

An Advanced Methodology for Processing Geo-Dispersed Big Data Using Map Reduce

Vasantha Kumari ¹, Swamy L N ², Sukruth Gowda M A ³, Rajesh S M ⁴

P.G. Student, Department of Computer Science & Engineering, VTU-CPGSB, Muddenahalli, Karnataka, India¹

Assistant Professor, Department of MCA, VTU-CPGSB, VIAT, Muddenahalli, Chickballapur, Karnataka, India²

Assistant Professor, Department of IS&E, SVIT, Bengaluru, Karnataka, India³

Assistant Professor, Department of CS&E, GITAM School of Technology, Bengaluru, Karnataka, India⁴

Abstract: Big Data focuses on a new era of data collection and execution. For the current world applications Big Data provides a large volume of uses and satisfactions towards expectation. A popular infrastructure of a Big Data in cloud computing allows processing of a Big Data applications. Practical real world applications are increasing with Big Data cover increase in usage of cloud computing, so it became part of many Big Data applications. The proposed new MapReduce-based framework is to process Geo-dispersed Big Data. The framework supports an accessing of vehicle details based on entering a vehicle color and vehicle license number to get vehicle details according to user requisition. In this paper we are managing about vehicle details in an emerging MapReduce framework, which is often data to be stored in HDFS file directories. We are saving login in the form of encrypted data. Here we propose a novel MapReduce framework a suite of mechanism for data access control to vehicle information stored in HDFS file directories.

KEYWORDS: Advanced MapReduce, HDFS file directories, Big Facts.

I. INTRODUCTION:

In recent years, a MapReduce operation on Geo-dispersed big data has emerged as vehicle details information for reference purpose. A MapReduce operations on Geo-dispersed big data with help of Advanced MapReduce framework(AMF) that allows to enter vehicle color, vehicle number and other like licence number of vehicle related data in one place through the HDFS, which has made the storage, retrieval(like accessing vehicle details) and if we need sharing the vehicle information more efficient. Especially each vehicle has its unique licence number or unique vehicle number to access the details about vehicle, but by entering the vehicle color, it is difficult to access because more number of vehicles had the same color. So that preferring of licence number or vehicle number that gives better result.

Data is already stored in the HDFS file by administrator through High Definition Camera at roadsides and junctions of high definition type for good quality. This application an authenticate person can only accessing for storage, retrieval and update of a vehicle

details. After successful it authentication go for work with an application. Already a data is stored in HDFS is to be accessed for work with an application. Big data applications are a large amount of data applications. For user or people convenience and benefits these applications are full-size facts applications because the records volume is huge and this updating facts is very rapid and data rate. Fast update data rate is to be need of a big data application because of daily updates to the data. This data is to be captured by high definition camera.

Due to the large amount of update data every data to big data makes difficult to extract. Multiple input makes extracting data is difficult. To improve the above solutions is, using an AMF the easy to extract every day update big data easily. First, an application work with the big data are large collection of facts that is storing, retrieval of a data for daily updating of a data makes difficult. There is a possibility for accessing that data related to vehicle information is to be possible so that a security purpose encryption and decryption method for every time users makes difficult. In this application for security purpose, sharing of details is to be avoided. Daily update of a data related to the vehicle information makes difficult to handle the data. For this purpose administrator always be active with storage activities. In this application using an AMF solving the more amount of updating data and gives the results will be reduced form.

II. LITERATURE SURVEY:

In paper [1], for group of gear a chart reducer is to be used for the distributed or data is to be displayed to the other systems of group to compute. So like this calculation is to be done in a loss amount of time by dispatching jobs. In this using techniques parallel and distributed operations. Disadvantages are, not using encryption and decryption methods and not available fully security for this data and hidden forms of parallelization, error charity, area group and fill matching huge volume of facts to handle difficult to immediate implementation of goods.

In paper [2], with the help of a cell cloud trust management provides the including of identity executive, key managing and security. By a user centric identity using techniques that is secure information processing

framework of cloud computing to mobiles is to be done. Disadvantages is, teenagers are especially prove to text and drive which can be dangerous. If the already fixed threshold value is to be beyond at present threshold value it is difficult track a vehicle details.

In paper [3], using a vehicle start time and end time, its travelled distance from start time from travel network that trying to find fuel consumption. In these using techniques that is Global Positioning System. The disadvantages are, not secure for the data and in this not find a vehicles information like start time and end time in geographically. In this, forget the user ID and password then can't access the information about the vehicles.

In paper [4], now a day, it is necessary to adopt a parallel data processing because of a large quantity of update data and increase in storage, MapReduce in storage. In this paper using map reduce but in main project using complex MapReduce structure. Disadvantages are MapReduce does not support any high level languages such as SQL in DBMS and any other reservation optimization methods.

