Original Article

DISTRIBUTED DATA CACHING MANAGEMENT USING ANALYTICS

K. Vishnu Vardhan[1], Unnam Pavithra [1], Dr. C. K. Gomathy[2], Dr. V. Geetha[2]

[1,] Student, [2] Assistant Professor, Department of CSE, SCSVMV (Deemed to be University), Kanchipuram, Tamil Nadu, India.

DOI: 10.47750/pnr.2022.13.S07.867

Abstract

Queries are the coronary heart of analytics. Without them, there may be no analytics. Data evaluation is assisting the business enterprise to make the very last choices on client developments and predictions and the queries could be massive and take a lot time to execute at the cluster. This task could assist queries to execute on distinctive systems primarily based totally at the configuration and cache the reaction for destiny reference through the use of a database. The task enables clients to research the information quicker and they could use it similarly for making the decision. In Enterprize information is important. In companies, Decisions are to be made through the massive extent of information. We cannot save the big extent of information in Relational databases and we can`t replace it due to the fact it's far vertical Scala. But Here Hadoop in HDFS is horizontal Scala and it shops a massive extent of information in a dispensed manner. Our task enables in shifting information from Relational databases like MySQL to Hadoop and it'll be accessed through the Java Database Connectivity (JDBC). The accessed information might be transformed to the JavaScript item notation (JSON) layout through the use of the Gson library. The transformed JSON layout might be saved in a database for destiny use.

Keywords: Caching, MariaDB, Sqoop, Json, Gson, Maven and JDBC, Query Optimization, Bigdata.

Introduction

It is produced in several operations, which includes the data of social networks, clinical information, etc. thus, massive information question processing is gambling the primary component in moment`s speedy information-pushed businesses. it`s now no longer insolvable for the everyday gadget to dissect this form of large complicated information that massive information device like Hadoop is employed, that is open - supply software. It shops and analyzes information in a dispensed terrain. Relational database operation structures (RDBMS) warrant excessive velocity due to the fact it is designed for consistent information manage in place of rapid-hearthplace boom.

Indeed if Relational database operation structures (RDBMS) are used to address and keep massive information, additionally it's going to grow to be assuredly expensive. The look of the Internet had redounded in velocity boom of the information length endlessly. Distributed way processing and storing of comparable large information units had became out to be a massive task for the database assiduity.' Big information' time period is normally used to keep large information units. Relational Database Management Systems (RDBMS) can't deal with the ones massive information units. For properly storehouse of information and analytics, we are moving to Hadoop dispensed educate structures (HDFS).

It's a body for massive information operation and analysis. For bulk information transfers we want a device in among known as Sqoop, which hundreds information from Relational database operation structures (RDBMS) to Hadoop dispensed educate structures (HDFS). Query optimization remains an open task for several organizations due to the quantity like large quantity and one-of-a-kind forms of datasets like based and unshaped. Processing massive- scale information withinside the several petabytes is a assuredly sensitive task.

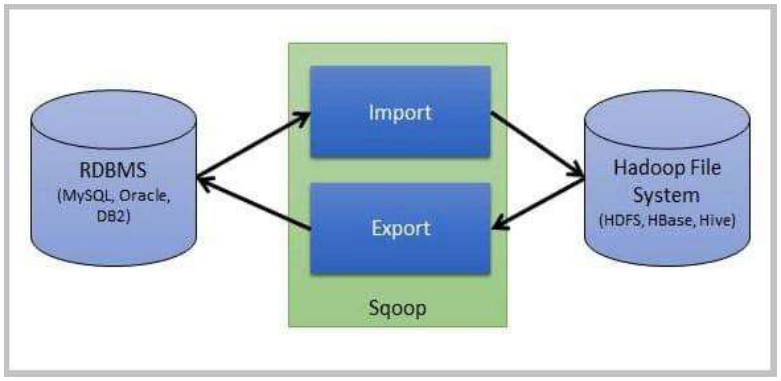
Existing System

The Execution of the Commands can be carried out with the aid of using Hadoop the usage of the shifting device known as Sqoop and the information ingestion can be carried out however later while there may be a want of shifting up to date information and newly delivered information, we need to execute the identical instructions once more. That's a number of time-ingesting as we recognise massive information way a massive quantity of information. That's simplest for shifting purposes. Later, we want to execute the Query in Hive for the transferred information. Big information queries can be complicated and massive due to the information. Sometimes there may be a hazard of Repetition of Queries withinside the destiny. Executing the identical queries once more additionally consequences in a number of time-ingesting. The present gadget isn't always powerful in terms of Time Efficiency.

