International Journal of Advanced Computer Research (ISSN (print):2249-7277 ISSN (online):2277-7970)
Volume-3 Number-2 Issue-10 June-2013

An Efficient Algorithm for Auto Upload and Chi-Square Test on Application Software

Sampada Kembhavi[1], Ravindra Gupta[2], Gajendra Singh[3]

Abstract

In this paper we proposed a novel algorithm for auto uploading the data in the cloud environment without any license consideration 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. Here we present the authentication process with auto uploading process. We also use the traditional method of data uploading, data gathering and sharing. So that we compare our statistics. We use object oriented programing (OOP) programs as a data which is used for data sharing and gathering. We use java programs for comparison. 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 execution time in different operating application with the help of software quality metrics.

Keywords

OOP, Chi-square test, FM, OR, PO

1. Introduction

The advantages of cloud computing over traditional computing include: agility, lower entry cost, device independency, location independency, and scalability. In order to solve the problem of data integrity checking, many schemes are proposed under different systems and security models [1], [2], [3], [4], [5]. In all these works, great efforts are made to design solutions that meet various requirements: high scheme efficiency, stateless verification, unbounded use of queries and retrievability of data, etc.

The fundamental principle is that the applications users need will be run on large-scale server cluster which is use as the internal points which is affected by the data source which is used for data transfer point.

Sampada Kembhavi, M.Tech Scholar, Computer Science SSSIST, Sehore, India.
Ravindra Gupta, Assistant Professor, Computer Science SSSIST, Sehore, India.
Gajendra Singh, HOD, Computer Science SSSIST, Sehore, India.

The data users process will be stored in the data center rather than in the local machine. Thus, users can access to any terminal unit on the computer at any time and in any place. Applications built on Cloud Architectures run in-the-cloud where the physical location of the infrastructure is determined by the provider [6]. We consider the advantage of simple APIs of computer accessible services that scale on demand, that are industrial-strength, where the complex reliability and scalability logic of the underlying services remains implemented and hidden inside-the-cloud which is application as a service [6]. Cloud computing is broken down into three segments: "application", "storage" and connectivity." Each segment serves a different purpose and offers different products for businesses and individuals around the world [7]. Cloud computing is an emerging technology that promises to change the paradigm of computer services [8]. The main usage of this type of resources in Cloud Architectures is as needed, sometimes ephemeral or seasonal, thereby providing the highest utilization and optimum for the provider.

The remaining of this paper is organized as follows. Literature Survey in section 2.In section 3 we discuss about proposed framework. Result analysis is discussed in section 4. The conclusions are given in Section 5. Finally references are given.

2. Literature Survey

In 2010, Hai Zhong et al. [9] investigate 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. Their simulation experiments indicate that dynamic scheduling policy performs much better than that of the Eucalyptus, Open Nebula, Nimbus IaaS cloud, etc.

In 2010, Shital Pati et al. [10] suggest that Cloud Computing has changed the way we think about technology. It is a computing model that provides web based software, middleware and other computing resources on demand by deploying technology as a service. Eucalyptus is one of the open source software computing model that

provides Infrastructure as a Service cloud. To increase throughput and gain maximum profit in cloud systems task scheduling plays very important role.

In 2012, Bharat Prajapat et al. [11] suggest that the Advancement in mobile computing and application of the J2ME applications are become out dated. J2ME mobiles are low cost, energy efficient and with long battery life. But these mobile having less processing capabilities then newly designed mobile devices. Thus required to enhance the processing capability of these mobile thus required a light weight cloud web service based application strategy for executing the large process. They design and develop rich document exploration using J2ME mobiles and Cloud Web Service (CWS).

In 2012, M.Malathi et al. [12] 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. [13] 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 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. [14] 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. The cloud server uses caches built information retrieval and caches run on the front-end network servers of your cloud server farm. Every cache upholds reproductions of data nearby on the front-end network server so that cloud manipulator supplies can be checked by using the cached data and wherever applicable, circumventing unnecessary circulation amid backend server and the front-end network servers and CPU data management on the front-end network server, thus refining performance and scalability.

