# Comparing time series with machine learning based prediction approaches for violation management in Cloud SLAs.

Mr.Saurabh Singhal (saurabh.singhal@gla.ac.in),
Assistant Professor, Dept. of CEA, GLA University, Mathura.

Nikhil Bansal (nikhil.bansal_cs16@gla.ac.in), Nikunj Gupta (nikunj.gupta_cs16@gla.ac.in),
Rishabh Agarwal (rishabh.agarwal_cs16@gla.ac.in), Tushar Verma (tushar.verma_cs16@gla.ac.in)

## 1. Abstract

The project will determine the SLA Violations between Cloud Service Provider and Cloud Consumer and on that basis we will calculate the penalties that cloud service provider will have to pay due to the violation of SLA terms and conditions. Time series algorithm of Machine Learning will be used to calculate the penalties that service provider will pay for the violation of SLA (Service Level Agreement) terms and conditions. This project can be widely used in IT industry to ensure service availability to consumer w.r.t. all terms and conditions and it will result in increased productivity.

## 2. Introduction

Cloud computing is being adapted by a growing number of individuals and enterprises due to its wide range of services, including the elastic scaling of resources, automatic service deployment and virtualized resources with its benefits of being economical and easily manageable in nature. Due to these features, cloud computing has become the first choice for many small to large organizations. Gartner Research states that cloud computing is a rapidly emerging technology on which organizations spent an estimated $677 billion from 2013 to 2016. According to a survey Conducted by an IT decision maker for large companies, more than half of the respondents (68%) expected that 50% of their I.T. resources would be migrated to cloud platform.

In cloud computing, service level agreement(SLAs) are legal agreements between a service provider and consumer that contain a list of obligations and commitments which need to be satisfied by both parties during the transaction.

Violation of such a commitment leads to penalties in terms of money and reputation and thus has to be effectively managed.

An SLA is a legal contract which includes service obligations, deliverability, service objectives and service violation penalties. An SLA is not only used to measure the performance of the provider, it also helps to resolve disputes regarding consumer duties. An SLA comprises one or more objectives, called service level objectives (SLO), which comprise one or many low-level metrics.

Violations in SLAs leads to penalties in form of money and loss of reputation and thus it needs to be effectively managed. Our analysis will help to test and report the accuracy of time series based machine learning prediction approaches and help the cloud service provider to choose an appropriate prediction approach and utilize the best method depending on input data patterns. We are going to propose a new framework in order to obtain the best method for predicting the QoS based on the input and output pattern, CPU, and memory. This approach ranks the different prediction approaches according to its predicting accuracy, and this determines the SLA Violations. Using this method, the service providers select the method since incorrect adoption causes SLA violation and penalties.

We will test and report the accuracy of the time series and machine learning based prediction approaches. Our analysis helps the cloud service provider to choose an appropriate prediction approach and further to utilize the best method depending on input data patterns to obtain an accurate prediction result and better manage their SLAs and to avoid violation penalties.

The project will determine the SLA Violations between Cloud Service Provider and Cloud Consumer and on that basis we will calculate the penalties that cloud service provider will have to pay due to the violation of SLA terms and conditions. For more precise results time series algorithm of Machine Learning will be used to calculate the penalties that service provider will pay for the violation of SLA (Service Level Agreement) terms and conditions. This project can be widely used in IT industry to ensure service availability to consumer w.r.t. all terms and conditions and it will result in increased productivity.

## 2.1 Cloud SLAs

A cloud SLA (cloud service-level agreement) is an agreement between a cloud service provider and a customer that ensures a minimum level of service is maintained. It guarantees levels of reliability, availability and responsiveness to systems and applications, while also specifying who will govern when there is a service interruption.

A cloud infrastructure can span geographies, networks and systems that are both physical and virtual. While the exact metrics of a cloud SLA can vary by service provider, the areas covered are uniform: volume and quality of work -- including precision and accuracy -- speed, responsiveness and efficiency. The document aims to establish a mutual understanding of the services, prioritized areas, responsibilities, guarantees and warranties provided by the service provider.

An SLA will commonly use technical definitions that quantify the level of service, such as mean time between failures or mean time to repair, which specifies a target or minimum value for service-level performance.

A typical compute and cloud SLA articulates precise levels of service, as well as the recourse or compensation the user is entitled to should the provider fail to deliver the service as described. Another area to consider carefully is service availability, which specifies the maximum amount of time a read request can take; how many retries are allowed; and so on.

The SLA should also define compensation for users if the specifications aren't met. A cloud storage service provider usually offers a tiered service credit plan that gives users credits based on the discrepancy between SLA specifications and the actual service levels delivered.

## 2.2 Importance of Cloud SLAs

- Overview, dates of contract, numbers, cost, and contact information, etc.
- Definition of the service you're receiving
- The performance you can expect
- How processes and performance levels will be monitored
- How data privacy or compliance concerns will be manage
- How to report any issues to the provider
- How long it takes to resolve issues
- A process for how to handle contract changes

## 2.3 Violation in Cloud SLAs

Cloud prefers or are obliged to have a certain Service Level Agreement (SLA) in place before delivering any service to its valuable customers. The SLA sets a definite time frame in which services are to be delivered to the valuable customers. If they are not delivered within the proposed time frame, SLA violation occurs and Service Provider is imposed by a penalty.

