THE EFFICIENT QUERY EXECUTION FRAMEWORK

Dr. C.K. Gomathy[1], Mr.K. Vishnu Vardhan[2], Ms.Pavithra[3]

---***---

ABSTRACT:

Data analysis helps enterprise to make decisions on customer trends and behavior prediction. Often the queries would be complex and takes more time to execute on the cluster. This project would help queries to execute on different platform based on the configuration and cache the response for future reference. The project helps customer to analyze the data faster and turn data into learning insights which they can use further for making the decision. In Enterprize data is important. In companies, Decision can be made through huge volume of data. We cannot store huge volume of data in RDBMS and we can't upgrade it because it is vertical Scala. But Here Hadoop is horizontal Scala and it stores huge volume of data in distributed manner. Whenever we try to process the data, it happens in only.one machine. But Our project helps to process the data in multiple machines.

- As part of our project, we try to use plug-in architecture where it is capable of configuring multiple execution frameworks. So, the execution configuration can be configured in different frameworks like MapReduce, Tez, Spark or any feature in the data analysis framework in present or future.

- Query execution framework also helps to measure the performance of any different framework.

- The Execution framework is also helping in capable of responding the output in the standard exchange format like JSON.

- The Framework also helps the fastened query execution and cashing the frequently executed query in MariaDB.

- The Query can also be implemented for Terabytes of data because of Hadoop.

Keywords: Efficient Query, Query Execution framework, analyze the data execution and data analysis framework

I. INTRODUCTION

Enterprises are adapting large-scale processing platforms, like Hadoop, to understand actionable insights from their "big data". Query optimization remains an open challenge during this environment because of the quantity and heterogeneity of data, comprising both structured and semi-structured datasets.

Processing large-scale data within the amounts of the many terabytes could also be a really difficult task. Solving the problems associated with high volume data requires is often achieved by dividing the data and work to many computers which can all work together in parallel to finish the task during a reasonable time. Map-Reduce and Hadoop have gained popularity in Parallel dataflow systems. These systems are extensively used for analytics. Big data has been produced in various applications, including the user information of social networks, sensor data, scientific data and sort of log data, etc... Therefore, big data query processing is playing a crucial role in today's fast-paced data-driven businesses. because the world is getting digitized the speed during which the quantity of knowledge is over throwing from different sources in several format, it's impossible for the normal system to compute and analysis this type of massive data that big data tool like Hadoop is employed which is an open-source software. It stores and computes data during a distributed environment. Data analysis helps enterprise to form decisions on customer trends and behavior prediction. Often the queries would be complex and takes longer to execute on the cluster.