In 2010, Zhengxiong Hou et al. [15] 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.

In 2012, Abdur Rahim Choudhary et al. [16] provide a closer look at the cloud computing services. First they provide and establish a baseline by specifying high level requirements for cloud computing services. Then they suggest improves upon the current architecture for the cloud computing services by adding new modules to the current architecture. The new modules by the author are gleaned from an analysis of the telecommunications cloud and security in distributed systems. The new modules include a management and control network, a set of trust domains, and a set of proxies. The author architecture is more ready for primetime use and supports a richer operational model.

3. Proposed Framework

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.

In this section we also present the algorithm of our proposed work. First we provide the algorithm for registration of users. If the user register in the cloud environment, then when the register user want to upload the data in the cloud it is auto uploaded and dynamically gather with sharing properties as per the user requirement or request. In our proposed approach we consider two type of entry for data uploading in the cloud environment. First is for non-registered users which can get the space according to the

requirement of the normal user. Normal user can upload their data, update it. Then we provide a framework for registered users which are already registered and there is no need of space while uploading. They have the privilege to auto upload their data, which save the time. Registered users having the privilege of uploading, updating and sharing the resources from the neighbour. Then we also provide the way of checking the quality of uploaded file. Here we apply the chi square test for the gathered data. After that by applying F Measure (FM), Power (PO) and odd ratio from the software construing quality management, we can check the quality constraints. We present the algorithm for registration of new user so that user can save time of licensing as we can observe in the previous cloud computing technique.

Algorithm 1

- Select admin\normal user
- Check the Authentication for normal user
- If(uid==userid&&pwd==password) [This is created by the admin of the cloud]

{Enter in the cloud environment}

- Else Exit
- If(admin)
- Admin (pwd)
- {Request(database);

CreateUser();}

- Request (database)

[Send request to the user]

- CreateUser()
- CreateUser(){

[Enter Your UID, Username, Mobile Number etc.]
If(Success)
{
Print("Registration Successful");
Print(your password is=);}
Else
[enter Missing Parameters]}

Algorithm 2

- Add(data)

{
FileWriter fstream = **new** FileWriter("out.txt",true);
BufferedWriter out = **new** BufferedWriter(fstream);
out.write("string");
}

- Delete(data)

{

- File f1 = new File(file);

and delete the file using delete function f1.delete();
}

- Share(data)

{

- File folder = new File("C:\\userfolder");
- File[] files = cloud.listFiles();
- for (int i = 0; i < files.length; i++) {
- list.add(i, files[i].toString());

}

When use uploaded the object oriented module on the cloud environment, he/she can dynamically set the sharing and gathering property of the data in the cloud. The pseudo code for sharing and gathering is shown above.

Then we create the hypothesis for testing for the execution speed. We can test our hypothesis by Chi-Square test. We will always have a null hypothesis which states that the observed distribution is not significantly different from the expected.

Algorithm 3: Chi-Square Test

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 0.05.
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 > 0.05, accept your hypothesis. 'The deviation is small enough that chance alone accounts for it.

If the p value for the calculated is p < 0.05, reject your hypothesis, and conclude that some factor other than chance is operating for the deviation to be so great.

The test statistic is $\chi^2 = \sum \frac{(E-O)^2}{E}$ where E and O are the expected and observed frequencies per category. The chi-square test will be used to test for the "goodness to fit" between observed and expected data.

4. Result Analysis

In this section we provide the basis of result analysis. Figure 1 shows the sample java files which is processed for cloud computing by any

registered user. The properties are also set dynamically.