## 2.4 Penalties on SLA Violation

Service level agreement penalties will vary from client to client and consumer to consumer. When they are being drafted, several parameters for these penalties should be considered. These are:

**Service availability: -** This involves factors like network uptime, database availability, and data center resources. Penalties should be included as a deterrent against service downtime, as such downtime could negatively affect business productivity.

**Service quality:** - This involves the guarantee of performance, the number of defects or errors allowed in a product or service, process gaps, and other quality issues. One approach to consider is to levy a penalty for every failure that was made to meet these objectives.

There are a variety of penalties that may be incurred from service level violations. The three most common are:

**Financial penalties:** - With these, the vendor will be required to pay back to the customer the amount of damages that was agreed upon in the contract. This may not amount to a full reimbursement of the service fee paid by the customer for the job.

**Service credits:**-With these, the vendor will reimburse the customer for the cost of the work that was done or offer credit for future work to be done. In either event, actually funds are not being transferred.

**License extension or support:**-With this, the vendor will be required to extend the license's term or offer further support to the customer without charge, which may include development and maintenance.

Such penalties must be set out in the language of the service contract, otherwise, they will not be enforceable. Furthermore, of these penalties, the service credit and license extension penalty may not be considered adequate compensation by some, as some might question the value of receiving the continued services of a provider that fails to meet its quality levels. Rather, employing a combination of penalties may be a better approach, while at the same time including an incentive like a monetary bonus for satisfactory or beyond satisfactory work.

## 2.5 SLA Levels

Service level agreements are also defined at different levels:

- **Customer-based SLA**: An agreement with an individual customer group, covering all the services they use. For example, an SLA between a supplier (IT service provider) and the finance department of a large organization for the services such as finance system, payroll system, billing system, procurement/purchase system, etc.
- **Service-based SLA**: An agreement for all customers using the services being delivered by the service provider**.** For example:
    - A mobile service provider offers a routine service to all the customers and offers certain maintenance as a part of an offer with the universal charging.
    - An email system for the entire organization. There are chances of difficulties arising in this type of SLA as level of the services being offered may vary for different

customers (for example, head office staff may use high-speed LAN connections while local offices may have to use a lower speed leased line).

- **Multilevel SLA**: The SLA is split into the different levels, each addressing different set of customers for the same services, in the same SLA.
    - **Corporate-level SLA**: Covering all the generic service level management (often abbreviated as SLM) issues appropriate to every customer throughout the organization. These issues are likely to be less volatile and so updates (SLA reviews) are less frequently required.
    - **Customer-level SLA**: covering all SLM issues relevant to the particular customer group, regardless of the services being used.
    - **Service-level SLA**: covering all SLM issue relevant to the specific services, in relation to this specific customer group.

## 2.6 MACHINE LEARNING

Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as e-mail filtering and computer vision, where it is difficult or infeasible to develop conventional algorithms to perform the needed tasks.

Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data Mining is a related field of study, focusing on exploratory analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

Machine learning involves computers discovering how they can perform tasks without being explicitly programmed to do so. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than have human programmers specify every needed step.

The discipline of machine learning employs various approaches to help computers learn to accomplish tasks where no fully satisfactory algorithm is available. In cases where vast numbers of potential answers exist, one approach is to label some of the correct answers as valid. This can then be used as training data for the computer to improve the algorithm(s) it uses to determine correct answers. For example, to train a system for the task of digital character recognition, the MNIST dataset has often been used.

## 2.7 MACHINE LEARNING APPROACHES

Early classifications for machine learning approaches sometimes divided them into three broad categories, depending on the nature of the "signal" or "feedback" available to the learning system. These are:-

**Supervised Learning**: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.

**Unsupervised learning**: **No** labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).

**Reinforcement learning**: **A** computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent) As it navigates its problem space, the program is provided feedback that's analogous to rewards, which it tries to maximize.

Other approaches or processes have since developed that don't fit neatly into this three-fold categorization, and sometimes more than one is used by the same machine learning system. For example topic modeling, dimensionality reduction or meta learning. As of 2020, deep learning had become the dominant approach for much ongoing work in the field of machine learning .

## 2.8 APPLICATIONS OF MACHINE LEARNING

- Agriculture
- Anatomy
- Adaptive Websites
- Affective Computing
- Banking
- Bioinformatics
- Cheminformatics
- Citizen Science

- Computer Networks
- Computer Vision
- Data Quality
- Information Retrieval

## 2.9 MACHINE LEARNING METHODS

- Regression
- Classification
- Clustering
- Dimensionality Reduction
- Ensemble Methods
- Neutral Nets and Deep Learning
- Transfer Learning
- Reinforcement Learning
- Natural Language Processing
- Word Embeddings

## 2.10 LIMITATIONS OF MACHINE LEARNING

Although machine learning has been transformative in some fields, machine-learning programs often fail to deliver expected results. Reasons for this are numerous: lack of (suitable) data, lack of access to the data, data bias, privacy problems, badly chosen tasks and algorithms, wrong tools and people, lack of resources, and evaluation problems.

In 2018, a self-driving car from Uber failed to detect a pedestrian, who was killed after a collision. Attempts to use machine learning in healthcare with the IBM Watson system failed to deliver even after years of time and billions of investment.