II.EXISITING SYSTEM

Query execution occurs over a query plan, which describes a sequence or expression evaluation tree of query operators (as well as the algorithms to use in executing each operator). The Query will be executed through any of the frameworks as per our wish and the output will be shown. Sometimes we need to execute the same query again and again. Small queries can be executed as many times as we need. But, here in the case of large and complex queries, executing again and again is very difficult and we know Bigdata means a huge amount of data and it needs large and complex queries for analysis.
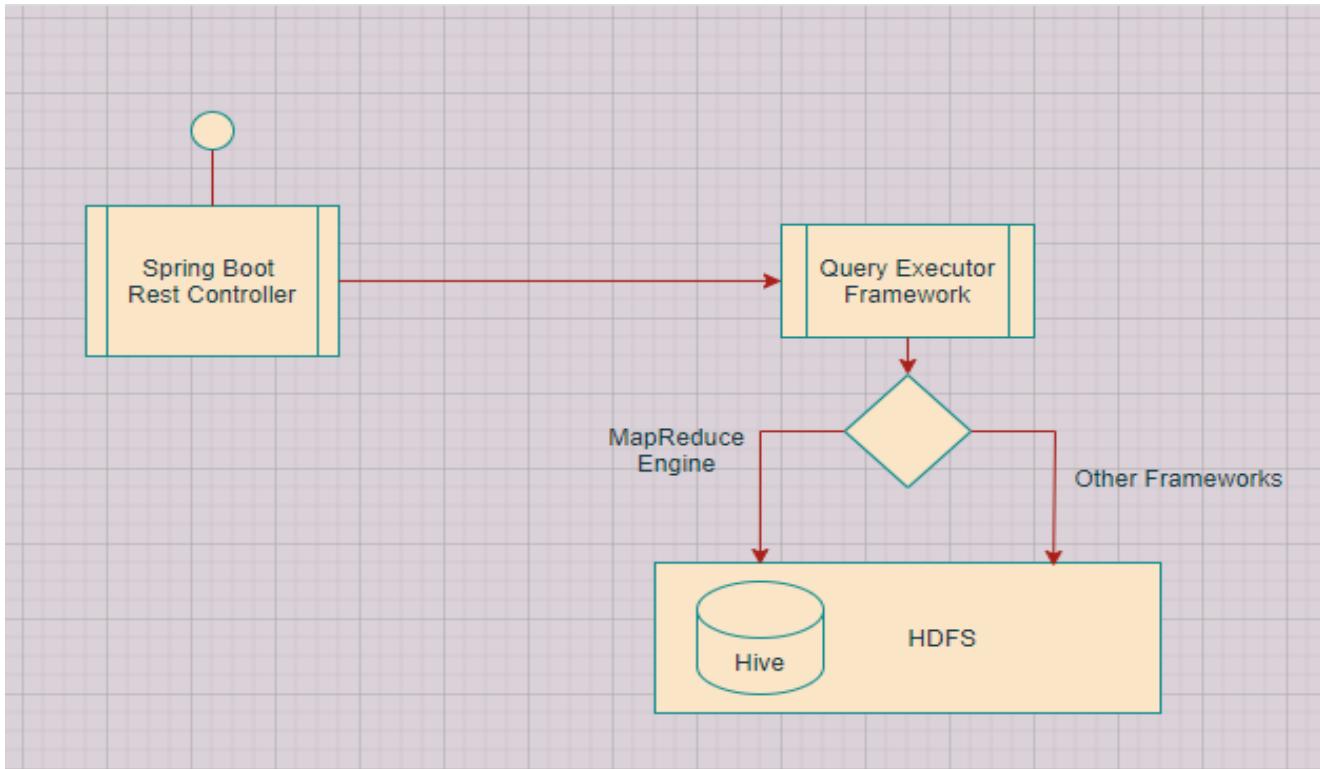

Fig 1: Existing System Architecture

III. METHODOLOGY

Let's see the process of implementation of the project:

First, we choose the Framework in which we want to execute the Query.

Suppose if we choose MapReduce.

