

A Trust Aware Product Recommending Scheme for Multiple Cloud using HADOOP Services

A. Owaise Ahmed, K.Abirami, S.Chandrika, M.Janani

Computer Science and Engineering, Anna University, SKP Engineering College, Tiruvannamalai, Tamil Nadu, India

ABSTRACT

Service recommender systems have been shown as irreplaceable tools for yielding worthy recommendations to client. In the recent years, the range of client, services and online information exchange has grown rapidly, producing the big data analysis issue for service recommender systems. Accordingly, the conventional recommender systems frequently suffer from scalability and problems related to efficiency most of existing recommender systems presents the same grades and rankings to various users without considering multiple users' preferences, which fails to meet users' individualize requirements. In this work, to mention the above challenges and presenting a personalized recommendation list for products and recommending the most relevant products to the users effectively. Particularly, keywords are used to point out users' preferences, and hadoop framework is used for storing and processing the data of the client and will generate appropriate recommendations.

Keywords : T-broker, cloud, keyword search, content based search, ranking using hadoop.

I. INTRODUCTION

Analyzing the historical facts, we have always been using the user-experience in determining the usability of any product, be it book or hardware. The recommendation methods have evolved through millennia and now we are using the algorithms that finds the frequency and patterns in the information mining. The user who uses a system may have different tastes. One user's preference may be same as on other user's interests in some cases. The opinions and recommendation from other system users may provide a prediction which can be useful for the user to decide on what product he wants to buy. It aims at finding the keywords from which the ranking is done.

Based on the feedback from end-users the keywords are extracted. To gain specific knowledge in particular product, the users can make use of this environment to gain information and knowledge in particular product. The recommendation is found by the rankings or likes and the search similarity in the textual keywords. By using the keyword-based search and content-based search the relevant information is filtered and can be used for overcoming the system difficulty. The

predictions of the users are found with this and the recommendations are given to the customers on using the preferences.

Hadoop service providers offer users efficient and scalable data storage service with a very lower marginal than conventional approaches. It is a routine for users to leverage Hadoop storage services to share the data with others in a group, as data sharing becomes a standard feature in most hadoop storage offerings, including Dropbox, iHadoop and Google drive. The integrity of data in the Hadoop storage, however, is subjected to doubt and inspection, as data stored in the Hadoop storage can easily be lost or depraved due to the impossible hardware/software failures and human errors. To make this matter even worse, Hadoop service providers may be opposing to inform users about these data error to maintain the name of their services and avoid the losing profits. Therefore, the integrity of Hadoop data should be checked before utilizing the data, such as searching over Hadoop data. The conventional approach for checking data correctness is for retrieving the entire data from the Hadoop, and then verifies data integrity by checking the correctness of signatures. Surely, this conventional approach is able to

successfully check the correctness of Hadoop data. But, the efficiency of using this conventional approach on Hadoop data is in doubt.

In this paper we focus on product recommendation for that the subscribed user will only give reviews on the product based on the reviews given by the customers the ranking for the particular product is made.

II. METHODS AND MATERIAL

A. Existing System

Hadoop service provider gives users efficient and resizable data storage services with a very lower edge cost than conventional approaches. It is regular for the users to leverage Hadoop storage services for sharing the information with others in a group. Public verification was proposed to efficiently perform checking without downloading the entire data from the Hadoop. Here, data is splitted into many small blocks, where every block is signed independently by owner ;

A public verifier could be a data user who would like to utilize the owner's data via the Hadoop or a third-party auditor who can provide expert integrity checking services. So that during public verification on Hadoop data, the substance of private data of a personal user is uncovered to any public verifiers. Unfortunately, current public verification solutions mentioned above only reflects on personal data in the Hadoop. Existing public verification mechanisms can actually be extended to verify shared data integrity. However, a recent significant privacy issue has been introduced in the case of shared data with the use of existing approach is the leakage of individual privacy to public verifiers.

Disadvantages

- ✓ Lack of rigorous performance analysis for constructed audit system greatly affects the practical application of this scheme.
- ✓ It is essential to develop a more efficient and secure approach for existing public verification mechanisms.
- ✓ In order to protect the confidential information, it is essential and critical to preserve identity privacy from public verifiers during public verification.

