International Journal of Technology and Emerging Sciences (IJTES)

www.mapscipub.com

Volume 01|| Issue 02|| July 2021 || pp. 19-21 E-ISSN: 2583-1925

Performance Measurement in Big-data Analytics Platform

Dr. Pragati Priyadarshinee

Chaitanya Bharathi Institute of Technology(A)

Assistant Professor, Dept. of Information Technology, CBIT, Hyderabad, India

---

Abstract - The aim of the study is to develop a system to increase the performance of Big-data Analytics framework. It provides the method for tool building for efficient processing of large dataset. It identifies the ways to make systems faster by inventing the shortest job completion time. The results will be generated faster. It also reduces the complexity of data retrieval from multiple systems in lesser time. Apache hadoop system also deals with big-data analytics efficiency. The study is a novel approach for solving data-access problem in a big-data system through three-tier architecture.

***Key Words*: Big-data Analytics, Performance Enhancement, Hadoop, Cloud Storage**

1. INTRODUCTION

In the period of huge information, a gigantic measure of information can be created rapidly from different sources. Towards these large information, customary PC frameworks are not capable to store and handle these information. Because of the adaptable and versatile registering assets, distributed computing is a characteristic fit for putting away and handling large information. With iCloud processing, end-clients store their information into the cloud, and depend on the cloud server to share their information to different clients [4]. To just share end-clients' information to approved clients, it is important to configuration access iControl components as indicated by the prerequisites of end-clients. While re-appropriating information into the cloud, end-clients lose the actual control of their information. Additionally, cloud specialist co-ops are not completely trusted by end-clients, which make the entrance control seriously testing. For instance, on the off chance that the customary access control instruments are applied, the cloud server turns into the adjudicator to assess the entrance strategy and settle on access choice ([1][2]). Hence, end-clients might stress that the cloud server might settle on off-base access choice purposefully or inadvertently, and unveil their information to some unapproved clients. To empower end-clients to control the entrance of their own information, some quality based admittance control plans are proposed by utilizing property based encryption. In trait based admittance control, end-clients initially characterize access strategies for their information and encode the information under these entrance arrangements. Just the clients whose ascribes can fulfill the entrance strategy are qualified to unscramble the information. Albeit the current quality based admittance control plans can ideal with the trait renouncement issue, they all experience the ill effects of one issue: the entrance strategy might spill protection. This is on the grounds that the entrance strategy is related with the scrambled information in plaintext structure. From the plaintext of access strategy, the foes might get some security data about the end-client. For instance, Alice scrambles her information to empower the "Brain research Doctor" to access', the approach might contain the qualities "Brain research" and "Specialist". Assuming anybody sees this information, despite the fact that he/she will be unable to decode the information, he/she actually can figure that Alice might experience the ill effects of some mental issues, which releases the security of Alice. To keep the protection spillage from the entrance strategy, a clear technique is to conceal the qualities in the entrance strategy. Nonetheless, when the characteristics are covered up just the unapproved clients yet in addition the approved clients can't realize which credits are associated with the entrance strategy, which makes the unscrambling a difficult issue. Because of this explanation, existing techniques don't conceal the qualities. All things being equal, they just conceal the upsides of each trait by utilizing special cases, Hidden Vector Encryption, and Inner Product Encryption. Concealing the upsides of characteristics can some way or another secure client protection, however the trait name may likewise release private data. Besides, the vast majority of these to some extent stowed away approach conspires just help explicit arrangement structures.

1.1 OBJECTIVES

a) To generate faster results.
b) It reduces the complexity of data access and retrieval.
c) The alternative to this is apache Hadoop, which deals with big data with efficiency.

2. LITERATURE REVIEW

2.1 EXISTING SYSTEM

To empower end-clients to control the entrance of their own information put away on untrusted far off servers, encryption-based admittance control is a compelling strategy, where information are scrambled by end-clients and just approved clients are given decoding keys. Thicken additionally forestall the information security during the transmission over remote organizations which are defenseless against numerous dangers. In any case, customary public key encryption strategies are not appropriate for information encryption since it might deliver numerous duplicates of code text for similar information when there are numerous information purchasers in the systemin request to adapt to this issue, some property based admittance iControl plans are proposed by utilizing characteristic based encryption which just creates one duplicate of code text for every information and doesn't have to know the number of planned information shoppers during the information encryption. Also, when the cloud information are encoded, some accessible encryption calculations are proposed to help looking through scrambled cloud information. Large information Analytics can be connected to Supply-chain the board [11].

2.2 PROPOSED SYSTEM

To propose an efficient and fine-gained big data access control scheme with privacy-preserving policy, where the whole attributes are hidden in the access policy rather than only the values of the attributes [3]. It also designs a novel Attribute Bloom Filter to evaluate whether an attribute is in the access policy and locate the exact position in the access policy if it is in the access policy [9].

3. RESEARCH METHODOLOGY

3.1 SYSTEM ARCHITECTURE

The system architecture represents flow of request from the user to database through the server. Whole system is divided into three tiers using three layers: presentation layer, business layer and data-link layer [6].

The architecture came into picture to remove the obstacles in two-tier architecture. The third tier includes the process management where business logic and rules are executed. It offers flexibility, reusability and increased performance of the system. The advantages of three tier architecture are separate functionality, better understanding, well defined components and effective network performance.