## 2.11 PANDAS LIBRARY

Pandas is quite a game changer when it comes to analyzing data with Python and it is one of the most preferred and widely used tools in data munging/wrangling if not THE most used one. Pandas is an open source, free to use (under a BSD license) and it was originally written by Wes McKinney .

What's cool about Pandas is that it takes data (like a CSV or TSV file, or a SQL database) and creates a Python object with rows and columns called data frame that looks very similar to table in a statistical software (think Excel or SPSS for example. People who are familiar with R would see similarities to R too). This is so much easier to work with in comparison to working with lists and/or dictionaries through for loops or list comprehension.

## 2.12 Sample SLA

### Cloud Service Level Agreement

This Cloud Service Level Agreement is part of your Cloud Terms of Service.

**1. DEFINITIONS**

The following definitions shall apply for purposes of this Cloud Service Level Agreement:

"**Availability**" means it is guaranteed services will be available 99.99%.

"**Affected Components**" means the Service(s) that have been affected by the failure to meet a Service Level Guarantee and include the initial Service that failed plus any Service(s) which become inaccessible or unable to perform their intended purpose under the Agreement as a result of the initial Service's failure.

"**API Error**" means (i) a HTTP server error response to a Valid API Request or (ii) no response to a Valid API Request because the API is down. Network errors or downtime outside of the Data Center Network do not constitute an API error.

"**Cluster**" means a group of Cloud Big Data Platform virtual servers preconfigured with Hadoop software, comprising of a Control Plane and the Data Plane.

"**Cloud Database Instance**" means your unique database instance and the availability of the storage volume provisioned for your instance, as well as the API you use to administer the Cloud Databases Service.

"**Monitoring Alert**" means a device being monitored using the Monitoring Services violates a predefined error condition, and some form of notification is generated.

"**Monitoring Services**" means the infrastructure, software, and services responsible for monitoring devices and generating or tracking Monitoring Alerts.

"**Control Plane**" means the API used to create, manage, and delete your Cloud Server or (for Cloud Big Data Platform Services) your Cluster(s).

"**Cloud Server**" means a unique virtual machine instance, save for OnMetal Cloud Servers Services where Cloud Server shall mean a physical server instance.

"**Credit Percentage**" means a credit amount calculated as a percentage of the Service Fees (and for Cloud Queues includes bandwidth charges, and for Cloud Servers includes hourly virtual machine usage and bandwidth charges where applicable) in the affected Region.

"**Data Center Infrastructure Downtime**" means a Host Server experiences a failure as a result of Power or heating problems.

"**Data Center Network**" means the portion of the cloud network extending from the network egress point of your Host Server, or for Cloud Files Services the cloud edge device, to the outbound port of the data center border router.

"**Data Center Network Downtime**" means a Host Server is not reachable as a result of a failure in the Data Center Network.

"**Data Plane**" means the Cloud Servers or (for Cloud Big Data Platform Services) Clusters created via the applicable Control Plane plus supporting systems and services required for the proper functioning and availability of those Cloud Servers or Clusters.

"**DB Server**" means the supplied and configured software and hardware for which DBA Services have been purchased, but specifically excluding customer provided software and any hardware or service which has been labelled "collocated", "non-standard", "unsupported" or with similar designation.

"**Host Server**" means the physical server which hosts your Cloud Server or Cloud Databases Service, or (for Cloud Big Data Platform Services) the group of physical servers which host your Cluster, as applicable.

"**Load Balancer**" means your unique cloud load balancer instance.

"**Monthly Availability**" means a monthly availability percentage calculated on as per customer, per Region basis, for a given monthly billing period, as follows: 1 – (total API Errors)/ (total Valid API Requests).

"**Power**" includes uninterruptible power supplies, power distribution units and cabling but excludes Host Server power supplies.

"**Region**" means a particular physical data center location where the Host Server exists. Control Planes and Data Planes are independent per region and do not span regions.

"**Server Error Response**" means an HTTP return status code between 500-599.

"**Service Fees**" means the fees paid for the Services to which the relevant Service Level Guarantee applies, in the monthly billing period in which the event giving rise to a credit first occurred. "**Valid API Request**" is defined as a well formed request that complies with the published API specification.

"**Volume**" means a logically identified container or dataspace provisioned within the Cloud Block Storage Service for your use.

Capitalized terms not defined in this Service Level Agreement shall have the meaning ascribed to them elsewhere in the Agreement.

**2. CLOUD AVAILABILITY SERVICE LEVEL GUARANTEE.**

We guarantee that the services will be available 99.9% of the time in a given monthly billing period. Volumes shall be deemed available unless: (i) the Cloud Service returns a Server Error Response to a

Valid API Request during two or more consecutive ninety (90) second intervals, or (ii) data stored on Volumes becomes inaccessible to the applicable Cloud Server. If we fail to meet this guarantee, you will be eligible for a credit calculated as a percentage of the Service Fee for the affected Volume, and based on the percentage of Volume availability in a given monthly billing period as follows:

| **Cloud availability** | **Credit amount** |
|---|---|
| 99.89% - 99.5% | 10% |
| 99.49% - 99.0% | 25% |
| 98.99% - 98.0% | 40% |
| 97.99% - 97.5% | 55% |
| 97.49% - 97.0% | 70% |
| 96.99% - 96.5% | 85% |
| Less than 96.5% | 100% |