In paper [5], we propose a integrate data in the vehicular network points and in this sharing all the resources like which data we want only that details only sharing and only authorized peoples can access the information but unauthorized persons can't access the data. The disadvantages are not using in a decrypt method and also encrypt and not provides the security for the data.

III. PROPOSED WORK

Our proposed system gives a mapper function and reducer function in distributed manner to different data about the vehicles information. This function is used to study the interchange and recognize a particular about motor vehicle information according to its vehicle color and vehicle licence number or vehicle number. The watch facts are composed from the lofty quantity of definition cameras at street fractious and roadsides. In this we are using AMF that is used for extracts the pictures that is stored in the more number of cloudlets from the original video surveillance facts according to filter the based on particular filter the vehicle data like specified color and vehicle number.

The Figure 1 shows the system architecture, in this section first user enters with the user name and password, after successful login only we can go further execution of a system. Once a login is not possible then authentication with the system is also not available to customer or administrator. The administrator is a authenticate persons in our project for handling of entering a vehicle information and retrieving of a required vehicle details. These vehicle details are to be stored in the HDFS directory. HDFS file directory that provides a result with respect to an input and if not an input is not matched, and then it is failed to access results from HDFS file directory. An application will be run on Hadoop

software for execution purpose. After successful AMF execution it provides a result which vehicle details to users.

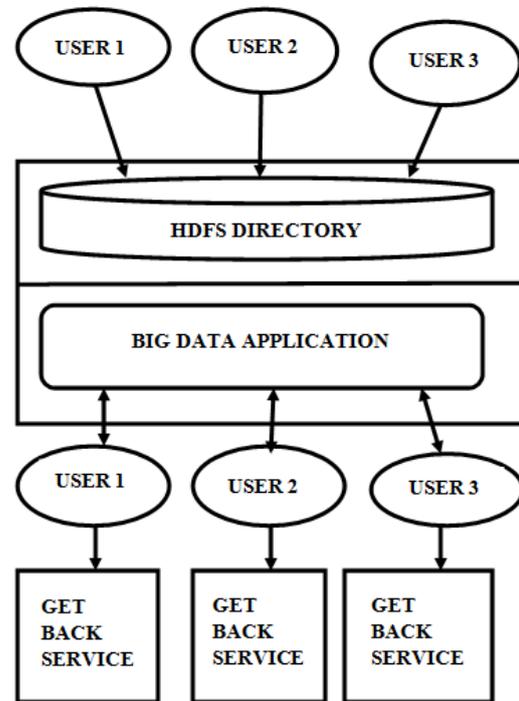


Figure 1: System Architecture

The various modules in the work are

A. CENTRAL SERVER MODULE

This module is used to save all the vehicle information gives response to user. In this what the user is entered a detail regarding a vehicle. For the response purpose to user when needed, these details regarding vehicle information. So that it is easy to get back when a user needs vehicle information with what details users stored in this module.

ALGORITHM

- Step 1:** Create class constructor with runnable constructor run().
- Step 2:** Upload file to server with details of vehicle.
- Step 3:** Receive details of file which contains vehicle information.
- Step 4:** Extract file content and activate the file content for view purpose with select all.
- Step 5:** If file content can't exists reload and start from step 2 to step 4.
- Step 6:** Access with object of class Central_Server class.
- Step 7:** Save file content if exists correctly.

B. CLOUDLET MODULE

The cloudlet module is interactive between the cloud server and user. The cloud server is having a data regarding vehicle details, these vehicle details are accessible when need to user. It helps to search vehicle details based on the vehicle color and vehicle number.

ALGORITHM

- Step 1:** Send a request to Central_Server for getting details of vehicle based on color.
- Step 2:** Create a cloudlet thread extending with Central_Server for getting vehicle information.
- Step 3:** Get users name and status history from Central_Server for adding components.
- Step 4:** Get request uploaded information of file to Central_Server.
- Step 5:** If exists filter it with matching color and display a vehicle details gathered from Central_Server.

C. SIGN IN AND REGISTER MODULE

This module is used to store user information and this module that allows us to with a username and password to go further in the system execution purpose. If a username and password are correct then only the sign in successfully, in case if it is not matching then it shown the user name or password incorrect like that that not allows go to access the data in a system.

The register module is refers to a vehicle registration like by entering a fields such as vehicle name, vehicle name and all the details which are mentioned in a registration. After entering all the details about a vehicle, then it allows to registration of vehicles.