Project Illustration

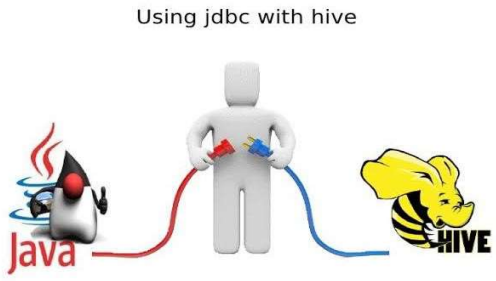
The most important intention of the undertaking is to lower time loss withinside the subject of bigdata analytics. For that we used a few gear and technology withinside the undertaking. Let's see approximately them in detailed.

SQOOP: It is a device that is specifically designed for the Bulk transformation of information among Relational databases and the Apache Hadoop. We used it to convert information from MySQL database to Hive, which affords in Hadoop Because, the database can not deal with massive quantity of information. Below is the architecture of the Sqoop.



HADOOP: Apache Hadoop is a framework, that is used for dispensed processing of massive information units throughout many computer systems in a cluster. It is a horizontal Scala. It can technique very massive length of information with the aid of using dispensing paintings to numerous hundreds of machines, it's far called HDFS known as Hadoop Distributed File Systems.

APACHE HIVE: It is constructed on pinnacle of the Hadoop. It permits customers to do moves at the massive datasets the usage of the Structured Query Language (SQL). Using the Hive, we will read, write and manipulate massive datasets. It is will paintings quicker even at the petabytes of information.



JAVA DATABASE CONNECTIVITY (JDBC): It is used to get entry to database control structures with the aid of using writing a java program. It is an utility programming interface (API) used to execute the question from the java program. We used it to get entry to the Hive the usage of Hive JDBC driver.

GSON Library: It is one of the Java Library. It can convert Java gadgets to Json Representation and Json Representation to Java Objects. We used it to transform the Hive information to Json layout after having access to thru JDBC program.

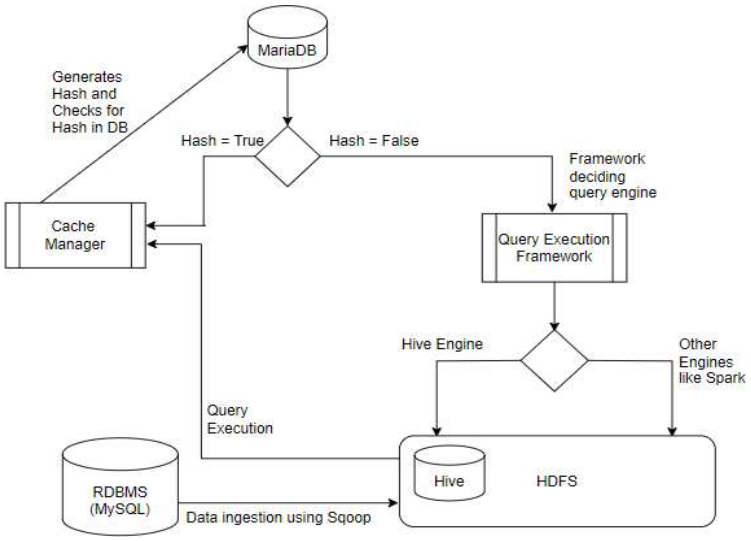
CACHE MANAGER: It is used to keep the cache. We used it to cache the json formatted information right into a database and retrieving the information from database withinside the destiny.

