 99.9% of the time in any calendar month. Hence availability can be to be an SLO here [15].

If a client wants one SLO to be more important than other, he can use urgency rating. For instance, if for an application, throughput is more important than availability, then the client can use urgency ratings to rate them. [9][10]

2.2 Service Level Agreement Template

There are a large variety of service level agreements offered by providers in market. One of the reasons responsible for this can be lack of standardized computing resources in market. Hence, the need of hour at present is to standardize resources. Each of the standardized resource can be defined by means of service level agreement template. SLA template contains the structure of SLA [9]. The use of service level agreement template becomes even more important when there is an automated service discovery, since it becomes comparatively easy to compare different agreement of different providers with help of service level agreement template [8].

Many researchers have distinguished SLA templates into two categories: public SLA template and private SLA template. [5][6]

Private SLA templates are those that are used to specify requirements and new offers made by market participants. These are created by market participant at the time of joining the market. On the other hand, public SLA templates are used to describe products offered for trade in market. These are automatically generated and managed by market. This terminology is mostly used when automatic SLA matching and provider selection is to be done. It matches private SLA template (defined by user) with a number of public SLA templates and selects the one that is the best. Different researchers have defined different mapping techniques for the mapping public and private templates. [5][6]

Recently, some work has also been done to make public SLA templates dynamic according to user requirements. In this, it is assumed that the accepted public templates are stored in market directory and they are updated regularly on the basis of user requirements. [6] The discussion of entire work is not in the scope of this paper.

2.3 Roles and Responsibilities of Cloud Provider and Cloud Consumer

SLAs hold both the consumer and provider, responsible for certain actions [10]:

- It is the responsibility of both consumer and provider to agree to every policy, process and procedure that has been defined in service level agreement.
- It is the responsibility of provider to clearly and unambiguously state all the services that he intends to provide in the service level agreement. The conditions where provider does not want to hold himself responsible for SLA violation should also be clearly listed.
- It is the responsibility of provider to pay the promised service credit to the client in case of any SLA violation (except if it is occurred due to reason mentioned in SLA exclusions) without causing any inconvenience or trouble to client.
- It is the responsibility of provider to make a customer representative available when resolving a service related incident or request.
- The exact duration when the service would not be available because of maintenance or any other reason should be specified in SLA. If the customer has important any business function interfering with this duration he should inform the provider well before time.
- The provider should create and maintain system related documentation.
- The client before claiming for any SLA violation should make sure that the violation occurred and should be able to prove the same to provider. The client should not force the provider for any service credit after the given duration specified in SLA and he must follow the procedure stated in SLA while applying for credit.

3. CONTENTS OF SERVICE LEVEL AGREEMENT

While the contents of service level agreement vary in all SLAs. Following are the main sections found in most SLA’s. [9]

- Definitions
- Performance Management
- Service Commitment
- Problem Management
- SLA Exclusion
- Remedies
- Service credit
- Credit Request/claim

3.1 Definition

The first section of most service level agreement documents is Definition. Definition section gives a precise description of how various services are delivered. Definition of same service varies in different provider’s SLA according to their application requirements.

For example, as per Google app engine SLA definition of Monthly uptime percentage is “total number of minutes in a month, minus the number of minutes of Downtime suffered from all Downtime Periods in a month, divided by the total number of minutes in a month”. [11]

And Terremark defines the same Monthly uptime percentage in its SLA as: “It is calculated by subtracting from 100% the
percentage of fifteen minute periods during the Service Month in which vCloud Express is not available.”[12]

Definition of different SaaS, PaaS and IaaS providers is explained in section 4.

3.2 Performance Management

In cloud computing cloud providers manage the performance of their clouds by guaranteeing to provide a certain level of performance of various services without any failure to their customers. These commitments are stated by providers in their service level agreements.