**3. CLOUD DATABASE INSTANCES SERVICE LEVEL GUARANTEE.**

We guarantee that your Cloud Database Instance(s) will be available 99.9% of the time in any given monthly billing period. If we fail to meet this guarantee, you will be eligible for a credit calculated as a percentage of the Service Fees, as follows: Five percent (5%) of the Service Fees for each 30 minutes of Cloud Database Instance unavailability, after the first 0.1% of unavailability during the month, up to one hundred percent (100%) of the Service Fees. You shall not be entitled to a credit if (ii) the event giving rise to the credit would not have occurred but for the resizing any Cloud Database Instance(s) upon your instructions (including by increasing or decreasing the amount of storage or compute resources dedicated to a given Cloud Database Instance), or (ii) where the applicable Cloud Database Instance(s) are root enabled.

**4. CLOUD DBA MONITORING SERVICE LEVEL GUARANTEE.**

Cloud Provider will notify you via text message, email, or the ticketing system of Monitoring Alerts on your DB Server(s) within five (5) minutes of the Monitoring Alert being generated. If we fail to meet this guarantee you are entitled to a credit in the amount of five percent (5%) of your Service Fees per event, up to one hundred percent (100%) of your monthly Service Fees. You shall not be entitled to a credit if the event giving rise to the credit would not have occurred but for your failure to implement the recommended configuration changes or upgrades provided to you by support teams (the "**Configuration Guidance**") or otherwise address the issue in a manner which permits to provide the DBA Services within five (5) days of your receipt of the Configuration Guidance, or such other time frame as stated in the Configuration Guidance.

**5. CLOUD FILES SERVICE LEVEL GUARANTEE.**

We guarantee that the Cloud Files Services will be available 99.9% of the time in a given billing cycle. The Cloud Files Services will be deemed available unless: (i) the Data Center Network is experiencing Data Center Network Downtime, or (ii) the Cloud Files Service returns a Server Error Response to a Valid API Request during two or more consecutive ninety (90) second intervals, or (iii) the global network of servers designed to expedite delivery of web content by serving the content from a location in geographic proximity to the user, known as a Content Delivery Network, fails to deliver an average

download time for a 1-byte reference document of 0.3 seconds or less, as measured by third party measuring service. If we fail to meet this guarantee you will be eligible for a credit calculated as a percentage of your Service Fees, and based on the duration of the availability in a given monthly billing period as follows:

| **Cloud Files availability** | **Credit amount** |
|---|---|
| 99.89% - 99.5% | 10% |
| 99.49% - 99.0% | 25% |
| 98.99% - 98.0% | 40% |
| 97.99% - 97.5% | 55% |
| 97.49% - 97.0% | 70% |
| 96.99% - 96.5% | 85% |
| Less than 96.5% | 100% |

## 6. CLOUD LOAD BALANCERS SERVICE LEVEL GUARANTEE.

We guarantee that your Load Balancers will be available 99.99% of the time in any given monthly billing period. Load Balancers will be deemed available unless network traffic is unable to reach the load balanced servers or endpoints as appropriate. In the event of a Load Balancer failure, the system shall failover to a partner device resulting in less than 30 (thirty) seconds of Load Balancer unavailability. If we fail to meet this guarantee you will be eligible for a credit calculated as a percentage of the Service Fee (including monthly instance, concurrent connections and bandwidth charges) as follows: Five percent (5%) of the Service Fees for each hour of unavailability or part thereof, up to one hundred percent (100%) of the Service Fees.

## 7. MONITORING SERVICES

7.1 The Monitoring Services consist of the Remote Monitoring Service and the Deep Server Monitoring Service as described below. The Remote Monitoring Check and Deep Server Check are together the "**Checks**".

7.2 **Remote Monitoring Service**. With the Remote Monitoring Service you can check the availability of remote web services by attempting to create an active connection to a specified port (a "**Remote Monitoring Check**"). You may use Remote Monitoring Checks on any IP address (or related hostname) that you own, whether hosted by Rackspace or at another location. You may not configure Remote Monitoring Checks against IP addresses that you do not own, and while you may configure Remote Monitoring Checks against dynamic IP addresses you must disable or modify the applicable Remote Monitoring Checks when the target IP is re-assigned (for this reason we suggest that you point any Remote Monitoring Check running against a dynamic IP target to a DNS entry that is updated when the target IP changes).

7.3 **Deep Server Monitoring Services**. With the Deep Server Monitoring Service, you can monitor your server resources, including Memory, CPU, Disk, and Network ("**Deep Server Check**"). You may not configure Deep Server Checks against servers that you are not authorized to monitor and you must disable or modify the applicable Deep Server Checks if your authority with the respect to the servers ends.

7.4 **Monitoring Services Service Level Guarantee**. We guarantee that the Monitoring Services will be available 99.9% of the time in any given monthly billing period. Monitoring Services are considered available if they perform properly configured Checks and notify you of any failed Check within thirty (30) minutes of performing the Check. If we fail to meet this guarantee in any given monthly billing period, you will be eligible for a one-time credit of ten (10) US dollars. If you are paying for your Services in a currency other than US dollars, then any credit that is expressed above as a fixed number of US dollars will be converted to the currency in which you pay for your services as of the date of the invoice on which the credit is applied. You will not be entitled to credits related to your misconfiguration of the cloud system or Monitoring Services agent (including if you are not using the most current version of the Monitoring Services agent), or for any failures relating to the networking issues between you and the Monitoring Services systems.