- Executing Query in Hive through Map Reduce Framework.
- Converting it to JSON Format.
- Storing it in Maria DB using Cache Manager.
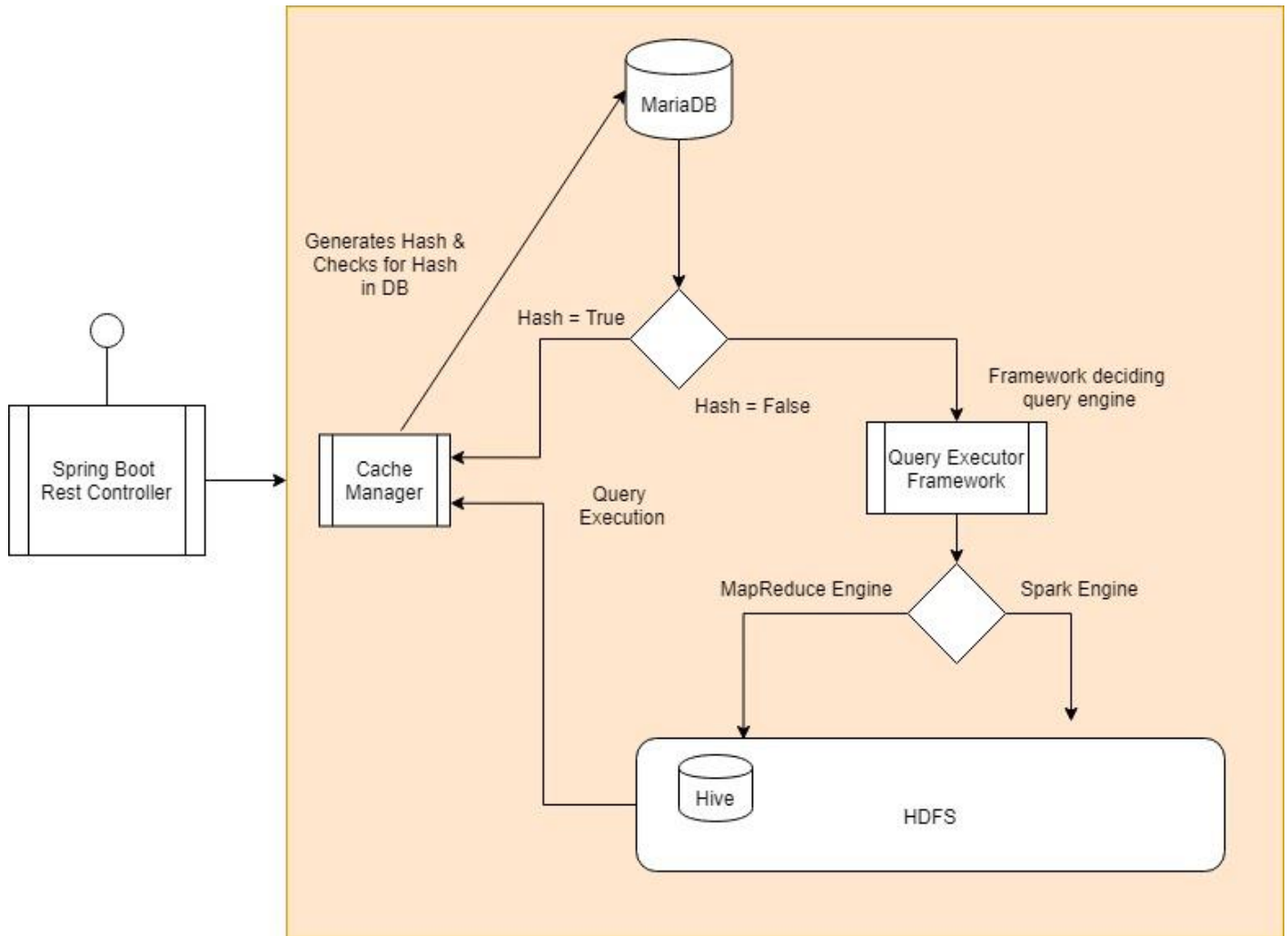- Cleaning of Data in Maria DB when some Queries are not being used frequently.

Fig 2: Proposed System Architecture

Here in our proposed system, when we want to execute the query, with the help of Cache Manager first it checks the MariaDB (that we added) whether the Query response that we want to execute presents in the Database or not. If not, then it will go for execution. The Database will store the frequently executed queries response with the help of the cache manager after the conversion to Json format. By this process, there is no need of executing large and complex queries again and time optimization is possible and query execution can be done in any framework as per our convenience in present technologies or upcoming technologies in the future.

- **Hash:** Hashing is that the method of converting a given key into another value. The results of a hash function are understood as a hash value or just, a hash.
- **Cache Manager**
- **MariaDB**
- **Tez:** It is an extensible framework for building high-performance batch and interactive processing data applications, coordinated by YARN in Apache Hadoop.
- **JSON:**
  JSON format is employed for serializing and transmitting structured
  data over a network connection. It is primarily used to transmit data

between a server and web applications. Web services and APIs use
JSON format to provide public data.

IV. CONCLUSION

In this work, we proposed an Efficient Query Execution based framework. The proposed system architecture will overcome the problems and difficulties in existing system and the performance of the proposed system exceeds the existing in terms of performance measures such as Query Response time. The main benefit of the framework is that, the execution configuration can be configured in different frameworks like Map Reduce, Tez, Spark or any framework in data analysis framework in present or future. To improve the overall performance of the framework, query need to be processed efficiently.

I am hopeful that many of the lessons I have learned from efficient query processing will translate to this new realm, and that much of the work of others will also be useful. It would be a significant. Achievement if the data management community really could manage most of the world's data, and if users and application programmers could store and query their data in a transparent, initiative way.

