

# Intelligent Big Data Processing On Multiple Cloud Environment

*Ms P. Indhuja MCA., Assistant Professor Mangyarkarasi Arts & Science College  
Mr A.SenthilKumar MCA., B.Ed., M. Phil., Assistant Professor, Prist University*

## Abstract

*The main purpose of this project is big data can be securely store and process this data in multiple cloud. The next goal is to ensure an efficient selection of trustworthiness cloud provider who eventually will guarantee high QoCS and fulfils key Big Data workflow requirements. This model is Cloud Computing, and among its main features has to stress its elasticity in the use of computing resources and space, less effort, and flexible costs. In this Intelligent Multi-dimensional trust model the security of cloud services can be easily measured. It provides the trust between the cloud service provider and user. The Security Strength of the particular cloud services can be measured for processing the big data on cloud. The Big data processing which include the process of Data selection, loading, storing, viewing, data preprocessing, test data quality, feature extraction, data analysis and quality evaluation.*

## Keywords

*Big Data, Big Data processing, cloud computing, cloud selection, trust model, Quality of Cloud Services, service evaluation, community.*

## Introduction

An Intelligent multi-dimensional trust model for Big Data workflow processing over different clouds. It ensures the trustworthiness of cloud providers based on: the most up-to-date cloud resource capabilities, the reputation evidence measured by neighbouring users and a recorded personal history of experiences with the cloud provider.

Big Data requires processing, analysis and storage. Big Data is not only defined by size, it is also characterized by multi V's; volume, variety, velocity, veracity, validity, volatility. These special characteristics of Big Data introduce several challenges, such as data collection and integration problems, due to the data being distributed across diverse geographical locations. Moreover, the management, processing and storage of Big Data also present significant challenges considering the enormous volume and heterogeneous nature of the datasets, and traditional processing platforms are unable to efficiently handle such massively heterogeneous data volumes.

Selecting the best cloud provider among this competing pool of cloud providers to store and process Big data is a challenging process. It is difficult for service consumers to decide which cloud provider to use as he/she may lack knowledge about whether the available cloud resource capabilities can handle Big Data tasks while satisfying a set of QoCS (Quality of Cloud Service) requirements.

A multi-dimensional trust model to evaluate the services of cloud providers based on:

- 1) The client's QoCS requirements,
- 2) The provider's current resources availability,
- 3) The historical records of his/her previous communications with the cloud service providers
- 4) The community members trust score evaluation based on their own historical records of previous communication

with the cloud service providers.

A comprehensive and multi-dimensional trust evaluation model to address these challenges and guarantee proven QoCS. Our solution uses a weighted average of three cloud selection strategies: current cloud characteristics, reputation and supported historical communications. Additionally, our trust model supports multiple crucial functional and nonfunctional requirements that guarantee reliable trust evaluation.