## 8. CLOUD SERVERS, ONMETAL CLOUD SERVERS, & CLOUD BIG DATA PLATFORM SERVICE LEVEL GUARANTEES

8.1 **Control Plane Service Level Guarantee.** This Control Plane Service Level Guarantee applies only to Cloud Server (Next Gen), OnMetal Cloud Server, and Cloud Big Data Platform Services. We guarantee the Control Plane shall have 99.9% Monthly Availability. If we fail to meet this guarantee you will be eligible for a credit, calculated as follows:

| **Monthly Availability** | **Credit Percentage** |
|---|---|
| < 99.9% - 99.5% | 10% |
| < 99.5% - 99.0% | 20% |
| < 99.0% | 30% |

8.2 **Data Plane Service Level Guarantees.** Save as expressly stated to the contrary below, these Data Plane Service Level Guarantees apply to Cloud Server (First Gen), Cloud Server (Next Gen), OnMetal Cloud Server, and Cloud Big Data Platform Services.

8.2.1 **Network Availability Service Level Guarantee.** It guarantees that its Data Center Network will be available one hundred percent (100%) of the time in any given monthly billing period. The Data Center Network will be deemed available unless a Cloud Server is not reachable due to Data Center Network Downtime. If we fail to meet this guarantee you will be eligible for a credit, as follows: Five percent (5%) of the Service Fees for each thirty (30) minutes of Data Center Network Downtime (or portion thereof), up to one hundred percent (100%) of the Service Fees.

8.2.2 **Data Center Infrastructure Service Level Guarantee.** It guarantees that data center HVAC (heating, ventilation, and air conditioning) and Power will be functioning one hundred percent (100%) of the time in any given monthly billing period. Data center HVAC and Power will be deemed functioning unless experiencing Data Center Infrastructure Downtime. If we fail to meet this guarantee, you will be eligible for a credit, as follows: Five percent (5%) of the Service Fees for each thirty (30) minutes of Data Center Infrastructure Downtime (or portion thereof), up to one hundred percent (100%) of the Service Fees.

8.2.3 **Cloud Server Hosts Service Level Guarantee.** It guarantees the functioning of all Host Servers hosting the Cloud Servers or Cluster (including the hypervisor, where applicable). If a Host Server hosting a Cloud Server or Cluster fails, It guarantee that restoration or repair will be complete within one (1) hour of problem identification. If we fail to meet this guarantee you will be eligible for a credit, as follows: Five percent (5%) of the Service Fees for each additional hour of unavailability (or portion thereof), after the first hour, up to one hundred percent (100%) of the Service Fees.

8.2.4 **Migration Service Level Guarantee.** This Migration Service Level Guarantee does not apply to OnMetal Cloud Server or Cloud Big Data Platform Services. If a Cloud Server migration is required because of Host Server degradation, we will notify you at least twenty four (24) hours in advance of beginning the migration, unless we determine in our reasonable judgment that we must begin the migration sooner. Either way, we guarantee the migration will be complete within three (3) hours of the time we begin the migration. If we fail to meet this guarantee you will be eligible for a credit, as follows: Five percent (5%) of the Service Fees for each additional hour of unavailability (or portion thereof), after the first three (3) hours, up to one hundred percent (100%) of the Service Fees.

## 9. CLOUD SITES SERVICE LEVEL GUARANTEE.

It guarantees that its Cloud Sites Services will be available one hundred percent (100%) of the time in any given monthly billing period. If we fail to meet this guarantee, you will be eligible for a credit, as follows: the equivalent of one (1) day's Service Fees for each sixty (60) minutes of Cloud Sites Services unavailability (or portion thereof), up to one hundred percent (100%) of the Service Fees.

## 10. CLOUD SUPPORT SERVICE LEVEL GUARANTEE.

This Cloud Support Service Level Guarantee applies to Managed Infrastructure and Managed Operations customers only. You are entitled to a credit of five percent (5%) of the Service Fees for the Affected Components for every half hour (or part thereof) of unplanned unavailability caused by Support activities, up to a maximum of a hundred percent (100%) of the Service Fees for such Affected Components. The Services shall be deemed unavailable if they fail to respond to the Monitoring Services and are unable to send and receive information for the purpose in which they have been provisioned. If the event for which a credit may be payable under this guarantee is an event which would also be eligible for a credit under another SLA or is an event contemplated by another SLA, you shall not be eligible for a credit under this guarantee. You shall not be entitled to a credit in the event that you fail to implement appropriate Monitoring Services; or disable, block, remove, or otherwise interfere with the Monitoring Services.