V. REFERENCES

[1] High-Efficient Fuzzy Querying with HiveQL for Big Data Warehousing Published in: IEEE Transactions on Fuzzy System Date of Publication: 29 March 2021

[2] A development of RDF data transfer and query on Hadoop Framework   December 2016: International Journal of Pharmacy and Technology

[3] Efficient query processing framework for big data warehouse: an almost join-free approach April 2015

[4] Optimization of Multiple Queries for Big Data with Apache Hadoop/Hive 2015 International Conference on Computational Intelligence and Communication Networks

[5] Multi-Q: Multiple Queries Optimization based on MapReduce in Cloud 2014 Second International Conference on Advanced Cloud and Big Data

[6] A multi-agent based intelligent query processing system for Hadoop with FIPA-OS using cooperating agent in cloud environment 2016 10th International Conference on Intelligent Systems and Control (ISCO)

[7] Analyzing the query performances of description logic-based service matching using Hadoop 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)

[8] Review of Hadoop performance optimization IEEE International Conference on Computer and Communications

[9] Dr.C K Gomathy, Article: An Effective Innovation Technology In Enhancing Teaching And Learning Of Knowledge Using Ict Methods, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) *E*-Issn: 2395-5325 Volume3, Issue 4,P.No-10-13, April '2017

[10] Dr.C K Gomathy, Article: A Semantic Quality of Web Service Information Retrieval Techniques Using Bin Rank, International Journal of Scientific Research in Computer Science Engineering and Information Technology (IJSRCSEIT) Volume 3 | Issue 1 | ISSN : 2456-3307, P.No:1563-1578, February-2018

[11] Dr.C K Gomathy, Article: A Web Based Platform Comparison by an Exploratory Experiment Searching For Emergent Platform Properties, IAETSD Journal For Advanced Research In Applied Sciences, Volume 5, Issue 3, P.No-213-220, ISSN NO: 2394-8442,Mar/2018

[12] Dr.C K Gomathy, Article: A Study on the Effect of Digital Literacy and information Management, IAETSD Journal For Advanced Research In Applied Sciences, Volume 7 Issue 3, P.No-51-57, ISSN NO: 2279-543X,Mar/2018

[13] Dr.C.K.Gomathy, A.V.Sripadh Kaustthub, K.Banuprakash, Article: An Effect of Big Data Analytics on Enhancing Automated Aviation , International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) *E*-Issn: 2395-5325 Volume 4, Issue 3,P.No-1-7.March -2018

[14] Dr.C K Gomathy, Article: A Semantic Quality of Web Service Information Retrieval Techniques Using Bin Rank A Cloud Monitoring Framework Perform in Web Services, International Journal of Scientific Research in Computer Science Engineering and Information Technology IJSRCSEIT | Volume 3 | Issue 5 | ISSN : 2456-3307,May-2018

[15] Dr.C K Gomathy, Article: Supply chain-Impact of importance and Technology in Software Release Management, International Journal of Scientific Research in Computer Science Engineering and Information Technology (IJSRCSEIT) Volume 3 | Issue 6 | ISSN : 2456-3307, P.No:1-4, July-2018

[16] Dr.C K Gomathy, Article: A Scheme of ADHOC Communication using Mobile Device Networks, International Journal of Emerging technologies and Innovative Research (JETIR) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:320-326, Nov-2018

[17] Dr.C K Gomathy, Article: A Study on the recent Advancements in Online Surveying , International Journal of Emerging technologies and Innovative Research (JETIR) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:327-331, Nov-2018

[18] C.K. Gomathy, V. Geetha, Venkata Nagendra, Venkata Sai Ram, Article: The Secured Proficient Smart Electronic Voting System, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-4, April 2020

[19] C K Gomathy,V Geetha,P.S.lavakumar,K.V.Rahul,The Efficient Automatic Water Control Level Management Using Ultrasonic Sensor, International Journal of Computer Applications (0975 – 8887) Volume 176 – No. 39, July 2020

[20]C K Gomathy, Student Location Tracking Inside College Infrastructure, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 06 | June 2020

AUTHOR'S PROFILE

1 **Dr.C.K.Gomathy** is Assistant Professor in Computer Science and Engineering at Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university,Enathur,Kanchipuram,India. Her area of interest is Software Engineering,Web Services, Knowledge Management and IOT.

2. **K.Vishnu Vardhan**, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university,Enathur,Kanchipuram,India.His Area of Interest image processing,Software Engineering.

3. **Pavithra**, student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university,Enathur,Kanchipuram,India Her Area of Interest is image processing,Software Engineering.