RAN METHOD THAT PERFORMS NESTED
SCALABLE NO SQL TRANSACTION AND DATA
ANALYTICS

D.S.C. Reddy
Professor & HOD, Dept. of CSE,
PBR Visvodaya Institute of Technology and Science, AP, India.

Kanamathareddy Reshma Reddy
M.Tech, Dept. of CSE,
PBR Visvodaya Institute of Technology and Science, AP, India.

ABSTRACT

Last decade records a large quantity of data transition occurring web. Today's world is connected by Social Networking sites that have immense size of databases. Currently, there's no utility existing for remodeling SQL databases into NoSQL databases. NoSQL could be a new trend; there's no software system or utilities that are designed for the specified conversion. Most of the utilities developed, were developed for changing one variety of SQL database to a different SQL database. To market Hybrid database design together with an electronic information service and NoSQL database, an information Adapter system is planned. It will affect database transformations employing a database synchronization mechanism. It focuses on sanctioning effective database Analytics on Scaled-out Object Storage Systems. There's implementation of a database. Analytics Layer on associate degree Object-based Storage Cluster to perform in-place Map-Reduce computation on existing data. It keeps the strengths of RDBMS as ACID properties and at an equivalent time by providing the advantages of NoSQL through a middle layer that keeps track of all running transactions and manages with alternative layers with co-occurring transactions. the info Adapter uses a SQL layer acceptive queries from application services, in order that original application remains intact. The info Adapter conjointly controls question flow throughout information transformations. Associate degree approach that performs computation on existing large-scale information in associate degree Object Storage system while not moving information anyplace and outcomes of this approach are analyzed. This define will deliver strict consistency while not poignant measurability and handiness of NoSQL Databases.

Keywords: SQL; NoSQL; data adapter; migration; dataanalytics; SVM

Cite this Article: D.S.C. Reddy, Kanamathareddy Reshma Reddy, RAN Method that Performs Nested Scalable no SQL Transaction and Data Analytics, International Journal of Advanced Research in Engineering and Technology (IJARET), 11 (1), 2020, pp 30-37.

http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=1

Electronic copy available at: https://ssrn.com/abstract=3534525
1. INTRODUCTION
To maintain the organization or business with success, information is needed. Structured command language (SQL) could be a distinctive programming language supposed for managing information control in an exceedingly electronic information service Management System (RDBMS), or in an exceedingly relative information Stream Management System (RDSMS) for stream process. huge information and Hybrid information systems are getting widespread as Cloud service blooms. NoSQL information bases are growing in quality for large Data applications. Most of the prevailing systems are supported electronic information service, however with the expansion of knowledge size, enterprise tends to handle huge information with NoSQL information for analysis or needs to induce quicker access on huge information. The info transformation mechanism for electronic information service to NoSQL information is needed once incorporating the initial system with NoSQL information. the appliance is forced to suspend and to attend for information synchronization thanks to the transformation processes. The transformations might take a protracted time if information is in giant scale. it's a essential issue for a few period of time, non-interrupting service like scientific analysis or internet applications. [1] Clusters and Cloud storage applications usement that job on large-scale information generally employ separate cipher and Storage Clusters, since the necessities of the cipher and Storage tiers are completely different from one another. However, the need to transfer immense amounts of knowledge from the Storage nodes to the cipher nodes so as to perform computation then to shift the results back to the Storage Cluster could be a significant issue of this design. Today, several storage systems store Petabytes of knowledge for varied applications, like climate modelling, astronomy, genetic science analysis etc., and therefore the quantity of knowledge keep within these systems is projected to succeed in Exabyte scale in the close to future. Therefore, moving huge amounts of information of knowledge between storage associate degreed cipher nodes isn't an economical manner of playacting computation on large-scale data any longer. In addition, storing information each at the storage and cipher sites will increase storage overhead and with information replicated multiple times at each sites for resiliency, this overhead becomes even worse. Moving information between storage and cipher nodes conjointly will increase the whole energy consumption and therefore the network load. Object-based Storage is a crucial effort during this respect and plenty of Scaled-Out Storage systems these days are supported the Object-based Storage abstraction.

Object-based Storage is another to the standard Block-based Storage. As a result, Object-based Storage will increase the interaction between the Storage system and therefore the end-user, simplifies the info management of a Storage system. [2] victimisation Object-based Storage options, the process applications in an exceedingly Cluster or Cloud application will take pleasure in the intelligence of the underlying storage system and eliminate information movement whereas sanctioning in-place Analytics Capabilities. Consequently, the Storage layer is scaled whereas the process layer remains light-weight.