Proposed System

In our proposed gadget, we triumph over all of the issues in Existing gadget specifically time performance. We lessen the time loss and we growth the time performance in each the factors like Sqoop instructions and the hive Queries. We will keep the Sqoop instructions, which might be used for information shifting from Relational databases to Hive in a Sqoop task for in addition use and what we must do withinside the destiny is that, we simply must execute that present Sqoop task on every occasion there may be want of the vintage instructions. By doing that we can store time.

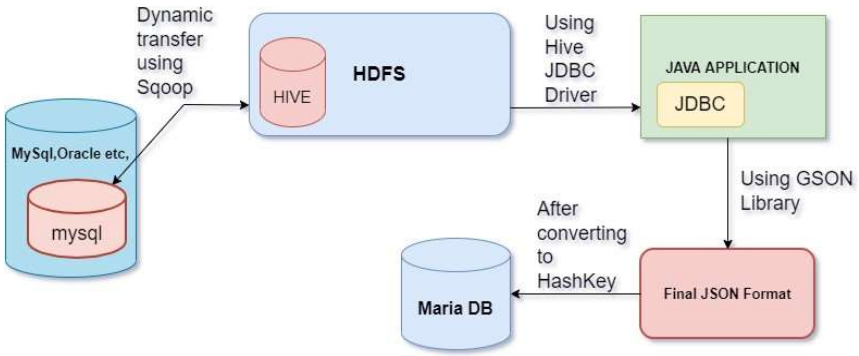
Later every other trouble is that repetition of Hive queries, which ends up in lot of time loosing. For that what we do is that, we get entry to the Hive the usage of java JDBC and we convert the question reaction to Json layout, which is straightforward manner to get entry to. Later we keep that transformed reaction in a database with the aid of using the usage of the cache manager. By doing that, we will retrieve that data withinside the destiny with the aid of using the usage of cache manager. This reduces the lot of time intake for repetition queries.

Architecture/Structure



'How the project flows from first point to last point?' We can see that by the above diagram.

With the below picture we can understand how process works from Relational databases to the Maria DB:



Implementation

The main part of the project is the implementation part. There are mainly three modules in the project. The first module is about the Sqoop, the second module is about the JSON conversion and the third module is about storing the output response in a database using cache manager. Let's see about it in detail.

- Selecting the table in MySQL, which we want to import to Hive. Later it will be imported to Hive using Sqoop commands. Later a Sqoop job will be created for the purpose of future updates and new rows which will be added to the table in the MySQL database. In the future, there is no need for Sqoop commands again we just have to execute the job.

- Now we should create a java maven project in any java platform. We used IntelliJ IDEA here. The transformed table will be accessed by the Java JDBC program. Before that, we should add the dependencies of Hadoop, hive, and Gson to the pom.xml file, which presents in the java maven project. We execute a query there. Later by using the Gson libraries, we convert every record in the query to Java objects and the java objects to JSON strings. The final JSON format will be done.

- We will generate a unique hash key for the final JSON formatted data. The hash key will be stored in a database by using the cache manager for future use.

- An example for the Json format data is given below: It will be in key and value pairs.

Here keys are id, fname, lname, state and values are remaining part.

"id": 1,

"fname": "Gomathy",

"lname": "CK",

"state": "Tamilnadu"

"id": 2,

"fname": "Pavithra",

"lname": "Unnam",

"state": "Andhra Pradesh"

"id": 3,

"fname": "Vishnu",

"lname": "Vardhan",

"state": "Andhra Pradesh"

- When there is a need in the future, we convert the requested parameters to SQL query and we generate a hash for the query. Later we write a code to check if the query response is already present in MariaDB or not. If it exists then we render the response from the MariaDB using cache manager. If the response is not found then it goes for the execution in the hive. In this way, we can reduce a lot of time in the bigdata analytics.