## System Analysis

### Proposed System

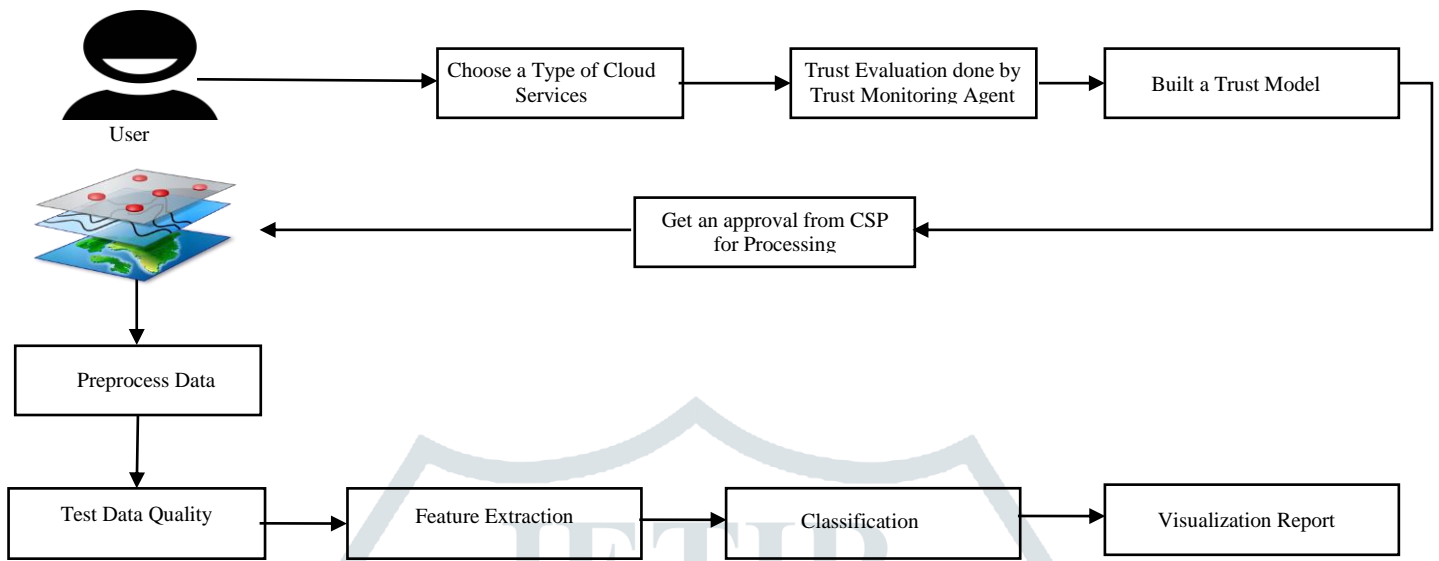
To overcome the security problems that are occurred in the existing system and effectively process the data over the cloud we introduce this system. It will provide better quality evaluation report based on the overall process of the big data.

### Advantages

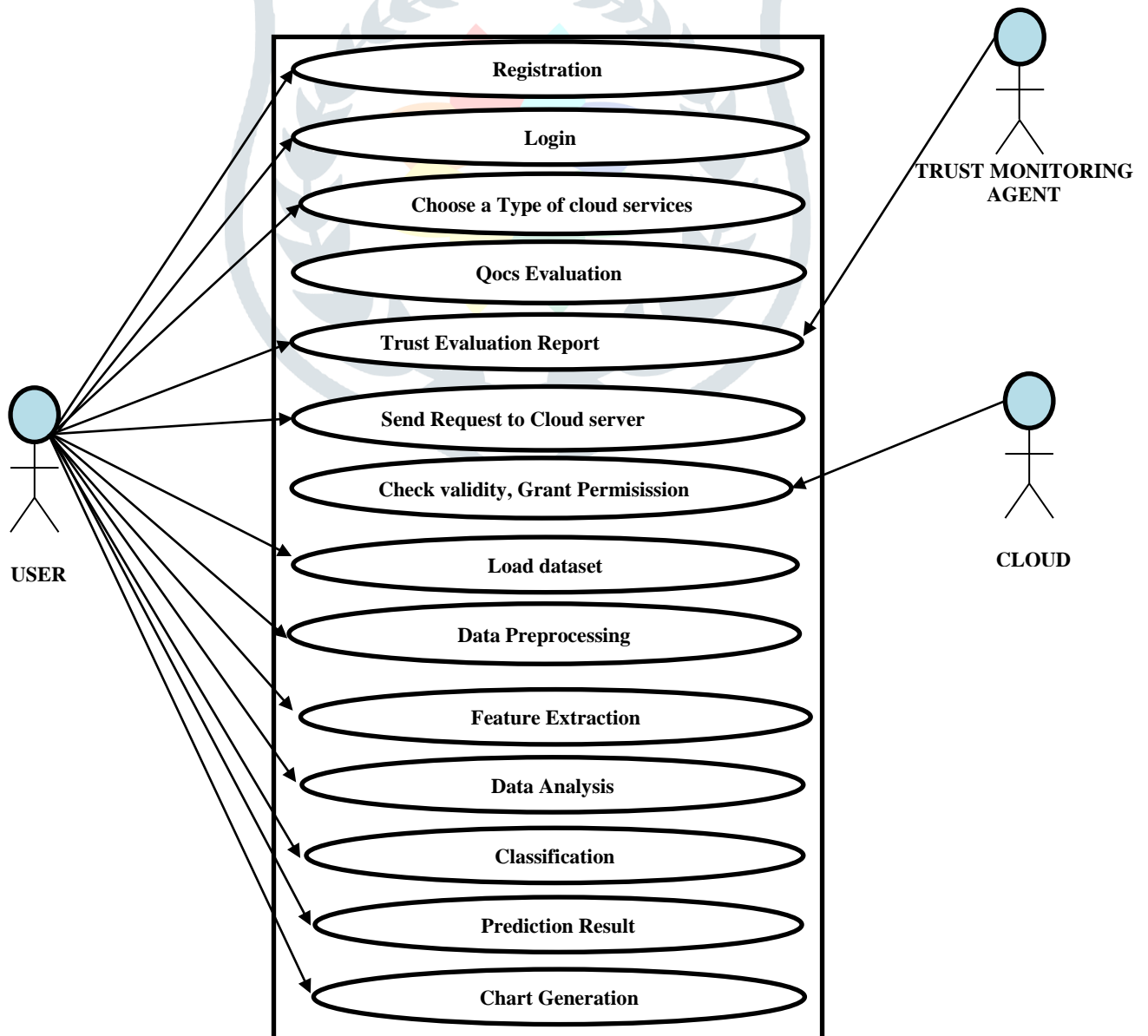
- Low Computational Cost.