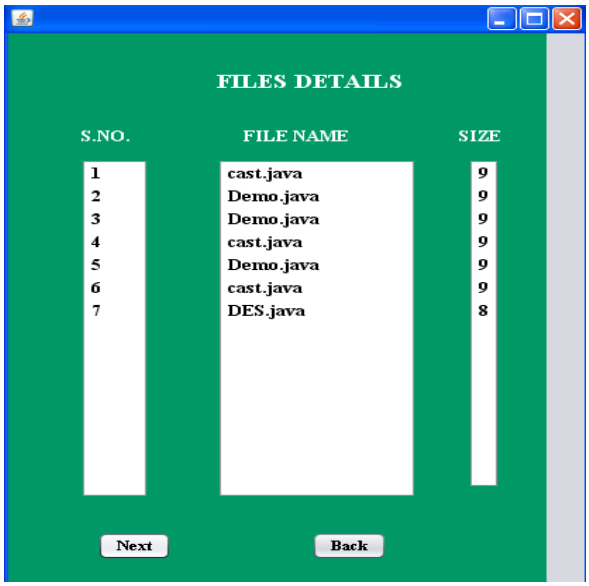

Figure 1: Java Files

We apply the Chi Square test for testing the hypothesis of execution time. The phenomenon is shown in figure 2. In this we first set the expected value which is constant for every program. Then we calculate the observed and the deviation. Based on the deviation we finally determine the p value. This is the main consideration point.

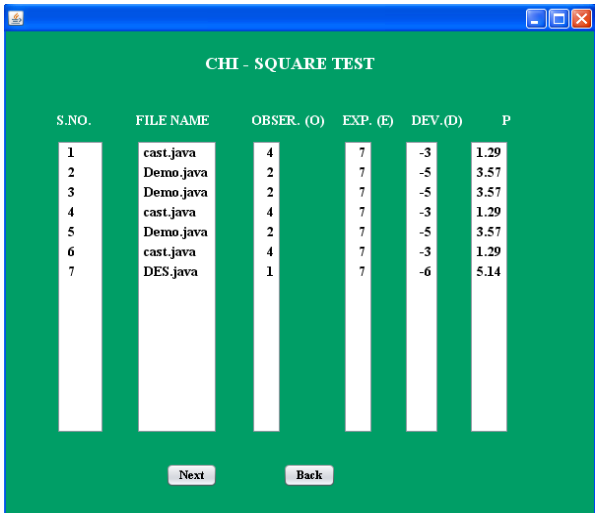

Figure 2: Chi Square Test

If the final p value is qualified then we calculate software parameters. We apply three parameters for finding the best among them. The parameters are F-Measure (FM)[Table 1], Odd Ratio OR[Table 2] and Power (PO)[Table 3].

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= $((1\text{-Precision})^k\text{-}(1\text{-Recall})^k)$

Table 1: FM calculation

sno	filename	class	object	inheritance	dma
1	cast.java	0.15	0.82	0.27	0.82
2	Demo.java	0.42	0.42	0.42	0.42
3	Demo.java	0.42	0.42	0.42	0.42
4	cast.java	0.15	0.82	0.27	0.82
5	Demo.java	0.42	0.42	0.42	0.42
6	cast.java	0.15	0.82	0.27	0.82
7	DES.java	0.6	0.07	0.82	0.07

Table 2: OR calculation

sno	filename	class	object	inheritance	dma
1	cast.java	0.15	0.82	0.27	0.82
2	Demo.java	0.42	0.42	0.42	0.42
3	Demo.java	0.42	0.42	0.42	0.42
4	cast.java	0.15	0.82	0.27	0.82
5	Demo.java	0.42	0.42	0.42	0.42
6	cast.java	0.15	0.82	0.27	0.82
7	DES.java	0.6	0.07	0.82	0.07

Table 3: PO Calculation

sno	filename	class	object	inheritance	dma
1	cast.java	-0.09	-0.09	-0.14	-0.09
2	Demo.java	-0.17	-0.17	-0.17	-0.17
3	Demo.java	-0.17	-0.17	-0.17	-0.17
4	cast.java	-0.09	-0.09	-0.14	-0.09
5	Demo.java	-0.17	-0.17	-0.17	-0.17
6	cast.java	-0.09	-0.09	-0.14	-0.09
7	DES.java	-0.17	-0.05	-0.09	-0.05

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. Graphical representation of FM and PO is shown in Figure 3 and Figure 4.