## 11. CLOUD QUEUES API SERVICE LEVEL GUARANTEE.

We guarantee that the Cloud Queues API shall have 99.9% Monthly Availability. The rate limit for the Cloud Queues Service is 300 requests per account per second. If you exceed this limit, your Cloud Queues Service may be throttled or stopped, and this guarantee shall not be applicable. If we fail to meet this guarantee you will be eligible for a credit, calculated as follows:

| **Monthly Availability** | **Credit Percentage** |
|---|---|

| | |
|---|---|
| < 99.9% - 99.5% | 10% |
| < 99.5% - 99.0% | 20% |
| < 99.0% | 30% |

## 3. System Architecture

### 3.1 System Modules

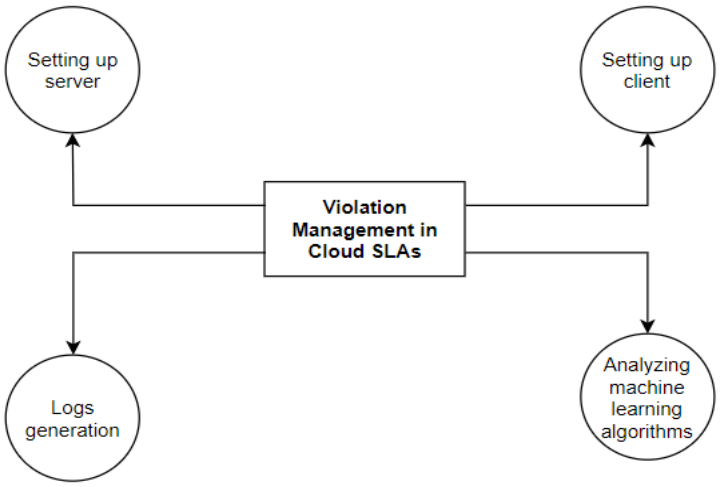


**Fig 3.1: System Modules**

## 3.2 Data Flow Diagram

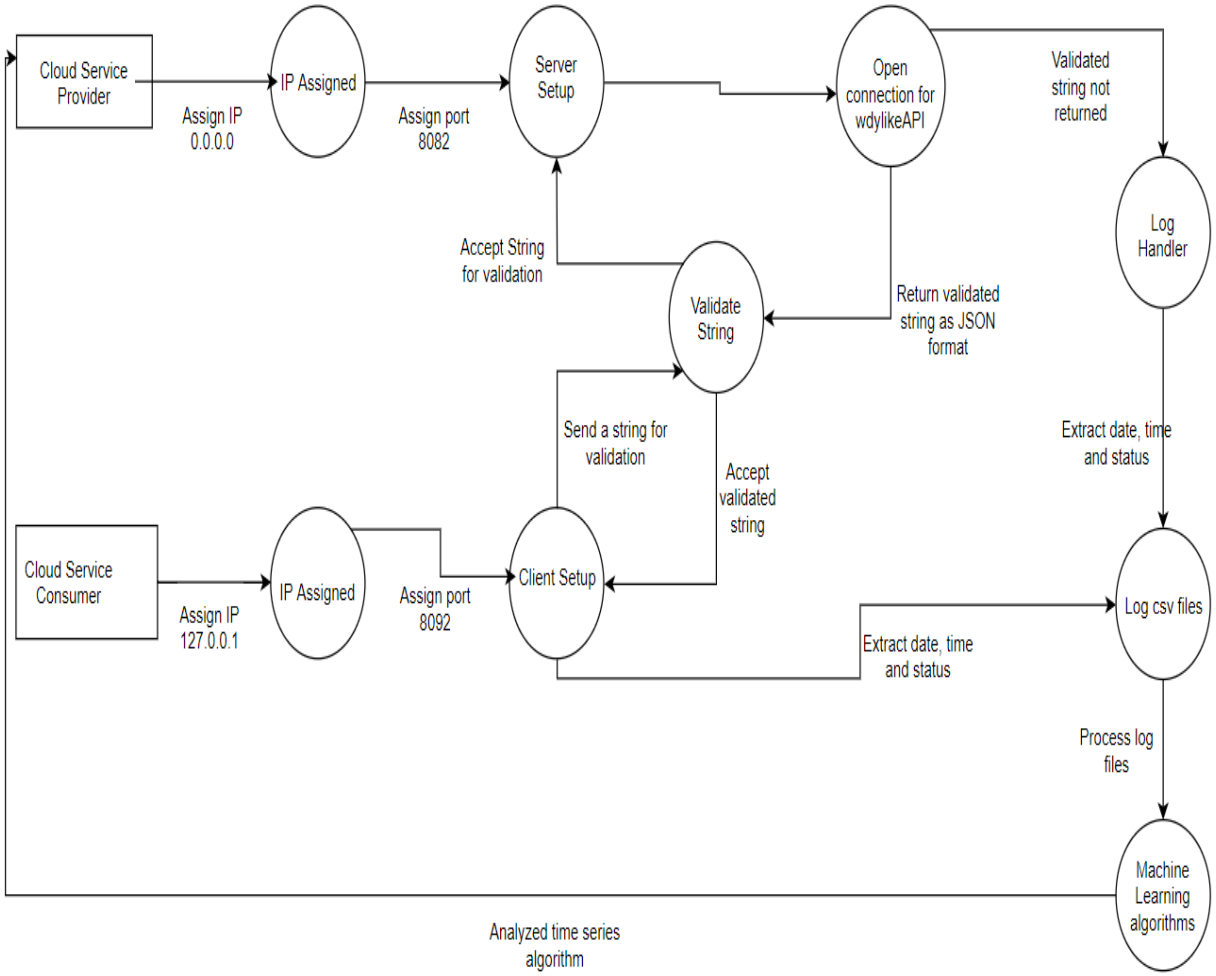


**Fig 3.2:Data Flow Diagram**

## 4. Task Model

We use flask technology to implement this project. In this project, we have 2 entities: - Cloud Service Provider and Cloud Service Consumer. Now we are assigning IP address and a port address to the Cloud Service Provider Flask Application to setup server to serve the request of the client. We also assigned IP address and port address to the Cloud Service Consumer Flask Application which sends a string for validation to the server by using Validate String Process. Now server connect to the WDYlike API. If that API return Validate String as JSON format, it should