- Increase the accuracy of the result.
- The data's are highly secured through the processing.

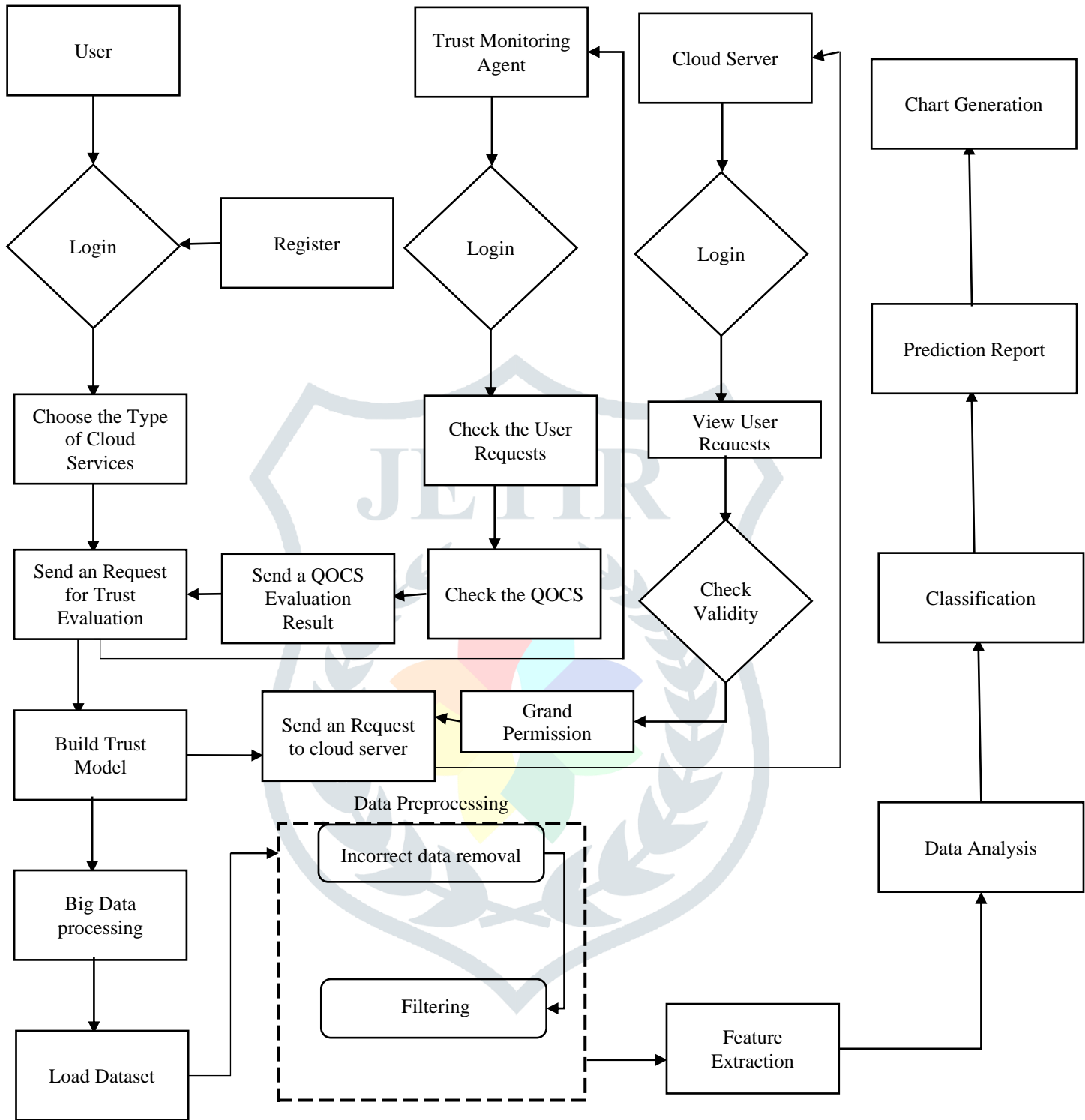
**Architecture Diagram**



**Use Case Diagram**



Flow Diagram



**Modules**

- Cloud and User Login
- Quality of Service Evaluation and Built a Trust Model
- Data Selection and Loading
- Data Preprocessing
- Test Data Quality
- Feature Extraction
- Data Analysis
- Data Visualization

**Cloud and User Login**

The user can login with their username and password if it is a new user, the user have to be registered before login. The user can choose the type of cloud services they want and send a request to the cloud service provider. The cloud service provider will login to their account, view the user request, built a trust model and send a response for processing the big data over competing cloud.

**Quality of Service Evaluation and Built a Trust Model**

It evaluates the trustworthiness of cloud providers. Building Trust model guarantee high QoCS and fulfils key Big Data workflow requirements

**Data Selection and Loading**

Data extraction techniques are part of the business intelligence domain and apply specific algorithms to discover hidden patterns and relationships in large data sets. In this module, the dataset is loaded for processing the big data.

**Data Pre-Processing**

Data preprocessing is the process of data cleaning and filtering activities to remove unwanted and noisy data. Initial step of this process is data pre-processing. The data is pre-processed to remove the unwanted data.

**Test Data Quality**

Assessment activities of a set of data quality attributes including data accuracy completeness and consistency.

**Feature Extraction**

Extract relevant features from the dataset for analyze. Feature extraction is the process of transforming the dataset into a set of features which can very well represent the input data.