3.2.1 The service commitments

The service level agreements form the building block of cloud computing. It is on the basis of these commitments only that a consumer is able to rely on a cloud. If there were no commitments from the provider side then no consumer would ever be able to risk their business on clouds. For example, consider a consumer who wants 100% availability for his application, then he would go for a cloud only if it guarantees 100% availability. Without such a guarantee there would always be a risk of loss for the cloud consumer. Hence service commitments are very essential part of SLA. While every cloud SLA contains service commitment, one example of service commitment from real world cloud is of App engine that guarantees 99.9% availability of Google Apps Covered Services web interface [11]. Generally, there is no such heading (service commitment) in SLA. However, such a guarantee is always present in every SLA. Service Commitments of leading SaaS, PaaS and IaaS are compared in table I in section 4.

3.3 Problem Management

One of the main motives of problem management is to prevent the problems from occurring. In SLA, problem management is done by means of SLA exclusions. In the SLA exclusion section, the provider informs the client about those problems for which he won’t be responsible and client has to himself prevent those from occurring. Thus, SLA exclusions help the provider to prevent the problems from occurring by informing the client to keep care of those conditions which are not his rather client’s responsibility.

3.3.1 SLA Exclusions

SLA Exclusions are present in almost every service level agreement. It lists all the possible points for which provider will not be responsible for SLA violation and thus would not pay any credit for the same. Following paragraph illustrates SLA exclusion from real world cloud:

Terremark promises monthly uptime of 100% and service credit in case it is not met. However, it lists in its exclusions that it won’t be responsible for unavailability when service suspension has occurred due to internet access problem, or any inactions caused by a third party. Thus, it won’t be entitled to pay service credit if the service is unavailable due to any point listed in service exclusions [12]

3.4 Remedies

Remedy can be defined as something that tends to repair an error or fault. In cloud computing SLA, service credit acts as remedy in case of any failure. Service credit is the penalty that the provider has to pay if he fails to meet the conditions that are specified in service level agreement. Service level agreement of many clouds states service credit as one of the sole and exclusive remedy for any failure. Service credit request / claim are explained in this section as they are a procedure for requesting service credit.

3.4.1 Service credit

If any cloud provider does not provide performance up to the level he had promised, then he has no other option but to pay penalty equivalent to that specified in service level agreement. This penalty is referred as service credit. However if the performance failure is due a reason mentioned in SLA exclusion, then the provider is saved from paying credit. This is so because provider is not responsible for the failure and the responsibility of failure lies on consumer himself. The service credits are clearly specified in service level agreement and vary in different documents. For example, Rackspace promises to pay service credit equivalent to 100% if the total block storage available time of any customer drops below 96.5% for the given billing cycle [13]. Most SLA’s also specify maximum credit they would pay. For example, Google App engine states in the definition of maximum financial credit that maximum credit it would pay won’t be greater than 50% of the amount pending in the monthly bills of given month[11].

Service Credits of SaaS, PaaS and IaaS are compared in column four of table I of section IV.

3.4.2 Service Credit Request /Claim

A customer is awarded a service credit only if he/she requests the same according to the procedure stated and within the duration specified in the Service Level Agreement. The customer is not entitled to receive any credit if he requests for the same after the duration specified in SLA. For example in Google App engine SLA, it has been specified that if a customer does not request for service credit by informing Google about SLA violation within thirty days of SLA violation he/she won’t be eligible anymore to receive any service credit[11]

While the procedure for the above is quite simple, the same is quite complex for a number of another cases

Credit request/claim with examples from SaaS, Paas and IaaS is explained with examples in section 4