2. RELATIONAL VS. NOSQL
The NoSQL models are completely different in their data-structure and the way they scale, the {needs} of those databases match completely different business needs. If a business wants sturdy transactional controls that guarantee information is up-to-date then, it's an honest argument for relative Databases that maintain strict ACID rules. If the business needs great deal of knowledge that don’t essentially got to be up-to-date and not all nodes within the Distributed system got to invariably maintain ACID rules than NoSQL is that the more sensible choice. If a business needs the most effective of each worlds, with ACID maintained
wherever required, associate degree a lot of potency in areas wherever this isn’t needed and an eventually consistent dataset is suitable, then each is used. [3]

3. A COMPARISON BETWEEN NOSQL DATABASES

MySQL was compared against these NoSQL Databases specifically, Cassandra, HBase, MongoDB, CouchDB and Redis. In addressing output, MySQL suffered once reading/writing over 7000 operations per second whereas the NoSQL implementations showed very little change; HBase being higher at writes, whereas prophetess showed to be a lot of read capable. Despite this MySQL and similar implementations retain sturdy read-oriented strength, however this is often solely true for smaller information sets. once the scale of the MySQL information grows overlarge there is performance penalties which will build the NoSQL implementations a lot of favourable. Prophetess and HBase retain higher write-optimizations, even in comparison at a smaller dataset size. however in comparison from price issue purpose of read, MongoDB proves to be price effective NoSQL information.

4. ARCHITECTURE FOR SCALED OUT OBJECT STORE

A typical Hadoop implementation is shown in Fig. 1. Storage Cluster stores the info on that a MapReduce application are going to be dead. cipher Cluster consists of nodes that type a Master-Slave relationship; wherever the Master node is answerable for sending jobs or I/O operations and watching the standing of Slave nodes. every Hadoop node has 2 processes running adore the Computation (MapReduce) and Storage (HDFS) layers severally. On Master nodes, the computation layer method could be a Job huntsman and therefore the Storage layer method could be a name node. On Slave nodes, the Computation layer method could be a Task huntsman and therefore the Storage layer method could be a information node. [2] the most important disadvantage of the everyday approach conferred in Fig. one is that they got to transfer information between the Storage and cipher Clusters before and when a MapReduce application. Considering the scale of knowledge being moved, the network interconnect is simply maxed-out and therefore the total energy consumption would increase as system resources are used for information transfer.

| S. No | Features | Cassandra | Couch DB | HBase | Mongo DB | REDIS |
|-------|----------|-----------|----------|-------|----------|-------|
| 1     | Database Model | Wide column Store | Wide column Store | Wide column Store | Key Value Store |
| 2     | License | Open Source | Open Source | Open Source | Open Source |
| 3     | Implementation Lang | Java | Erlang | Java | C++ | C |
| 4     | Typing | Yes | Yes | No | Yes | Partial |
| 5     | Secondary Indexes | Restricted | No | No | Yes | No |
| 6     | APIs & other access methods | CQL, Apache thrift API | RESTful HTTP API | Java API, RESTful HTTP API, thrift | Proprietary protocol using JSON | Proprietary protocol |
| 7     | Server-side scripts | No | JavaScript | Yes | JavaScript | Lua |
| 8     | Triggers | Yes | Realtime notifications | Yes | No | No |
| 9     | Partitioning methods | Sharding | Sharding | Sharding | Sharding |
| 10    | Replication methods | Selectable replication factor | Master-master, Master-Slave | Selectable replication factor | Master –Slave replication |

http://www.iaeme.com/IJARET/index.asp

Electronic copy available at: https://ssrn.com/abstract=3534525
The planned approach implements Hadoop on the cipher Cluster. By doing therefore, the network interconnect between separate Clusters is totally eliminated and therefore the cipher Cluster is really a Storage Cluster at an equivalent time. Ceph stores object information and it’s tied to the info node processes in Hadoop. As shown in Fig. 2, Ceph object store and Hadoop information node processes are tied along and there is multiple ways to integrate Hadoop with the underlying Ceph storage system. the foremost straight-forward approach would be to transfer object information directly from Ceph to HDFS. Another approach is victimisation Ceph because the backend storage of Hadoop and implementing Hadoop information management policies there.
5. PROPOSED SYSTEM

5.1. Theoretical Analysis

A Data Adapter system is planned to support the Hybrid information design together with an electronic information service (RDB) and NoSQL information. The info Adapter system is extremely modularized, superimposed between the appliance and databases. It's answerable for playacting queries from application and information transformations between databases at an equivalent time. The system provides a SQL interface parsing question statements to access each an electronic information service and a NoSQL information. A mechanism is planned to regulate the information transformation method and let the appliance to perform the queries whether or not target information (table) is being reworked or not. When information is reworked, it's planned to produce a patch mechanism to synchronize inconsistent tables.

Effective information Analytics is planned on Scaled-out Object Storage systems. victimization Object-based Storage options, the process applications in an exceedingly Cluster or Cloud application will take pleasure in the intelligence of the underlying Storage system and eliminate information movement whereas sanctioning in-place Analytics Capabilities. Consequently, the Storage layer is scaled whereas the process layer remains light-weight. The planned system can implement a middle layer which is able to support transactional properties for the shoppers like ACID and maintains the motives of NoSQL information like handiness and partitioning provided already by the NoSQL Database Engine. The center layer can rely solely on however we are able to get or update information on the information Engine. Therefore, with minor modifications, we are able to build it appropriate for several NoSQL information Engines. Also, we are able to management the center layer through turning it ON in things once dealings management is very important and turning it OFF in things once dealings management isn't important.