**Data Analysis**

Some techniques are applied to dataset in order to extract the meaningful information. It support the decision making. The Naïve Bayes Classification algorithm is used to classify the normal and abnormal person.

**Data Visualization**

Some techniques are applied to dataset in order to extract the meaningful information. It support the decision making. Visualization task generates reports based on the overall process.

**Future Enhancement**

Data Classification based on Security: A cloud computing data centre can store data from various users. To provide the security level based on the importance of data, classification of data can be done. This classification scheme should be various aspects like access frequency, update frequency and access by various entities etc. based on the type of data. Once the data is classified and tagged, then security level associated with this specific tagged data element can be applied. Security level includes confidentiality, encryption, integrity and storage etc. that are selected based on the type of data.

**Conclusion**

This project ensure an efficient selection of trustworthiness cloud provider who eventually will guarantee high QoCS and fulfils key Big Data workflow requirements. It will provide better data analysis report.

**Screenshots**

## Trust Evaluation

[View User Details](#)

**Username**

Jessy

**Type of Cloud Services**

Public Cloud

[Trust Evaluation](#)

**Security Level**

[Send Evaluation Report to User](#)

-> Next



### Send an Request to Cloud Server

EEG.csv Choose the file

```
0 7 41 50 71 135000 356000 105000 9190 14400 110000 120000 22200 1 0
0 7 27 48 -56 1040000 484000 90700 32700 21200 25100 94000 18000 1 0
0 7 30 48 11 818000 116000 36700 31400 25100 98300 42500 11200 1 0
0 7 37 47 67 575000 351000 45900 24000 11300 128000 118000 24600 1 0
0 7 26 38 -69 88900 120000 8550 4870 2790 9930 7750 5230 1 0
0 7 50 41 28 459000 54300 32500 5490 10700 58100 32700 18600 1 0
```

**File Details**

File Details	File Format	File Size
EEG.csv	csv	66356.0 bytes

Send an Request to Cloud server

View Request Status In Progress

