International Journal of Advanced Technology and Engineering Exploration
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
Volume-1 Issue-1 December-2014

Auto Upload and Chi-Square Test on Application Software as a Service for Cloud Computing Environment

Sampada Kembhavi[1*] and Gajendra Singh[2]
M.Tech Scholar, CSE, SSSIST, Sehore, Bhopal, India[1]
HOD (CSE/IT), SSSIST, Sehore, Bhopal, India[2]
sampada.parulkar@gmail.com

Abstract

The cloud delivery model has been suggested as the panacea for complex as well as inflexible IT systems that can be ill afforded by small and large enterprises alike. However the lack of mature operational models for Application software as a Service (Asaas) meant that the cloud capabilities have not yet been developed and researched to a level that allows their exploitation to a full degree. In this paper we provide a survey for the cloud computing and suggest a novel framework where data is auto upload after a proper authentication and processing their task (gathering and sharing) in the cloud environment. Then we apply Chi-Square test, to test the hypothesis for correctness. Then we check the program capability under three parameters; first is F-measure (FM), second is odds ratio (OR) and third is power (PO). Based on the three parameters we can find the better application software as a service approach. Based on the above phenomena we can compare the cost and the time.

Keywords

Asaas, Chi-square test, FM, OR, PO.

1. Introduction

Cloud computing has emerged as one of the most promising and challenging technologies of our time. This new paradigm utilizes two separate technological development utility computing and service oriented architecture to provide the users (individuals, SMEs and enterprises) with highly scalable, pay-per-use, everything as a service model for IT delivery. Some of the properties that characterize the cloud computing service delivery model are scalability/elasticity, on-demand service provisioning, shared resource pooling, multi-tenancy hosting, utility pay-as-you-use pricing and abstraction of lower layers [1]. These characteristics give rise to several business drivers that make cloud computing an attractive service delivery model from a customer's point of view. They include capital expenditure reduction, increased IT agility, faster return on investment, resilient infrastructure leading up to better business continuity. From a deployment model point of view, a cloud computing taxonomy can be divided into the following types: Public clouds: where the IT capabilities that are offered by cloud providers to any customers over the internet. Private clouds: where IT capability is offered to a select group of consumers who are part of an enterprise. The cloud service provider may be an internal IT organization (i.e., the same organization as the consumer) or a third party.Hybrid clouds: in which the environment is created through the usage of a combination of private and public cloud offerings by an organization. Internal clouds: is a subset of the private cloud model, where the cloud is an IT capability offered as a service by an IT organization to its own business.

External clouds: is IT capability offered as a service to a business that is not hosted by its own IT organization. An external cloud can be public or private, but must be implemented by a third party. From a point of view of architectural service layers based on the services provided using the cloud model, the ecosystem can be broadly divided into three:

Software as a Service (SaaS): forms the top layer featuring a complete application provided in a multi-tenant environment. One prominent example of SaaS is Salesforce [2].

*Author for correspondence

Platform as a Service (PaaS): providing a development and deployment middleware layer. Key players include the Microsoft Azure platform [3] as well as Google App Engine [4]. Infrastructure as a Service (IaaS): the lowest layer delivering services like compute storage and network. One prominent example of IaaS is Amazon EC2 service [5]. The work reported here mainly deals with the IaaS service delivery model. In this paper we provide a framework for the betterment of (SaaS), by utilizing as an Application software as a Service (Asaas).

The remaining of this paper is organized as follows. In Section 2 we discuss about problem statement. Literature Survey in section 3.In section 4 we discuss about proposed framework. The conclusions are given in Section 4. Finally references are given.

2. Problem Statement

SaaS's strength comes from economies of scale as providers can consolidate the support, update, and server infrastructure for all the users of a specific service. However, along with the introduction of SaaS came the expressions: "Software on demand" and "Service on demand". These expressions implied that users subscribed for a service and expected to use this service the same day. This expectation allows rapid growth in demand and reduces administrative costs for the users, but can lead to a problem if the SaaS provider is not properly prepared to meet the growing demand. Inadequate server or support infrastructure by the provider can lead to a poor user experience. The poor user experience can be devastating for companies depending upon the service that they are expected to be provided. Poor management by either the provider or their customers can lead to organizational issues for both (here the customers are companies that have contracted for service with the SaaS provider). The trend that SaaS introduced was expanded upon by Amazon in 2007 when they launched their "Elastic Cloud Computing" service (here after referred to as Amazon EC). In this model customers can rent raw server capacity as "infrastructure as a service"(IaaS). Raw server capacity refers to the ability of the customer to specify the virtual machine images that are to be run when they are needed.