send to Validate String process which further transfer the result to the Cloud Service Consumer, else, the string will be send to Log Handler process. According to the Validate String process, Uptime and Downtime status of the server is appended in the log.csv file and that csv file will be used as the dataset for the further analysis. On that dataset, we apply Pandas Data Analysis Library to predict the SLA Violation.

## Step 1: Data Collection

In this step, time series data is collected and is divided into two parts, namely input data and testing data. Input data constitutes 80% of the collected data and is used by the prediction algorithms to predict the future QoS of each input. The predicted values are then tested against the testing data to determine the accuracy of the predicted results.

## Step 2: Input selection

In this step, the inputs required from the input data to predict the future QoS are selected. This step needs to be applied to each of the time series prediction methods under consideration as their required inputs vary. For example, simple moving average, one of the most basic prediction methods, considers data from previous N time intervals, averages them and then uses the result to predict the future time interval. On the other hand, the extrapolation method unlike interpolation which considers previous data between two known data points, considers data beyond the range of known data points. Hence, data needs to be selected appropriately according to the specifics of the time series method considered to obtain an accurate prediction result.

## Step 3: Implementing analysis methods

In this step, the six different time series prediction approaches considered in this work are applied on the input data to predict the future QoS. MATLAB is used to apply the different algorithms in this step.

## Step 4: Comparing methods and result analysis

In this final step, root mean square error (RMSE) is used to examine the accuracy of the prediction methods on the inputs dataset in determining the future QoS

# 5. Code

In [1]:
```
import pandas as pd
df = pd.read_csv('/content/cloud_server.csv')
print(df)
```

```
     Date_Stamp Time_Stamp Server_Alpha_Status  Server_Numerical_Status
0    2019-12-13    1:10:14           server up                        1
1    2019-12-13    1:10:14           server up                        1
2    2019-12-13    1:10:15           server up                        1
3    2019-12-13    1:10:15           server up                        1
4    2019-12-13    1:10:16           server up                        1
..          ...        ...                 ...                      ...
728  2019-12-14   12:33:35           server up                        1
729  2019-12-14   12:33:36           server up                        1
730  2019-12-14   12:33:36           server up                        1
731  2019-12-14   12:33:37           server up                        1
732  2019-12-14   12:33:38           server up                        1

[733 rows x 4 columns]
```

In [5]:
```
import matplotlib.pyplot as plt
server_status = df.groupby('Server_Alpha_Status')['Server_Numerical_Status'].count()
print(server_status)
```

```
Server_Alpha_Status
server down    585
server up      148
Name: Server_Numerical_Status, dtype: int64
```

In [7]:
```
server_status_unique = list(df.Server_Alpha_Status.unique())
server_status = server_status.reindex(server_status_unique,axis=0)
print(server_status)
```

```
Server_Alpha_Status
server up      148
server down    585
Name: Server_Numerical_Status, dtype: int64
```

In [9]:
```
server_status = server_status.to_frame()
print(server_status.head(12))
server_status.reset_index(level=0,inplace=True)
```

```
                    Server_Numerical_Status
Server_Alpha_Status
server up                              148
server down                            585
```

In [15]:
```
plt.figure(figsize=(13,10))
plt.bar(server_status['Server_Alpha_Status'],server_status['Server_Numerical_Status'],color=(0.5,0.1,0.5,
0.6))
plt.suptitle('Cloud Up Time / Down Time Analysis',fontsize=10)
plt.title('Using Data from 13 DEC 2019 to 14 DEC 2019 (734 Rows | 4 Cols)',fontsize=20)
plt.xlabel('Server Status',fontsize=20)
plt.ylabel('Server Status Time In Seconds',fontsize=20)

for i,num in enumerate(server_status['Server_Numerical_Status']):
  plt.text(i,num,num,ha='center',fontsize=12)
```

## 6. Conclusion

The data is analyzed according to the terms and conditions mentioned in the SLA like Server uptime and downtime, the data set is compared according to the SLA, if there is any discrepancies found then the service provider will pay the penalty as per the mentioned violations in the SLA.