4. COMPARISON OF SLA OF DIFFERENT CLOUD PROVIDERS

This section briefly compares SLA’s of different IaaS, PaaS and SaaS providers. It compares SLA of IaaS provider: Rackspace; PaaS provider Engine Yard and the SLA of SaaS provider: Google Apps. The first row of the table lists the cloud provider’s name, the second row lists the service commitment that the provider guarantees. The third row is for the definition section of SLAs, the third row is for claim in which the maximum duration under which cloud client can claim for a service credit is described and the fourth column is of credit i.e. penalty that the provider would have to pay for SLA violation and last column is for SLA exclusions. The definition section of each provider is discussed in detail after completing the table.
<table>
<thead>
<tr>
<th>Cloud Provider</th>
<th>RACKSPACE CLOUD SERVER</th>
<th>ENGINE YARD</th>
<th>GOOGLE APPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SERVICE COMMITMENT</td>
<td>1) Network—Rackspace guarantees availability of data center network 100% of the time (in any given monthly billing period) Maintenance period is not to be included&lt;br&gt;2) Data Center Infrastructure—Rackspace guarantees functioning of data center HVAC and power 100% of the time in any given monthly billing period. Maintenance period is not to be included&lt;br&gt;3) Cloud server Host—Rackspace guarantees proper functioning of all cloud server hosts including compute, storage and hypervisor and in case of any failure of cloud host, complete restoration will be done within one hour of problem identification.&lt;br&gt;4) Migration—Migration will be complete within three hours and in most cases customer would be notified at least twenty-four hours in advance.[13]</td>
<td>Engine Yard guarantees to provide at least 99.9% System Availability in any given month. Scheduling maintenance periods are not included in this.[15] System availability will be monitored by an external service. For this the customer will have to specify one or more URL. These URLs will be monitored by external service in order to track system availability. If there is any customer initiated maintenance then the customer will inform engine yard of the same by means of Engine yard supplied tool or API[14]</td>
<td>Google apps guarantees fully available and operational covered service web interface to every customer for least 99.95% of the time in any calendar month [15]</td>
</tr>
</tbody>
</table>

**Table I**

<table>
<thead>
<tr>
<th>DEFINITION</th>
<th>CREDIT REQUEST/CLAIM</th>
<th>SERVICE CREDIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Cloud server&lt;br&gt;2) Cloud server fees&lt;br&gt;3) Cloud server host&lt;br&gt;4) data center network&lt;br&gt;5) power&lt;br&gt;6) scheduled maintenance[13]</td>
<td>Client must claim for credit by means of Rackspace cloud control panel within thirty days following end of downtime and must also prove that SLA violation has adversely affected him in some way or another[13]</td>
<td>1) Network—For each 30 minutes of network downtime, Rackspace assures 5% of affected cloud server fees back&lt;br&gt;2) Data center infrastructure -- For each 30 minutes of infrastructure downtime, it promises to return 5% of affected cloud server fees back&lt;br&gt;3) Cloud server Host -- For each additional hour of downtime, it promises to return 5% of affected cloud server fees cloud server fees back&lt;br&gt;4) Migration-- For each additional hour of downtime, it assures back 5% of affected cloud server fees cloud server fees back [14]&lt;br&gt;• Up to 100% of cloud server fees can be returned for any of the above failures&lt;br&gt;• Maximum credit would not be more than 100% of cloud server fee for that particular month</td>
</tr>
</tbody>
</table>
Below the definition column of rackspace cloud is briefly explained.

1) Cloud Server – If it is possible to change Virtual private server at run time, it is referred to as cloud server. In Rackspace it is referred as unique virtual machine instance [13]

2) Cloud server fees - The fees for the client’s cloud servers for that monthly billing period in which failure took place is cloud server fees. It includes virtual machine instance and bandwidth charges [13]

3) Cloud server host—According to Rackspace SLA, that physical server that hosts client’s cloud server is cloud server host [13]

4) Data center Network— The entire network between the network egress point of client’s cloud server host to the outbound port of the data center border router is Data center Network[13]

5) Power—As per rackspace SLA, power includes UPSs, PDU sand cabling, and excludes the power supplies in cloud server hosts [13]

6) Scheduled maintenance—It would be announced in advance by rackspace at least ten business days and would not exceed sixty minutes in any calendar month[13]

Below the definition of engine yard are discussed

1) Scheduled maintenance—According to engine yard, it refers to certain infrequent maintenance actions done on the Engine Yard System. These can include adding resources, upgrading software, installing security patches, etc [14].