Amazon EC is based on virtualization and the ability to move virtual instances of a running VM to another host computer without interrupting (or only briefly interrupting) the operation of the VM. This facilitates consolidating hardware requirements for many different companies via resource pooling. This resource pooling provides increased resilience and greater efficiency, while lowering capital requirements. Today there are a large number of cloud providers that offer server capacity at a competitive price. Many SaaS providers see cloud computing as their solution to scalability, but there are many issues that SaaS providers have to consider before deciding if a cloud is the right platform for their infrastructure. Some aspects that must be considered are: reliability, lock in effects, legislative requirements, and data security.

3. Literature Survey

In 2010, Hai Zhonget al. [6] investigates the possibility to allocate the Virtual Machines (VMs) in a flexible way to permit the maximum usage of physical resources. They use an Improved Genetic Algorithm (IGA) for the automated scheduling policy. The IGA uses the shortest genes and introduces the idea of Dividend Policy in Economics to select an optimal or suboptimal allocation for the VMs requests.

In 2010, Xiaohui Wei et al. [7] suggest that Meta-scheduling is proven to be an essential tool for resource management in cross-domain environments like Grids. They take one step further in combining meta-scheduling with application management tool to present a novel multi-domain cloud environment for biological research. With the scalable architecture and stable performance, Community Scheduler Framework (CSF) is a mature meta-scheduler for traditional Grids. However, as a scheduler, CSF does not support flexible application management. They integrate CSF with another useful tool OPAL, which is a toolkit for wrapping scientific application as web services, to take advantage of both grid and cloud resources. They first talk about the integration of CSF4 with Opal to form a user-transparent and configurable scientific cloud platform. Next, some enhancements of CSF4 are presented to improve the usability of the OPAL-CSF system. In addition, they will describe a virtual screening web service in OPAL-CSF system. The virtual screening service compares multiple receptor/ligand protein pairs using Autodock. Performance analysis of this service shows that the integration of Opal and CSF can efficiently facilitate large-scale scientific applications.

In 2010, Chunye Gong et al. [8] propose the characteristics of this area which make cloud computing being cloud computing and distinguish it from other research areas. The cloud computing has its own conception, technical, economic and user experience characteristics. The service oriented, loose coupling, strong fault tolerant, business model and ease use are main characteristics of cloud computing. Clear insights into cloud computing will help the development and adoption of this evolving technology both for academe and Industry.

In 2010, Xin Bai [9] explore ways to maximize the power of mobile learning technologies to increase enrollment, reduce attrition, improve teaching efficiency, and enable green technology without sacrificing the quality and placing extra burden on faculty. As mobile learning technologies pose challenges to most educational practitioners, a generic mobile learning framework was developed to address this issue. Their hybrid application approach utilizes mobile devices' native functions and existing Web 2.0 applications. Additional applications to fit in the framework are created to allow non-technical instructors to develop mobile learning content to engage student learners in informal settings.

In 2010, Dheresh Soni et al. [10] o generalize the formulation of data mining techniques with cloud computing environment. In data mining we want to find useful patterns with different methodology. The main issue with data mining techniques is that the space required for the item set and there operations are very huge. They combine data mining techniques with cloud computing environment, and then we can rent the space from the cloud providers on demand. This solution can solve the problem of huge space and we can apply data mining techniques without taking any consideration of space. They survey and analyze the utility for solving the above situation. Performance measurement is suggested in [11].

In [12] combine base classifiers are used for ensemble fusion methods namely Decision Template, Dempster-Shafer and Bayes to compare the accuracy of the fusion methods on the brown cloud dataset. Their results for the Bayes fusion method performs better classification accuracy of 95% than Decision Template of 80%, Dempster-Shaferof 85%, in a Brown Cloud image dataset.

