

Cloud Query Language for Cloud Databases

Mansaf Alam and Bashir Alam

Abstract- This paper deals with the cloud query language for manipulating the cloud data. The cloud definition and cloud data manipulation language are proposed as component of cloud query language. This paper is also focus on retrieval and manipulation cloud data dispersed on different data centre.

Keyword- Cloud Database, Engine, Retrieval, Manipulation, Cloud Query

I. Introduction

The cloud database management system is emerging right now. The cloud query will be used in cloud database management system. The data are spread over the network as a cloud. The fast growing rate of internet technology and data are major problem to manage it. The big problems are faced by the people around world to manage the data in cloud. The Claremont Report on Database Research [3] stated that Cloud Services can provide efficiencies for application providers, both by limiting up-front capital expenses, and by reducing the cost of ownership over time. Such services are typically hosted in a data center, using shared commodity hardware for computation and storage. In the Claremont report also discussed that a varied set of cloud services available today, including application services (salesforce.com), storage services (Amazon S3), compute services (Google App Engine, Amazon EC2) and data services (Amazon SimpleDB, Microsoft SQL Server Data Services, Google's Datastore). The Mike Hogan[4] have discussed that one of the benefits of cloud databases is that they shift much of the low-level DBA functions to experts who are managing the databases in a centralized manner for all of the users. However, tuning a shared-nothing database requires the coordinated involvement of both the DBA and the application programmer. This significantly

increases support costs. Shared-disk databases cleanly separate the functions of the DBA and the application developer, which is ideal for cloud databases. A globally robust method has been developed for retrieving cloud properties from high-resolution imager data during day and night, Patrick Minnis et. al. [5]. A break in the clouds: towards a cloud definition, L. Vaquero et. al.[8], discusses the concept of Cloud Computing to achieve a complete definition of what a Cloud is, using the main characteristics typically associated with this paradigm in the literature. Gook-Pil Roh et. al.[12] support new types of queries that search for matching trajectory patterns in a database. While such pattern matching queries have been successfully adopted for querying time series data, Qiuqing Lia, et. al.[14] proposes the concept of coordinating transaction. They describe that Coordinating transaction allows group members to access data cooperatively and also satisfies the requirement of dynamic change in cooperation to some extent. The Frank Munz[15] discussed Rackspace Cloud Server and Rackspace Cloud Files for Windows and Linux in his book. He has explained designing principles, designing for availability and designing for scalability. The cloud database is also discussed by him. Recently Toad [18] software is used for Cloud Databases, this software provides a SQL-based interface for generation of queries, migration, browsing, and editing the data, as well as creates reports and tables like SQL view. Toad is also helpful for loading and exporting data from cloud databases to various external data sources, like RDBMS. The community edition of Toad for Cloud Databases [17] is a free Eclipse plug-in that enables users to use SQL to manipulate the data in a cloud database. It is also use to join a cloud database to a relational database on the personal computer as well as copy various table from one cloud database to cloud database or between a cloud database and a relational database on the stand alone computer. Ken North [19], Cloud Databases: Connectivity and Platform Options has discussed the developing new applications or services, and the public, private, and hybrid cloud options for hosting big data and databases. He has also focused on APIs and Protocols, Clients ship queries over a computer network for processing by the database server and how the cloud

Mansaf Alam
Department of Computer Science
Jamia Millia Islamia, New Delhi, India

Bashir Alam
Department of Computer Engineering
Jamia Millia Islamia, New Delhi, India

database will support various available applications. Snia[20], Cloud Data Management Interface, mentioned private cloud for delivery of SaaS, PaaS, IaaS, and DaaS to a restricted set of customers, around one organization and Public Cloud to delivery of SaaS, PaaS, IaaS, and DaaS to a relatively authorized group of clients. In Cloud Data Management Interface [20] a network of a virtualized programming environment is used. This is consisting of an application deployment stack based on a virtual computing environment for delivery over the network.

II. Cloud Definition Language

In this paper cloud definition language is proposed as a component of cloud query language. The cloud definition language defines the structure of cloud data for storage and management. In RDBMS the Data Definition Language used to define the structure of data. In OODBMS ODL is used to define the structure to data as in object form. Similarly in this paper a CDL is used define the structure to store the cloud data. According to Cristina B̄asescu et. al. [7] a suitable storage system has to be able to scale to a large number of cameras, each of them concurrently writing huge amounts of data to different files. To leverage these needs for storage capacity, the data can be hosted directly on the Cloud. David Chappell [9] state that the Cloud platforms are changing how we should build and run applications. They are also changing how we store and access data.