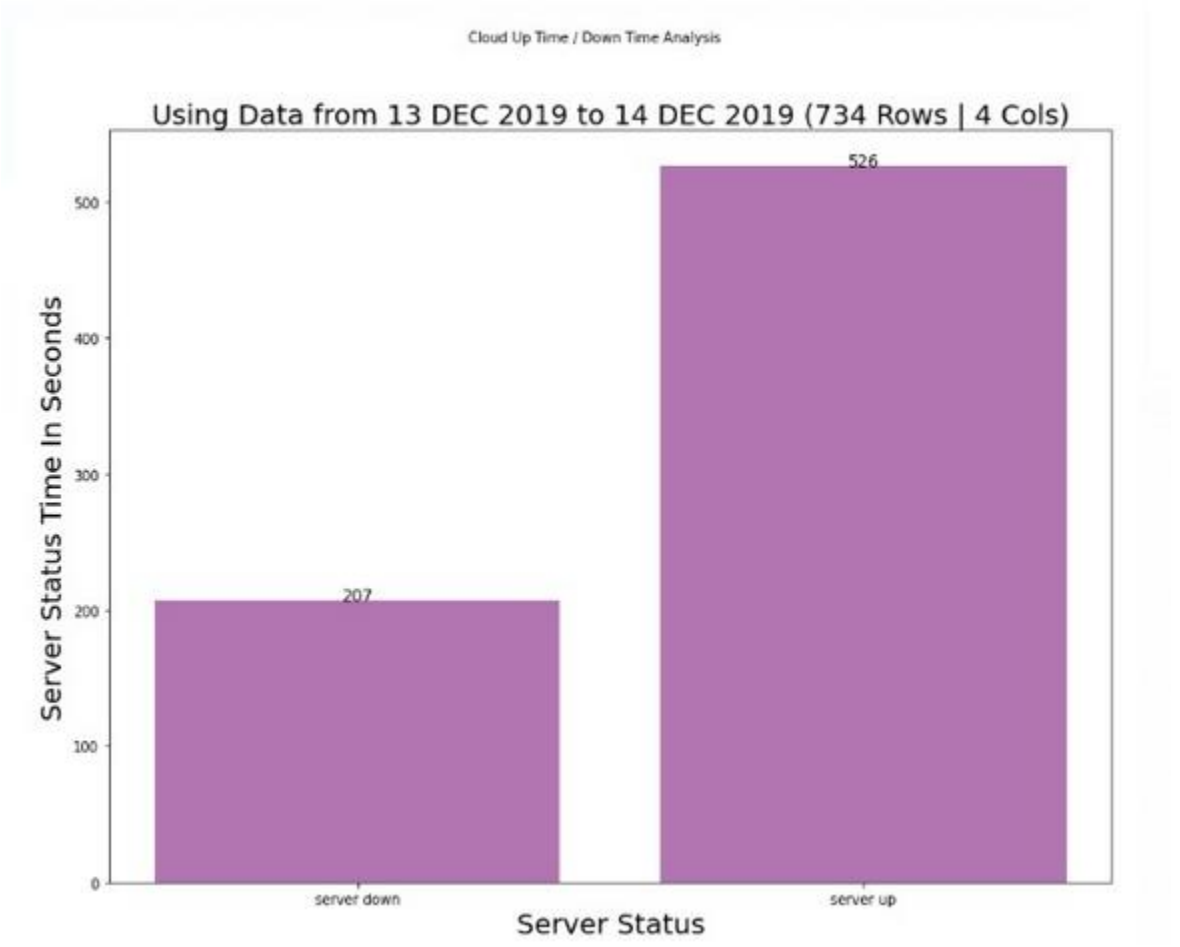


**Fig 6: Data Analysis Graph**

Following Data Analysis Graph is depicting the Server Availability and Server Unavailability for a given period of time (1 Day i.e. 13 Dec 2019 to 14 Dec 2019) using Pandas Library of Python and on that basis violation determination of SLA Violation can be made and Cloud Service Provider will have to pay specified penalties to its customer (end user) as per SLA Terms And Conditions.

## References

[1] M. Cunha, N. Mendonça, and A. Sampaio, "Cloud Crawler: a declarative performance evaluation environment for infrastructure-as-a-service clouds," Concurrency and Computation: Practice and Experience, vol. 29, no. 1, 2017.
[2] I. Gartner, "Forecast: Public Cloud Services, Worldwide, 2013-2019," ed: Gartner 2016.
[3] B. Narasimhan and R. Nichols, "State of cloud applications and platforms: The cloud adopters' view," Computer, no. 3, pp. 24-28, 2011.
[4] D. Greenwood, G. Vitaglione, L. Keller, and M. Calisti, "Service level agreement management with adaptive coordination," in International conference on Networking and Services, Slicon Valley, CA, USA, 2006, pp. 45-45: IEEE.
[5] B. R. Kandukuri, V. R. Paturi, and A. Rakshit, "Cloud security issues," in IEEE International Conference on Services Computing, Bangalore, India, 2009, pp. 517-520: IEEE.
[6] W. Hussain, F. K. Hussain, and O. K. Hussain, "Maintaining Trust in Cloud Computing through SLA Monitoring," in Neural Information Processing, 2014, pp. 690-697: Springer.
[7] J. Yan, R. Kowalczyk, J. Lin, M. B. Chhetri, S. K. Goh, and J. Zhang, "Autonomous service level agreement negotiation for service composition provision," Future Generation Computer Systems, vol. 23, no. 6, pp. 748-759, 2007.
[8] V. C. Emeakaroha, I. Brandic, M. Maurer, and S. Dustdar, "Low level metrics to high level SLAs-LoM2HiS framework: Bridging the gap between monitored metrics and SLA parameters in cloud environments," in International Conference on High Performance Computing and Simulation (HPCS), Caen, France, 2010, pp. 48-54: IEEE.