B. PROPOSED SYSTEM

In the previous systems data mining and cloud is used for recommendation system for that no mathematical model is presented and no verification is done. Some

results have been given to establish higher correctness of the model, but fails in examination. In the proposed system, Hadoop framework is used for processing the reviews. Here cloud act as a middleware between the Hadoop framework and the ranking application. The cloud is used for providing secure access of the data by the users. The file about the product which is uploaded by the customers can be viewed in the cloud by the authorized users. The Hadoop framework is used ranking the products based on the reviews given by the customers. For selecting the Hadoop service providers, data mining techniques and aggregation methodologies may apply for combining the trustworthiness and abilities in order to estimate risk of interaction and compute the Trustworthiness through direct interactions or from feedbacks related to status of merchant. The application if developed as web services, then many applications can make use of the records. The data integrity in cloud environment is not considered. The error situation can be recovered if there is any mismatch in the data. The web site and the database is hosted in real cloud place during the implementation. Use of cloud for processing result in cost maximization hence using hadoop will reduce time and cost. It provides a relationship among recognized interaction risk, trustworthiness and competence of the recommended product. This suggests a mechanism by which trustworthiness of a product may be estimated.

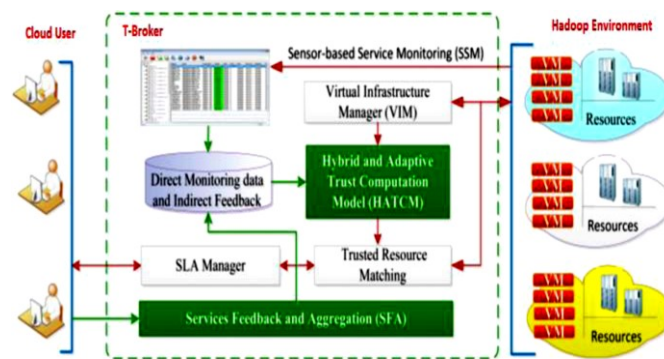


Figure1. Audit System Architecture

Processing Steps:

Step 1:

Initially the admin has to generate the product. That can be viewed by the customers. The product request from the customers can also be trusted or rejected by the admin.

Step 2:

The cloud user has to sign up in both cloud and the application. A security code is generated in cloud for every user. That security code is entered in the application. So that the user can sign up successfully. This security code is for authenticating the cloud user for accessing the cloud resources.

Step 3:

The customer has to give review for the product he/she have bought. While giving review the customer has to upload a file about the product. The file is stored in cloud and can be viewed by other cloud user. The file uploaded in the cloud is captured by hadoop and it divides the file into nodes. Each node is processed and the ranking is given. The file uploaded is sensed that whether it is word document, pdf or text document. This done by SSM. According to the type of file uploaded in the cloud. The processing is done.

Step 4:

The ranking is generated based on the review given by customers for each products. These ranking can be viewed by the users. The review can be given only by the customers who have purchased the particular product.

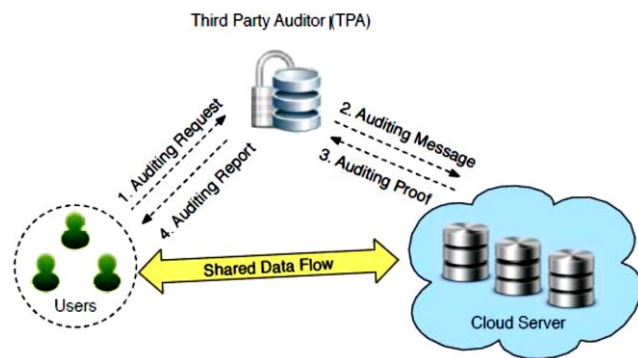


Figure 2. Architecture Base

C. MODULE

MODULE DESCRIPTION

There are four modules to implement the product recommendation in our system. The modules are as follows:

- 1. Hadoop User Module
- 2. Hadoop Resources Module (Admin)
- 3. T-Broker Module
- 4. Ranking