![Proposed Data Adapter](image)

**Figure 3** Proposed Data Adapter

5.2. Proposed System Flow

**Algorithm:** Proposed System Flow

**Input:** Relational big data

**Output:** NoSQL data and the big data analytics

- Relative information is chosen for process.
- Consequently information adapter helps in getting ready the required scripts for the chosen relative information.
RAN Method that Performs Nested Scalable no SQL Transaction and Data Analytics

- Hadoop associate degree Ceph configuration is finished for largescale information in an scaled out object storage system.
- Transformation from relative information to NoSQL information is finished by the info adapter.
- The reworked information is then prepared for the info analytics.
- Applicable information Analytics algorithmic rule is enforced.
- Results are generated for the analytics done.

5.3. Data Analytics

Data Analytics is analyzing method for information which has reviewing, filtering, converting, and molding information with the goal of encountering valuable data, recommending conclusions, and favoring call formulation. It will examine huge information sets to reveal secret patterns, unknown relationships, market trends, shopper preferences and accessorial valuable business data. information Analytics has varied aspects and methodologies, comprising various techniques through style of classes, in several businesses, science, and plenty of a lot of areas of influences. This on the market wide selection of information of knowledge of information analytics algorithms are applied on the reworked huge data. 1. SVM (Support Vector Machine): SVM could be a supervised algorithmic rule used for classification. A separating hyper plane is employed to outline SVM classifier. the choice surface is being sought for aloof from any information, so margin of the classifier is being calculated by taking distance between surface distance and therefore the nearest purpose. Here the SVM classifier is employed to classify the info that's being reworked into NoSQL, i.e., nonrelational information over Hadoop. For classifying information victimisation SVM over Hadoop, Apache driver was getting used.

5.4. Results

The take a look at setting is as below. A Linux machine illustrates MongoDB illustration. Machine has Intel i5 mainframe with eight GB RAM. The Fig. four displays a typical example table from MySQL that is to be reworked to MongoDB. a question was written to convert that MySQL table to JSON object. Than that JSON object was being foreign to MongoDB victimization mongoimport command in MongoDB. Fig. five shows usage of mongoimport command in MongoDB. to look at the reworked sample information table, notice command is employed that is displayed in Fig. 6. Fig. four Sample MySQL table

![Figure 4 Sample MySQL table](https://ssrn.com/abstract=3534525)
Figure 5 Illustration of mongoimport command

Figure 6 Transformed data in MongoDB

Figure 7 Data Analytics using SVM Algorithm over Hadoop
6. CONCLUSION

A flexible and extremely modularized information Adapter for Hybrid information system is planned. The info Adapter uses general SQL layer acceptive queries from application services, in order that original application doesn't got to modification the planning, the info Adapter conjointly controls question flow throughout information transformations. Associate degree approach that performs computation on existing large-scale information in associate degree Object Storage system while not moving information anyplace and outcomes of this approach are analysed. Several applications like bank application, transactions and real time systems need sturdy information consistency. This application will get sturdy information consistency if it's operated over RDBMS. There ar some limitations that RDBMS face like measurability. NoSQL databases offer solely weak consistency, which makes it not appropriate for applications that need sturdy consistency. This frame work will offer strict consistency while not poignent measurability and handiness of NoSQL Databases. In future approach any of the on the market information analytics algorithms is used on the massive information, counting on the relevance of the massive information.

REFERENCES

[1] Ying-Ti Liao, Jiazheng Zhou, Chia-Hung Lu, Shih-Chang Chen, Ching-Hsien Hsu, Wenguang Chen, Mon-Fong Jiang ,Yeh-Ching Chung, "Data Adapter for Querying and Transformation between SQL and NoSQL Database", Future Generation Computer Systems 65 (2016)111–121, © 2016 Elsevier Inc. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[2] Cengiz Karakoyunlu , John A. Chandy , Alma Riska, "Adding Data Analytics Capabilities to Scaled-out Object Store“, The Journal of Systems and Software 121 (2016) 16–27 , © 2015 Elsevier Inc.

[3] Ayman E. Lotfy , Ahmed I. Saleh, Haitham A. El-Ghareeb, Hesham A.Ali, “A Middle Layer Solution to Support ACID Properties for NoSQL databases”, Journal of King Saud University – Computer and Information Sciences (2016) 28, 133–145, © 2016 Elsevier Inc.

[4] Deka Ganesh Chandra," BASE Analysis of NoSQL Database", Future Generation Computer Systems 52 (2015) 13–21, © 2015 Elsevier Inc.

[5] Amitkumar Manekar, G. Pradeepini,"Cloud Based Big Data Analytics A Review", 2015 International Conference on Computational Intelligence and Communication Networks, © 2015 IEEE.