Figure 3: FM Representation

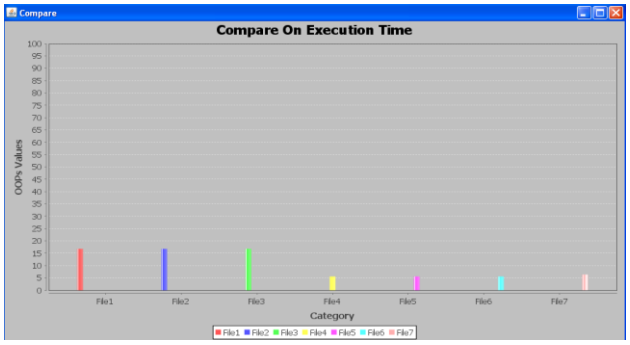

Figure 4: PO Representation

5. Conclusion

In this paper we present an efficient algorithm which is used for auto uploading the object oriented module by dymacially setting the properties of data sharing and gathering. This is the way 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] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable Data Possession at Untrusted Stores," Proc. 14th ACM Conf. Computer and Comm. Security (CCS '07), pp. 598-609, 2007.

[2] A. Juels and B.S. Kaliski Jr., "Pors: Proofs of Retrievability for Large Files," Proc. 14th ACM Conf. Computer and Comm. Security (CCS '07), pp. 584-597, 2007.

[3] H. Shacham and B. Waters, "Compact Proofs of Retrievability," Proc. 14th Int'l Conf. Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT '08), pp. 90-107, 2008.

[4] K.D. Bowers, A. Juels, and A. Oprea, "Proofs of Retrievability: Theory and Implementation," Report 2008/175, Cryptology ePrint Archive, 2008.

[5] M. Naor and G.N. Rothblum, "The Complexity of Online Memory Checking," Proc. 46th Ann. IEEE Symp. Foundations of Computer Science (FOCS '05), pp. 573-584, 2005.

[6] Ashutosh Kumar Dubey, Animesh Kumar Dubey, Mayank Namdev, Shiv Shakti Shrivastava," Cloud-User Security Based on RSA and MD5 Algorithm for Resource Attestation and Sharing in Java Environment", CONSEG-2012.

[7] Ajey Singh,Maneesh Shrivastava, "Overview of Security issues in Cloud Computing", International Journal of Advanced Computer Research (IJACR) Volume 2,Number 1,March 2012.

[8] Igor Ruiz-Agundez, Yoseba K. Penya and Pablo G. Bringas, "Cloud Computing Services Accounting", International Journal of Advanced Computer Research (IJACR) ,Volume 2, Number 2, June 2012.

[9] 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.

[10] Shital Patil ,R. A. Kulkarni, "Improving Performance Guarantees in Cloud Computing through Efficient and Dynamic Task Scheduling", International Journal of Advanced Computer Research (IJACR),Volume-2, Number-4, Issue-6, December-2012.

[11] Bharat Prajapat and Dr.Manish Shrivastava, "Mobile Cloud Computing through J2ME application: Cloud Enabled Web Services", International Journal of Advanced Computer Research (IJACR), Volume-2, Number-4, Issue-6, December-2012.

[12] M.Malathi, "Cloud Computing Issues-A Survey", International Journal of Advanced Computer Research (IJACR) Volume 2, Number 2, June 2012.

[13] 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.

[14] 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.

[15] 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.

[16] Abdur Rahim Choudhary," Baseline Requirements and Architecture for Cloud Computing Services", International Journal of Advanced Computer Research (IJACR), Volume-2, Number-4, Issue-7, December-2012.