The below screenshots will help you understand more:

Fig. 1. (Rows count and Schema of the table in MySQL)

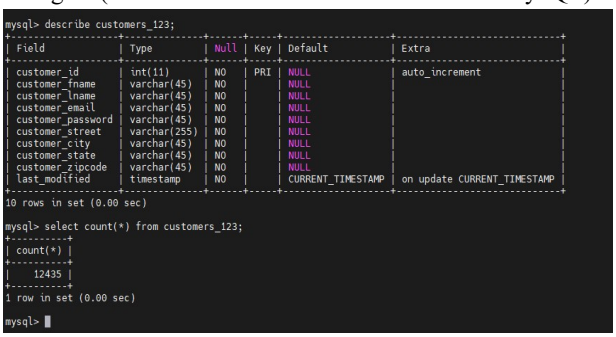


Fig. 2. (Sqoop Job Creation for future use)

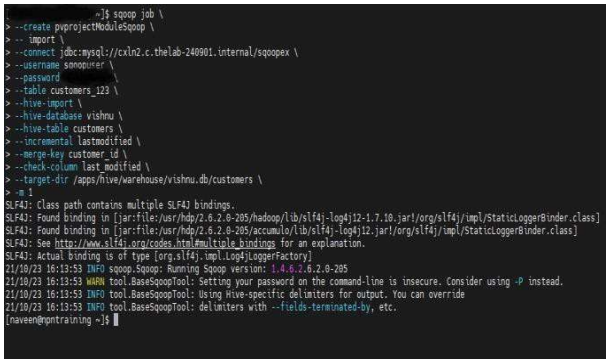


Fig. 3. (Accessed data from Hive using Java JDBC Program)

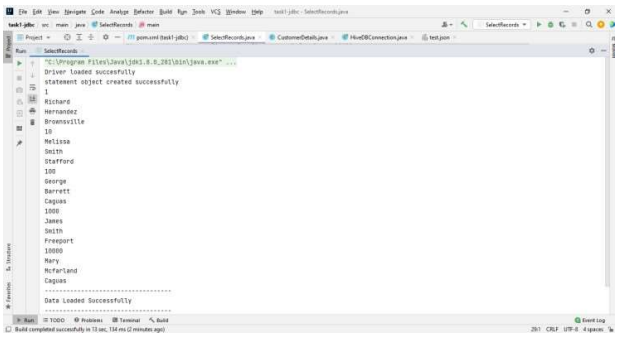


Fig. 4. (Final Json converted data using Gson Library)

Social Benefits

- Saves lot of time and energy
- Easier Query execution in big data analytics
- One time execution but many times retrieving
- Dedicated query storing database
- Data engineers and data analysts will be more benefited

List of Abbreviations

RDBMS	Relational Database Management System
JDBC	Java Database Connectivity
HDFS	Hadoop Distributed File Systems
JSON	Java Script Object Notation
SQL	Structured Query Language

Conclusion

In this work, the system that we proposed will help the big data engineers to make the process of query execution easy. This work will overcome all the problems in the present query execution and query response time will be faster than the existing system. Caching is much required in the analytics field. The main benefit of this work is the time efficiency. The data ingestion is now very easy by using the Sqoop job. Sqoop job ingests the data in less time and the performance of that is very high. As data is being large in our data to day life, the difficulty in processing and managing that large data will increase. This work helps in decreasing that difficulty. The work we are doing helps a lot in analytics.