3.2 SYSTEM DESIGN

The System Design Document describes the system requirements, operating environment, system and subsystem architecture, files and database design, input formats, output layouts, human-machine interfaces, detailed design, processing logic, and external interfaces. It consists of Use-case diagram, Sequence diagram, Class diagram and Activity diagram.

4. IMPLEMENTATION

The project contains four modules as end user, consumer data, cloud server and attribute authority [3].

The User has to register. He will be able to login only if he is activated by the attribute authority. The end user may “View Files” and “Upload Files” on the cloud. The Data Consumer also has to register. He will be able to login only if he is activated by the attribute authority. The Data Consumer may “View Files” and “Download Files” which are uploaded by the various End Users. If the Data Consumer wants to download a file from cloud then Key Verification should be done. If the both Keys are Matched then only the Data Consumer can download file [7].

Cloud Server is the Public Account where anyone can store data securely. In Cloud Server, files are stored and downloaded based on the end user and data consumer’s request and also view the “View End Users” and “Data Consumers” who all are using the Cloud Server. The files Uploaded by End User are stored in “Cloud Files”. The files which Data Consumer wants to download have to get request from “User Requests”. The Request is sent to the Data Consumer’s mail ID. The verification of the Attribute Key is done by using the key sent to the mail [8].

Attribute Authority plays the main role which acts as admin. He will Activate and Deactivate End Users and Data Consumers. There are different roles for different attribute keys where we have “Attribute Assigning” which displays the roles and keys respectively. “Graph” shows the rank of files which are uploaded onto the cloud by the End User ([9] [10]). The source code is provided in the annexure. The output is as follows.

Home Page: Home page contains four fields they are “End User”, “Data Consumer”, “Cloud Server” and “Attribute Authority”.

Figure: Home-Page

5. CONCLUSION

In this paper, Big Data Analytics with Privacy Preservation using CP ABE Technique, where the access policy will not leak any privacy information. Different from the existing methods which only partially hide the attribute values in the access policies, our method can hide the whole attribute in the access policies [5]. However, this may lead to great challenges and difficulties for legal data consumers to decrypt data. To cope with this problem, we have also designed an attribute localization algorithm to evaluate whether an attribute is in the access policy. In order to improve the efficiency, a novel Attribute Bloom Filter has been designed to locate the precise row numbers of attributes in the access matrix. We have also demonstrated that our scheme is selectively secure against chosen plaintext attacks. Moreover, we have implemented the ABF by using Murmur Hash and the access control scheme to show that our scheme can preserve the privacy from any LSSS access policy without employing much overhead. In our future work, we will focus on how to deal with the offline attribute guessing attack that check the guessing "attribute strings" by continually querying the ABF.

REFERENCES

[1] B. Waters, "Ciphertext-policy attribute-based encryption: An expressive,efficient, and provably secure realization," in Proc. of PKC'11. Berlin,Heidelberg: Springer-Verlag, 2011, pp. 53–70.
[2] C. Dong, L. Chen, and Z. Wen, "When private set intersection meets big data: an efficient and scalable protocol," in Proc. of CCS'13. ACM, 2013, pp. 789–800.
[3] D. Boneh and B. Waters, "Conjunctive, subset, and range queries onencrypted data," in Theory of cryptography. Springer, 2007, pp. 535–554.
[4] H. Li, D. Liu, K. Alharbi, S. Zhang, and X. Lin, "Enabling fine-grainedaccess control with efficient attribute revocation and policy updatingin smart grid," KSII Transactions on Internet and Information Systems(TIIS), vol. 9, no. 4, pp. 1404–1423, 2015.
[5] H. Lin, Z. Cao, X. Liang, and J. Shao, "Secure threshold multi authorityattribute based encryption without a central authority," in Proc. OfINDOCRYPT'08. Springer, 2008, pp. 426–436.
[6] J. Li, K. Ren, B. Zhu, and Z. Wan, "Privacy-aware attribute-basedencryption with user accountability," in Information Security. Springer, 2009, pp. 347–362.
[7] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supportingdisjunctions, polynomial equations, and inner products," in Advancesin Cryptology–EUROCRYPT'08. Springer, 2008, pp. 146–162.
[8] J. Lai, R. H. Deng, and Y. Li, "Fully secure cipertext-policy hiding cpabe,"in Information Security Practice and Experience. Springer, 2011,pp. 24–39.
[9] K. Yang and X. Jia, "Expressive, efficient, and revocable data accesscontrol for multi-authority cloud storage," IEEE Trans. Parallel Distrib.Syst., vol. 25, no. 7, pp. 1735–1744, July 2014.
[10] K. Yang, Z. Liu, X. Jia, and X. S. Shen, "Time-domain attribute-basedaccess control for cloud-based video content sharing: A cryptographicapproach," IEEE Trans. on Multimedia (to appear), February 2016.
[11] Raut, R. D., Mangla, S. K., Narwane, V. S., Gardas, B. B., Priyadarshinee, P., & Narkhede, B. E. (2019). Linking big data analytics and operational sustainability practices for sustainable business management. *Journal of cleaner production*, *224*, 10-24.