1. Hadoop User Module

In the first module, Hadoop users can send request to the T-broker for accessing the Hadoop resources .The feedback system collects locally-generated users’ reviews and aggregates these ratings to give the global evaluation ranking. After completing a transaction, the user will provide his or her rating as a reference for other users in future transactions .In this module, the user will register their personal details into the database. Each user has unique id, username and password. If the users enter the wrong password it doesn’t allow the user to access the Hadoop. If the new user enters then sign up with their details, for every unique user a verification code is generated.In this three types of user are registered. They are Admin, user and public verifier. The mode of operation of the Admin is they will upload the data and will provides access permission to all other users. The other user will login with their corresponding username and password. After using these details he can request for the specific file from server. The data will be converted to the encoded format and it will be stored back to the database. The use of public verifier is to audit the shared data. The public auditor doesn’t know the details about the data. The auditor audits the data once in a day. Here the log file is generated for the future references.

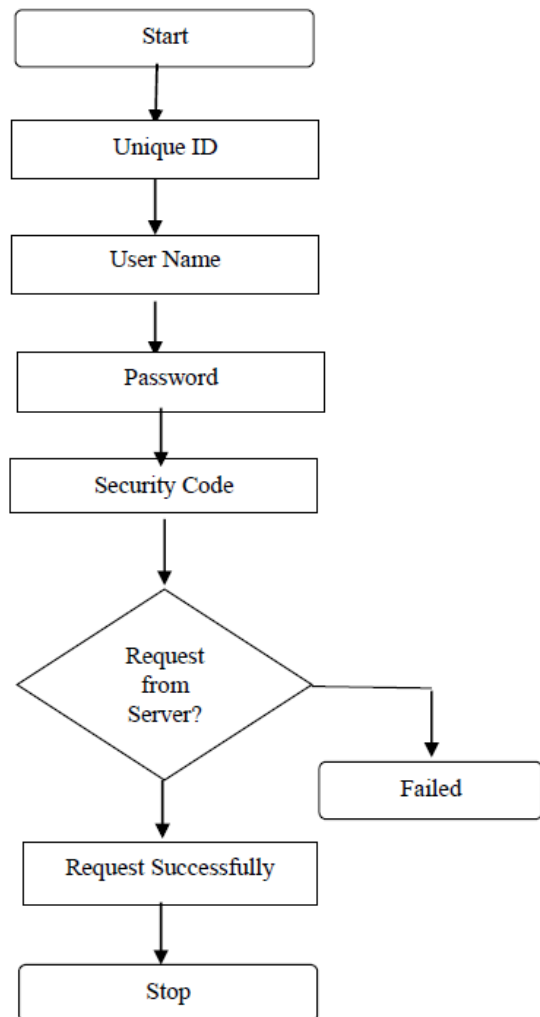


Figure 3. Flow Chart - Dynamic User Registration

User will download data later without any loss in data. The user can give the rating and the feedback about the Hadoop service provider once the user is added in the campaign. The user can view all the rating about the particular Hadoop service provider in the campaign, for the rating purpose separate password and username will be generated and by using that user can give the rating in the registration form. It also contains the queries about the Hadoop service provider and the user further requirements.