In 2012, M.Malathi et al. [13] suggest that the Cloud Computing is being adopted by many companies because of its capacity to use computing and storage resources on a metered basis thereby reducing the investments in infrastructure. With all its benefits, cloud computing also brings along concerns about the security, privacy and jurisdiction because of its size, structure, and geographical dispersion. They try to explore these concerns and gives suggestions which may help companies to take security initiatives before they actually move into the cloud.

In 2012, Astha Pareek et al. [14] discuss that the data mining is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. It is the extraction of information from huge volume of data or set through the use of various data mining techniques. The data mining techniques like clustering, classification, neural network, genetic algorithms help in finding the hidden and previously unknown information from the database. Cloud Computing is a web-based technology whereby the resources are provided as shared services. The large volume of business data can be stored in Cloud Data centres with low cost. Both Data Mining techniques and Cloud Computing helps the business organizations to achieve maximized profit and cut costs indifferent possible ways. The main aim of the work is to implement data mining technique in cloud computing using Google App Engine and Cloud SQL.

In 2012, Deepak Mishra et al. [15] observe that a cloud server connection consists of an occurrence of shared database architecture server and at least one front-end network server. When users request data from cloud server, the cloud application, running on the front-end network server, retrieves all the relevant data from backend to handle the manipulator request. In [16] cloud auditing services is also suggested.

In 2010, Zhengxiong Hou et al. [17] provide a web services portal, an on-demand software license service for the users. Application software is wrapped as web services on the basis of underlying computational resources. With a pay-for-use mode, there is no limitation for the licenses any more. The instant service rate, average job response time, and cost are analyzed for an evaluation. A case of implementation and the evaluation show that ASAAS can bring a much better effect than traditional mechanism.

4. Proposed framework

Our proposed framework is shown in figure 1. We provide the flexibility of uploading the data in two ways either by user or auto upload. We consider java files for uploading the data. Then data transaction including data collection and sharing will be provided in both of the model. Then we apply chi- square Test for cost computation. A chi-square ( ) test for goodness of fit is performed when the question is whether or not an observed pattern or a distribution of numbers is significantly different from an expected pattern or a distribution of numbers. An example of this idea could be the expected or claimed number of cars rented out per category (like small, medium size, large, SUV etc) versus the actual number of cars rented out per category. In our example we test for the cost calculation which provides a barrier for non-convincing results.

We will always have a null hypothesis which states that the observed distribution is not significantly different from the expected distribution and of course use words relevant to that particular problem.

The decision rule for this test will always be $\chi^2 < \chi_c{}^2$ where the critical value has to be read from the χ^2 distribution table. The only two numbers needed to look up this critical value are the level of significance α and the number of degrees of freedom. The degrees of freedom for this test will be defined as the number of categories minus 1. A partial χ^2 distribution table is presented below. This is how we find the critical value for a particular problem: suppose that we use $\alpha = 0.05$ and have 5 degrees of freedom (6 categories). We can find the critical value by looking at the point where the two arrows in the table meet.

The test statistic is $\chi^2 = \sum \frac{(E-O)^2}{E}$ where E and O are the expected and observed frequencies per category.

Step-by-Step Procedure for Testing Your Hypothesis and Calculating Chi-Square:

1. State the hypothesis being tested and the predicted results. Gather the data by conducting the proper experiment.
2. Determine the expected numbers for each observational class. Remember to use numbers, not percentages. Chi-square should not be calculated if the expected value in any category is less than 5.
3. Calculate p using the formula. Complete all calculations to three significant digits. Round off your answer.
4. Use the chi-square distribution table to determine significance of the value.
   - Determine degrees of freedom and locate the value in the appropriate column.
   - Locate the value closest to your calculated value on that degrees of freedom row.
   - Move up the column to determine the p value.
5. State your conclusion in terms of your hypothesis.
   - If the p value for the calculated is p > 0.5, accept your hypothesis. 'The deviation is small enough that chance alone accounts for it.
   - If the p value for the calculated is p < 0.5, reject your hypothesis, and conclude that some factor other than chance is operating for the deviation to be so great.