References

1. DR.C.K. Gomathy, V. Geetha, S. Madhumitha, S. Sangeetha, R. Vishnupriya, Article: A Secure With Efficient Data Transaction In Cloud Service, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, March 2016, ISSN: 2278 – 1323.
2. Dr.C.K. Gomathy, C.K. Hemalatha, Article: A Study on Employee Safety and Health Management. International Research Journal of Engineering and Technology (IRJET) - Volume: 08 Issue: 04 | Apr 2021.
3. 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.
4. 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 Volume 3, Issue 4, P. No. 10-13, April '2017.
5. 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.
6. C.K Gomathy and V. Geetha. Article: A Real Time Analysis of Service based using Mobile Phone Controlled Vehicle using DTMF for Accident Prevention. International Journal of Computer Applications 138(2): 11-13, March 2016. Published by Foundation of Computer Science (FCS), NY, USA, ISSN No: 0975-8887.
7. C.K. Gomathy and V. Geetha. Article: Evaluation on Ethernet based Passive Optical Network Service Enhancement through Splitting of Architecture. International Journal of Computer Applications 138(2): 4-17, March 2016. Published by Foundation of Computer Science (FCS), NY, USA, ISSN No: 0975-8887.
8. C.K. Gomathy and Dr.S. Rajalakshmi. (2014), "A Software Design Pattern for Bank Service Oriented Architecture", International Journal of Advanced Research in Computer Engineering and Technology (IJARCET), Volume 3, Issue IV, April 2014, P. No: 1302-1306, ISSN: 2278-1323.
9. C.K. Gomathy and S. Rajalakshmi, "A software quality metric performance of professional management in service oriented architecture," Second International Conference on Current Trends in Engineering and Technology - ICCTET 2014, 2014, pp. 41-47, DOI: 10.1109/ICCTET.2014.6966260.
10. Dr.C.K. Gomathy, V. Geetha, T.N.V. Siddartha, M. Sandeep, B. Srinivasa Srujay, Article: Web Service Composition In A Digitalized Health Care Environment For Effective Communications, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, April 2016, ISSN: 2278 – 1323.
11. C.K. Gomathy. (2010), "Cloud Computing: Business Management for Effective Service Oriented Architecture", International Journal of Power Control Signal and Computation (IJPCSC), Volume 1, Issue IV, Oct - Dec 2010, P.No:22-27, ISSN: 0976-268X.
12. 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.
13. Dr.C.K. Gomathy, C.K. Hemalatha, Article: A Study on Employee Safety and Health Management International Research Journal of Engineering and Technology (IRJET) - Volume: 08 Issue: 04 | Apr 2021.
14. Dr.C.K. Gomathy, V. Geetha, T. Jayanthi, M. Bhargavi, P. Sai Haritha, Article: A Medical Information Security Using Cryptosystem For Wireless Sensor Networks, International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) E-ISSN: 2395-5325 Volume 3, Issue 4, P. No. 1-5, April '2017.
15. C.K. Gomathy and Dr.S. Rajalakshmi. (2014), "Service Oriented Architecture to improve Quality of Software System in Public Sector Organization with Improved Progress Ability", Proceedings of ERCICA-2014, organized by Nitte Meenakshi Institute of Technology, Bangalore. Archived in Elsevier Xplore Digital Library, August 2014, ISBN: 978-9-3510-7216-4.
16. Parameshwari, R. & Gomathy, C.K. (2015). A Novel Approach to Identify Sullied Terms in Service Level Agreement. International Journal of Computer Applications, 115, 16-20, 10.5120/20163-2253.
17. C.K. Gomathy and Dr.S. Rajalakshmi. (2014), "A Software Quality Metric Performance of Professional Management in Service Oriented Architecture", Proceedings of ICCTET'14, organized by Akshaya College of Engineering, Coimbatore. Archived in IEEE Xplore Digital Library, July 2014, ISBN: 978-1-4799-7986-8.
18. C.K. Gomathy and Dr.S. Rajalakshmi. (2011), "Business Process Development In Service Oriented Architecture", International Journal of Research in Computer Application and Management (IJRCM), Volume 1, Issue IV, August 2011, P. No. 50-53, ISSN : 2231-1009.
19. Dr.C.K. Gomathy, Dr.V. Geetha, G.S.V.P. Praneetha, M. Sahithi Sucharitha. (2022). Medicine Identification Using Open Cv. Journal of Pharmaceutical Negative Results, 3718–3723. https://doi.org/10.47750/pnr.2022.13.S09.457

20. Dr.V. Geetha, Dr.C.K. Gomathy, Kommuru Keerthi, Nallamsetty Pavithra. (2022). Diagnostic Approach To Anemia In Adults Using Machine Learning. Journal of Pharmaceutical Negative Results, 3713–3717. https://doi.org/10.47750/pnr.2022.13.S09.456
21. Vishnupriya C.K. and et al, Dimensional and Morphologic Variations of palatal Rugae-a hospital based study among Chennai populations, International Journal of Science Research, ISSN No: 2277-8179 Volume 7, Issue 7, P. No. 19-20, July ’2018.