2. Hadoop Resource Module

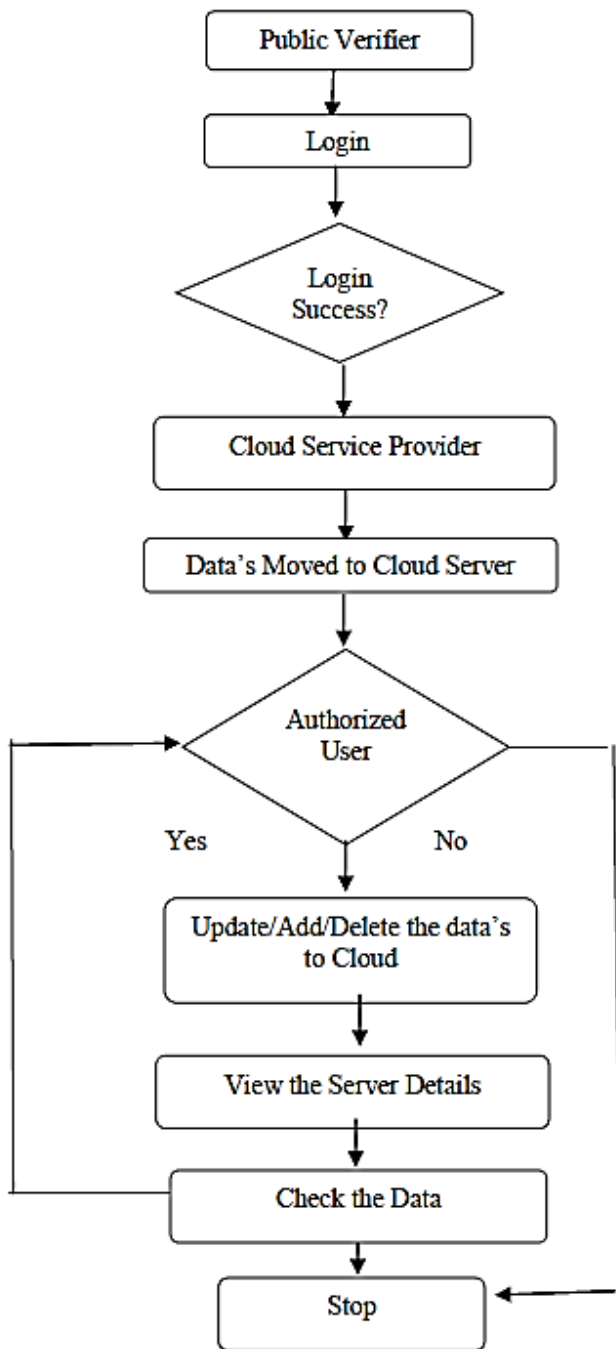


Figure 4. Flow Chart – Public Verifier and Audit Services

In the second module, Hadoop resource module will provide the Hadoop resources. Web based Hadoop computing managing tool for managing Hadoop infrastructure from multiple providers. Here the admin will generate the product and also the products requested by the customers. Hadoop provides a structured marketplace where the product providers will sell the extra products they have and the buyers can take advantage of cheap costs selecting the best service provider at each moment. After purchasing the product the customer can give their own feedbacks.

3. T-Broker Module

In the third module T broker acts as middleware for the hadoop service and the ranking application. Then it uses adaptive and hybrid cloud for computing the degree of trust. In most of the cases T broker produces best results. A hybrid and adaptive trust model to calculate the entire trust degree of service resources, in which trust is the fusion evaluation result from adaptively combining the dynamic product behavior with the social feedback of the product. The HATCM allows Hadoop users to specify their requirements and opinions when accessing the trust score of Hadoop providers. That is, the users can also specify their own preferences, according to their requirements and business policy, to get a customized trust value of the Hadoop providers. A maximizing deviation method is used for computing the direct trust of service resource, which can prevail the limitations in the conventional trust models, in which the trusted attributes have been weighted manually. Meanwhile, this method has a faster convergence than other existing approaches. Sensor-Based Service Monitoring (SSM) is used for monitoring real-time service data of the allocated resources to guarantee the SLA (Service Level Agreement) with the users. In the interactive process, this module will dynamically monitor the service parameters and it is responsible for getting run-time service data. The monitored data is stored in the evidence base, which is maintained by the broker. To calculate the QoS-based trustworthiness of a resource we mainly concentrate on five kinds of trusted attributes of Hadoop services, which consisting of node specific profile, and average of resource usage information, average response time taken, average of task success ratio, and the number of malicious access. Each Hadoop provider offers several VM configurations, are often referred as instance types. An instance type is given in terms of hardware metrics such as CPU