The chi-square test will be used to test for the "goodness to fit" between observed and expected data. Then we apply three parameters for finding the best among them. The parameters are F-Measure (FM), Odd Ratio OR and Power (PO). A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

FM= (2* Precision *Recall)/(Precision + Recall)

The odds ratio is a measure of effect size, describing the strength of association or non-independence between two binary data values. It is used as a descriptive statistic, and plays an important role in logistic regression.

OR= 2* Recall (1-Precision) /(1-Recall*Precision)

Power (PO) is defined as:

PO= ((1-Precision)^k-(1-Recall)^k)

Finally based on the above parameters we can proceed to the comparison which is based on cost and time. We are here to improve the effectiveness of application as a software service. By the above approach we can find the better result.

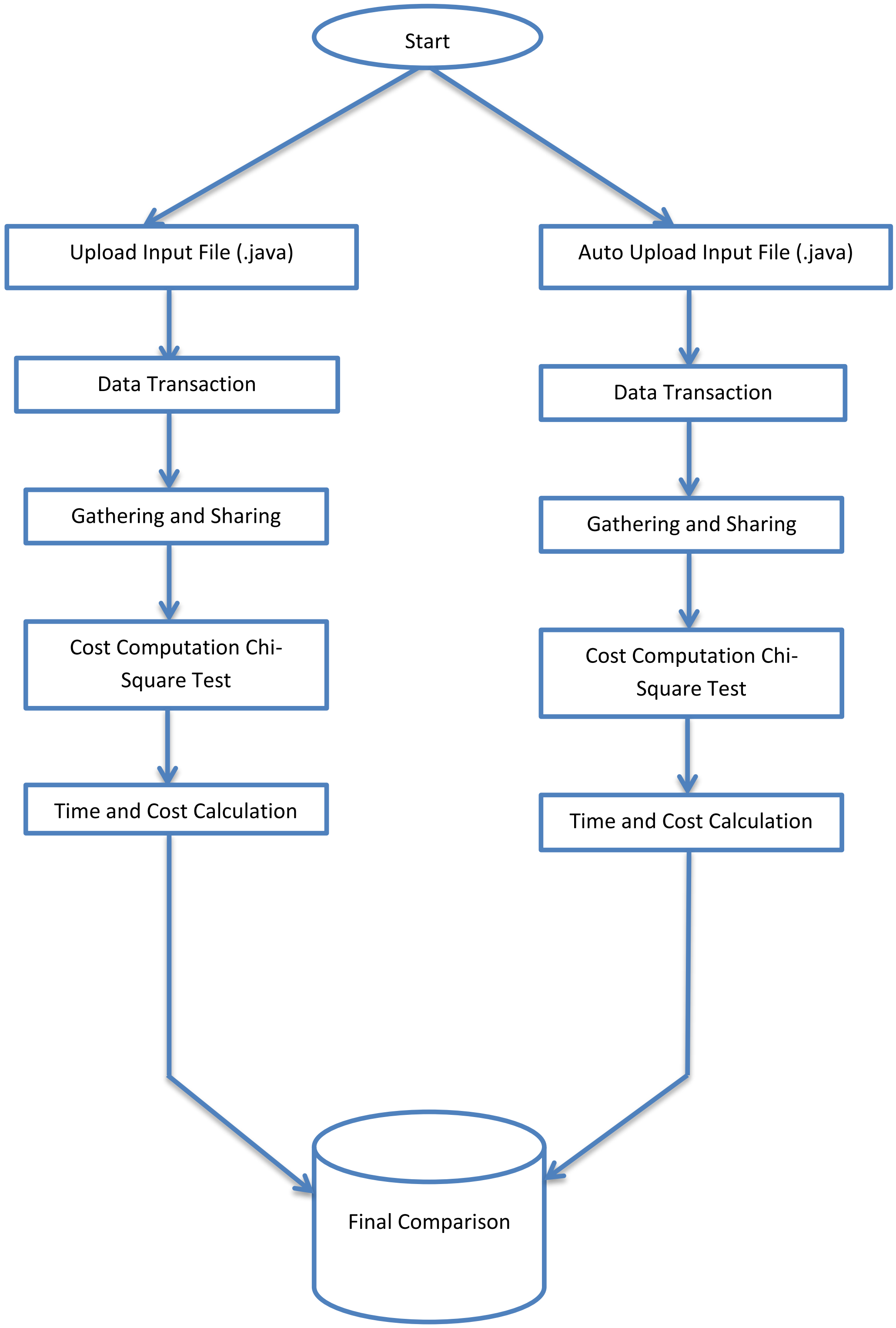

Figure 1: Proposed framework

5. Conclusion

Many of the activities loosely grouped together under cloud computing have already been happening and centralized computing activity is not a new phenomenon: Grid Computing was the last research-led centralized approach. However there are concerns that the mainstream adoption of cloud computing could cause many problems for users. So in this paper we find the ways for betterment as an application cloud service. We also provide testing of hypothesis based on that we can compare the test results. In future we apply the test hypothesis with net beans environment for testing the betterment.

References

[1] NIST Definition of Cloud Computing v15,http://csrc.nist.gov/groups/SNS/cloud-computing/clouddef-v15.doc.
[2] CRM - salesforce.com, http://www.salesforce.com/.
[3] Windows Azure Platform,http:///www.microsoft.com/windowsazure/.
[4] Google App Engine, http://code.google.com/appengine/.
[5] Amazon Elastic Compute Cloud(Amazon EC2),http://aws.amazon.com/ec2/.
[6] Hai Zhong, Kun Tao, Xuejie Zhang, "An Approach to Optimized Resource Scheduling Algorithm for Open-source Cloud Systems", The Fifth Annual China Grid Conference, 2010.