ALGORITHM OF SIGN IN

- Step 1:** Create class constructor new sign in main function in java package geo-dispersed big data.
- Step 2:** Declaration and implementation of sign in module.
- Step 3:** Create label box and text box as user name and password.
- Step 4:** Perform action with ActionListener.
- Step 5:** If action performed go next.
- Step 6:** Else action not performed go to step 2.
- Step 7:** Display message box successful otherwise unsuccessful.

ALGORITHM OF REGISTER MODULE

- Step 1:** Create class in the package Geo-Dispersed big data registration constructor.
- Step 2:** Create a registration form fields like name, vehicle number etc with JLabel text field and set bound.
- Step 3:** Perform action on events if true set valid else go to step 2.
- Step 4:** Performed actions are true save with valid data to dataset.
- Step 5:** Otherwise clear dataset reinsert go through step 1 and step 2 until step4.

D. INSERT VEHICLE INFORMATION MODULE

This module which gives vehicle details depends on what customer enters in the field to access details of vehicle. It is get vehicle information based on the vehicle color and vehicle number. By both the options user can access details about a vehicle. We can get a good result from user entering a vehicle number rather than a color because many of the vehicles had a same color. So that by entering a vehicle number, which get in a quality of result expectation from a system.

The main advantages of the proposed system are :
 First, all input data is to be stored in the storage area like are full vehicle details are to be stored in the cloudlet. If we want to check the vehicle details about our required input data so that a input is to be stored first after that a mapping with a input data and stored data is to be done. Storage of input data in different node and stored total data in different node results easy to access in the aggregation using MapReduce because a workload is to be divided among the data. Dividing workload to distributed data manner gives low time taken to produce a result. Reducer is to be always deals with the required data of a dataset as a output gives to a customers as a result in the reduced form.

ALGORITHM FOR INSERT VEHICLE

- 1:** Create class constructor of type InsertVehicleInfo() in java package.
- Step 2:** Create boxes with JLabel and bounds with .setfield as vehicle information, name, color, date of registration and place of registration.
- Step 3:** If inserted values are valid set save, if not valid it display a message box like invalid data.
- Step 4:** If invalid data, select clear or go to step.
- Step 5:** Valid data save with insert vehicle information into dataset.
- Step 6:** connect with dataset from JSON.

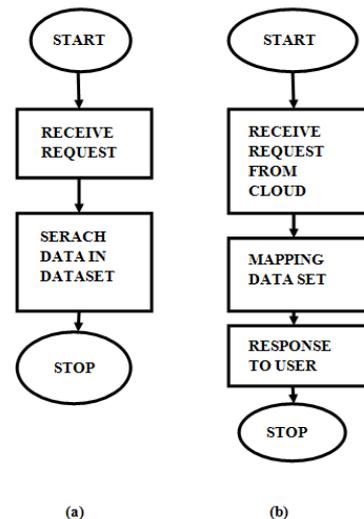


Figure 2: Data Flow Diagrams.

The Figure 2 (a) shows a service provider for the data flow diagram and Figure 2(b) shows the user side for the data flow diagram. In this receive request from the user to an application and try to match a input to an existing data, once a matching data is to be found then sends a result in an output form as reduced by main framework. In this request is sent to the HDFS file directory according to an input a matching output will be displayed in the reduced form.

IV. CONCLUSION

The challenges are in utilizing on Geo-Dispersed big data. We propose a new and flexible framework based on MapReduce to support first mapping an input of user and to display an output to user in a reduced form to user as a result. The proposed framework, Advance MapReduce Framework allows an input data to store first and then mapping with on existing data and input data is done. After comparing an input data with existing data if matches then a result will be produced in the reduce form. The AMF adopting uses multiple of inputs to user. An AMF is supports real time application, it produces a minimized reply time to user is achieved. For non valid instance applications, AMF makes a cut- off between interaction cost and reaction time.

V. REFERENCES

- [1] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Commun. ACM, vol. 51, no. 1, Jan. 2008, pp. 107-13.
- [2] D. Huang et al., "Secure Data Processing Framework for Mobile Cloud Computing," Proc. IEEE INFOCOM Wksp. Cloud Computing Shanghal, China, 2011, pp.614-18.
- [3] S. Shekhar et al., "Spatial Big-Data Challenges Intersecting Mobility and Cloud Computing.
- [4] K. H. Lee et al., "Parallel Data Processing with MapReduce: A survey," ACM SIGMOD Record, vol. 40, no. 4, Dec. 2011, pp. 11-20.
- [5] R. Yu et al., "Toward Cloud-Based Vehicular Networks with Efficient Resource Management," IEEE Network, vol. 27, no. 5, Sept.-Oct. 201, pp.48-55.