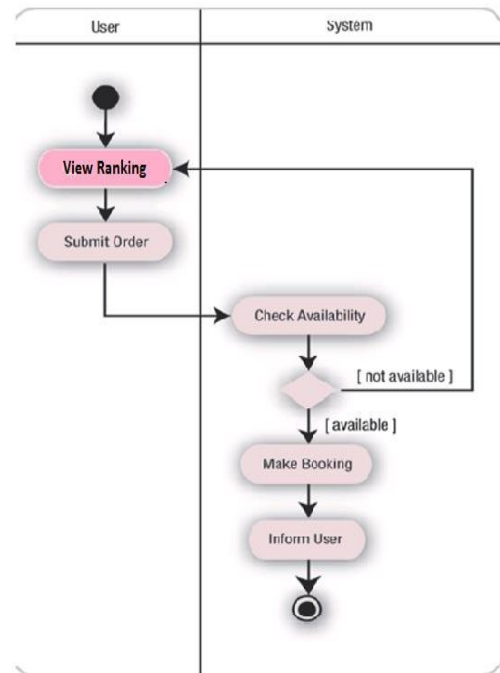
frequency, the size of the memory, the capacity of hard disk, etc. In this paper, the VIM device on the basis of the OpenNebula virtual infrastructure manager module is used to collect and index all the resources information from multiple Hadoop providers. It will get the information from each particular Hadoop provider and will act as resource management interface for the monitoring system. In the multiple Hadoop computing environment, SLA can offer an appropriate surity for the service of quality of sellers, and it serves as the foundation for the expected level of service between the users and the providers. An SLA is a contract agreement of a user with a provider and vice-versa which defines the service quality characters continuously. Adding trust mechanism into the SLA management Hadoop brokering system can prepare the best trustworthiness resources for each request in advance, and provide the best products to users.

MULTIPLE Hadoop theories and technologies are the hot directions in the Hadoop computing industry, which a lot of companies and government are putting much interest to make sure that they have benefited from this new innovation. However, compared with traditional networks, multiple Hadoop computing environment has many unique features such as resources belonging to each Hadoop provider, and such resources being completely distributed, heterogeneous, and fully virtualized; these features indicate that unmodified conventional trust mechanisms cannot be used in multiple Hadoop computing environments. A lack of trust between Hadoop users and providers has hindered the universal acceptance of Hadoop as outsourced computing services

4. Ranking

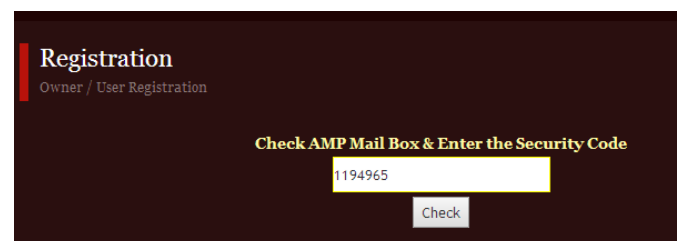
In the fourth module, the ranking of the product is done using the Hadoop framework. The ranking is done based on the review given by the customers who have purchased and used the product. The customers will give the description of the product, its merits, demerits and file about product is uploaded by him which can be viewed in the cloud that can be viewed by the user of the application. Only authorized user can give the review into the system.

D. ACTIVITY DIAGRAM



III. RESULTS AND DISCUSSION

The user will sign up in cloud for accessing the resource. The user and owner will have to register in the cloud. Then the user or owner has to register in the application. After registering the application will ask for OTP which is generated by the admin in the cloud then that OTP is entered in the application and then the registration is successful. By using cloud, only authorized users can get the required resources.



The admin will generate keyword then the customer will give description, merits, demerits and have to upload a file about the product. Here the customer also request for a product then that request will be viewed by the admin and he will accept or reject the request. Each individual customer can view his own review and he can also delete his own review.

Resource Maintenance Portal
Owner's Data Maintenance Port-based on the Given Keyword

Product: BOOK

Title: Java Programming

Description: It contains all the basics of core Java.

Merits: More examples with output. More exercise are given for practice.

DeMerits: Nil

Resource File: Choose File java book.pdf

Submit Reset

The admin will generate ranking based on the product reviews given by the customers. The ranking is done by hadoop

Product	RANKING
HI	4
BOOK	4
SKP	3
PROGRAMMING	3
FRUITS	3
AUTOMOBILES	3
FASTRACK	2
PERFUME	1
BABYSOAP	1
DRESS	1
CAR	1

The user will search for the product keyword. The search is based on keyword and content. After keyword search it will produce the details of the product and file about the product can also be viewed by the user. When we click on view file, the file is downloaded from the cloud or viewed in the cloud. It is based on the format of the file the customer uploads.