[7] Xiaohui Wei, Lixian Xi, Hongliang Li, Haibin Wang , "A Case of Meta-scheduling for Biological Researches in Cloud Environment", 2010 Fifth International Conference on Frontier of Computer Science and Technology.
[8] Chunye Gong, Jie Liu, Qiang Zhang, Haitao Chen and Zhenghu Gong, "The Characteristics of Cloud Computing", 2010 39th International Conference on Parallel Processing Workshops.
[9] Xin Bai, "Affordance of Ubiquitous Learning through Cloud Computing", 2010 Fifth International Conference on Frontier of Computer Science and Technology.
[10] Dheresh Soni ,Atish Mishra,Satyendra Singh Thakur , Mr. Nishant Chaurasia, "Applying Frequent Pattern Mining in Cloud Computing Environment", International Journal of Advanced Computer Research (IJACR),Volume 1, Number 2, December 2011.
[11] Pritam Fulsoundar, Rajesh Ingle, "Prediction of Performance Degradation in Cloud Computing", International Journal of Advanced Computer Research (IJACR), Volume-3, Issue-13, December-2013, pp.126-129.
[12] M.Krishnaveni, P.Subashini, A.Vanitha, "A Bayes fusion method based ensemble classification approach for Brown cloud application", International Journal of Advanced Computer Research (IJACR), Volume-4, Issue-14, March-2014, pp.299-304.
[13] M.Malathi, "Cloud Computing Issues-A Survey", International Journal of Advanced Computer Research (IJACR), Volume 2, Number 2, June 2012.
[14] Astha Pareek, Dr.Manish Gupta, "Review of Data Mining Techniques in Cloud Computing Database", International Journal of Advanced Computer Research (IJACR) Volume 2 Number 2 June 2012.
[15] Deepak Mishra, Dr. Manish Shrivastava, "Economic Price Estimation for cloud business", International Journal of Advanced Computer Research (IJACR) Volume-2 Number-4 Issue-6 December-2012.
[16] Igor Ruiz-Agundez, Yoseba K. Penya, Pablo G. Bringas," Cloud Computing Services Accounting", International Journal of Advanced Computer Research (IJACR), Volume-2, Issue-4, June-2012 ,pp.7-17.
[17] Zhengxiong Hou, Xingshe Zhou, Jianhua Gu,Yunlan Wang, Tianhai Zhao, "ASAAS: Application Software as a Service for High Performance Cloud Computing", 2010 12th IEEE International Conference on High Performance Computing and Communications.