2) Scheduled Service Uptime — According to engine yard, it means the total minutes of System Availability in a month excluding the time for Scheduled Maintenance and Customer-initiated maintenance. [14]

3) Engine Yard System— It refers to the Engine Yard’s platform as a service systems that provides cloud Services to Customer. These include external DNS resolution, Internet connectivity, HTTP load balancing, etc [14]

4) System availability—According to engine yard, it is defined as its ability to answer client’s requests without failure. [14]

Below the definition section of Google app engine is discussed

1) Downtime—In Google apps downtime is said to occur if there is more than five percent user error rate. [15]

2) Google apps covered service—It includes Gmail, Google Calendar, Google Talk, Google Docs and Drive, Google Groups, Google Sites, and Google Apps Vault components of the Service. This excludes the Gmail Labs functionality, Google Apps - Postini Services, Gmail Voice or Video Chat components of the Service. [15]

3) Monthly Uptime Percentage —It is calculated in Google apps by

<table>
<thead>
<tr>
<th>SLA Exclusions</th>
<th>number of service credits that Google can issue to a customer will never be more than 15 days of service which will be added to end of customer’s term of service. Except for the customers who are on monthly plan, service credit may not be exchanged for or converted to monetary amounts[15]</th>
</tr>
</thead>
</table>
| 1) Rackspace does not consider itself responsible for any violation if the downtime is resulting due to denial of service attack, virus activity, hacking attempts or any other kind of circumstances that are not under Rackspace control. No credit will be given in above circumstances.[13] | It would not be Engine yard’s responsibility if the system availability falls (below the stated level in SLA) because of any of following[14] :
1) If the system availability falls due to factors which are not under Engine Yard’s control, including any force majeure event.[14]
2) It results from activity or fault of customer or any contractor or any third party that is acting on behalf of customer.[14]
3) Customer’s application or any equipment, software or technology that does not belong to Engine Yard.[14]
4) Scheduled maintenance[14]
5) If there is any suspension or termination of customer’s right to use a service according to terms of engine yards managed terms of service.[14]
|
| 2) No credit will be given if there is a discord between client and rackspace regarding client’s service agreement. The service credit will not be given until the client has cured the discord.[13] | No credit will be paid in above cases[14] Google Apps SLA is not applicable to
1) Services that expressly exclude the Google apps SLA[15]
2) Any performance issue that occurs due to factors stated in “Force Majeure” section of the agreement[15]
3) That resulted from any equipment not within Google’s primary control like customer’s equipment or any 3rd party equipment.[15] |
| 3) If any downtime occurs due to client’s misuse of service, Rackspace would not consider itself responsible and hence no credit would be given to customer.[13] | It refers to the Engine Yard’s managed terms of service. That resulted from any
1) Cloud Server – If it is possible to change Virtual private server at run time, it is referred to as cloud server. In Rackspace it is referred as unique virtual machine instance [13]
2) Cloud server fees - The fees for the client’s cloud servers for that monthly billing period in which failure took place is cloud server fees. It includes virtual machine instance and bandwidth charges [13]
3) Cloud server host—According to Rackspace SLA, that physical server that hosts client’s cloud server is cloud server host [13]
4) Data center Network— The entire network between the network egress point of client’s cloud server host to the outbound port of the data center border router is Data center Network[13]
5) Power—As per rackspace SLA, power includes UPSs, PDU sand cabling, and excludes the power supplies in cloud server hosts [13]
6) Scheduled maintenance—It would be announced in advance by rackspace at least ten business days and would not exceed sixty minutes in any calendar month[13] |
Points = rank (1st priority parameter) * N + rank (2nd priority parameter) * N-1 + ... + rank (last priority parameter) * 1 (this equation goes for all clouds)

- Then, at the final step the final rank of each cloud is calculated. The cloud getting the maximum points is given highest rank. The above steps are repeated for each user requirement.

This approach will make sure that the cloud assigned to the user will be the one that fulfills all his requirements in the best possible way.

6. CONCLUSIONS

In this paper the concept of service level agreements in cloud computing is discussed with examples of services offered by present day clouds. The main motive of taking examples of present day clouds is to give an idea of how various cloud providers offer service level agreements in real world. An effort to compare Service level agreements of various present day successful cloud providers are also is also made. In spite of the efforts made by many researchers to automate provider selection in SLA’s, these are still being used as documents today. As the number of providers is increasing in industry, the need for automatic provider selection is increasing even more. The proposed approach tries to find an answer to the same problem. We hope that this contribution of ours would help those working in area of cloud computing to get an idea of the different clouds offering services and help them further in their future work.

7. REFERENCES


[10] Kevin Buck, Diane Hanf, Cloud SLA consideration for the government consumer, September 2010