IV. CONCLUSION

In this thesis work problem for a cloud customer is to select an appropriate product from the cloud marketplace to support its business needs. However, service guarantees provided by merchants through SLAs contain vague words which make the selecting an ideal provider much more difficult. As customers use cloud services to view the product file and store their individual review and the product file, guarantees related to product quality level is of at most importance. For this purpose, it is imperative from a customer's view to establish trust relationship with a provider. In this proposed system is competence and trust worthiness assessed based on transparency of the provider.

V. FUTURE ENHANCEMENT

The ranking is based on the reviews given by the customers. In future the quality of the product may be checked. Ensuring the quality of the product the review may be given.

VI. REFERENCES

- [1]. Shunmei Meng, Wanchun Dou, Xuyun Zhang, Jinjun Chen, "KASR: A Keyword-Aware Service Recommendation Method based on Map Reduce for Big Data Applications" IEEE Transactions On Parallel And Distributed Systems, TPDS-2013-12-1141.
- [2]. X. Yang, Y. Guo, Y. Liu, "Bayesian-inference based recommendation in social networks," IEEE Transactions on Parallel and Distributed Systems, Vol. 24, No. 4, pp. 642-651, 2013.
- [3]. G. Kang, J. Liu, M. Tang, X. Liu and B. cao, "AWSR: Active Web Service Recommendation Based on the Usage History," 2012 IEEE 19th International Conference on Web Services (ICWS), pp. 186-193, 2012.
- [4]. Yan-Ying Chen, An-Jung Cheng, "Travel Recommendation by Mining People Attributes and Travel Group Types from Community-Contributed Photos" IEEE Transactions on Multimedia, Vol. 15, No. 6, October 2013.
- [5]. M. Alduan, F. Alvarez, J. Menendez, and O. Baez, "Recommender System for Sport Videos Based on the User Audiovisual Consumption," IEEE Transactions on Multimedia, Vol. 14, No.6, pp. 1546-1557, 2013.
- [6]. Zibin Zheng, Xinmiao Wu, Yilei Zhang, Michael R. Lyu, Fellow ,and Jianmin Wang," QoS Ranking Prediction for Cloud Services" IEEE Transactions On Parallel And Distributed Systems, Vol. 24, No. 6, June 2013.
- [7]. G.Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item to Item Collaborative Filtering," IEEE Internet Computing, Vol. 7, No.1, pp.76-80, 2003.
- [8]. Fuzhi Zhang, Huilin Liu, Jinbo Chao, "A Two-stage Recommendation Algorithm Based on K-means Clustering In Mobile E-commerce", Journal of Computational Information Systems, Vol. 6, Issue 10, pp. 3327-3334, 2010.
- [9]. Brian Mc Fee, Luke Barrington and Gert Lanckrieo, "Learning Content Similarity for Music Recommendation" IEEE Transactions on

Audio, Speech, and Language Processing, Vol. 20, No. 8, 2012.

- [10]. Z. D. Zhao, and M. S. Shang, "User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop," In the third International Workshop on Knowledge Discovery and Data Mining, pp. 478-481, 2010.
- [11]. D. Agrawal, S. Das, and A. El Abbadi, "Big Data and Cloud Computing: New Wine or Just New Bottles?" Proc. VLDB Endowment, vol. 3, no. 1, pp. 1647-1648, 2010
- [12]. J. Dean and S. Ghemawat, "Map Reduce: Simplified Data Processing on Large Clusters," Comm. ACM vol. 51, no. 1, pp. 107-113, 2005.
- [13]. S. Ghemawat, H. Gobioff, and S. T. Leung, "The Google File System," Proc. 19th ACM Symp. Operating Systems Principles , pp. 29- 43, 2003
- [14]. Z. Luo, Y. Li, and J. Yin, "Location: A Feature for Service Selection in the Era of Big Data," Proc. IEEE 20th Int'l Conf. Web Service, pp515.