It should be assume that another client wants to access same interface for the other application’s data. The client needs to retrieve customer data by geographic area, however, rather than by customer number. To allow this, the creator of the application’s interface likely adds a new operation, perhaps called GetCustomerByGeo. When yet another client appears, it might well want to retrieve customer data in some other way, requiring still more changes to the interface. The appropriate data center is located automatically where data is to be defined as a cloud data and also space is created for storing the cloud data. Algorithm to define the storage structure of data in cloud:

- i. Find the appropriate data center out of n data centre with help of internet
- ii. If appropriate data center is found then locks the space in data center while define the structure of data.
- iii. Use the create command to create the data structure to store the particular information.
- iv. Insert the information as per structure of data defined in step ii
- v. Release the data centre so that other user can use that data center.

The users can interact with data using data source name (DSN). The cloud data are bound with DSN as shown in the figure-1. The DS lies between users and

cloud data. Any user wants to interact with the particular data. The DSN will establish the connection first with the data. After connection is established the user can manage the storage structure of cloud data.

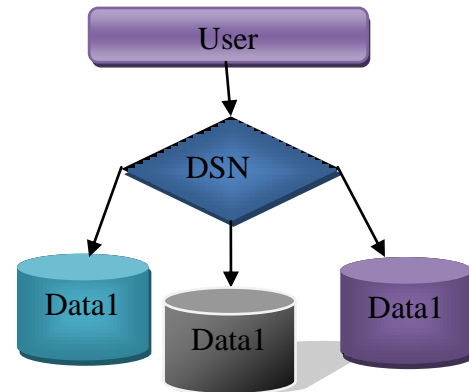


Figure-1: Users-data-Interaction

Let f(u) is define as a set of users and it is computed as

$$f(u) = \sum_{i=1}^n (u_i)$$

The f(d) is defining as set of data and computed as

$$f(d) = \sum_{i=1}^n (d_i)$$

The DSN is defined as f(c) is for connection between data and users. The f(c) is computed as cross product of f(u) and f(d).

The f(c) = f(u) x f(d)

$$f(c) = \sum_{i=1}^n (u_i) \times \sum_{i=1}^n (d_i)$$

where user u and data d running from 1 to n.

III. Cloud Data Manipulation Language

The cloud data manipulation language is proposed in this paper as another component of cloud data query. The DML is used to manipulate the data in relational databases (RDBMS) and OML is used to manipulate the data in object form in Object Oriented Databases (OODBMS). Similarly here the CML is used here to manipulate the Cloud data in Cloud Database Management system. Here the data are spread over the internet in the cloud form so it is very difficult to manage the data using DML and OML as in RDBMS and OODBMS. Y. Cao et. al.[11] introduce a data access optimizer within the data manipulator component, which dynamically chooses the best data access scheme for a specific query, relying on the statistics stored in the meta-data catalog. They also describe that the meta-data catalog provides schema storage, update and retrieval, statistics

storage, update and inquiry, and runtime statistics collected via daemon processes.

Here a new concept is introduced to manipulate the data in different data centre connected with the internet. The Cloud Manipulation concept is as consider the D is a set of data centre connected via internet. D define as $D=\{d_1,d_2,d_3,-----d_n\}$ where $d_1,d_2, d_3,-----$ are data centers.

The data center is defined $f(d)$, the $f(d)$ is computed as

$$f(d) = \sum_{i=1}^n (d_i)$$

Where $i=0,1,2,---n$ are data in tiny data centers.

Let define d_1 is a set of data at first data centre $d_1=\{x_1,x_2,x_3,-----x_n\}$ where x_1,x_2,x_3 are different types of data at data centre one. The definition can be represent mathematically as

$$f(d_1)=\sum_{i=1}^n (x_i). \text{ And similarly } d_2=\{y_1,y_2,y_3,----y_n\} \text{ where}$$

$y_1, y_2, y_3,-----$ are the data at data centre two and so on.

$$d_3= \{z_1,z_2,z_3,-----z_n\}$$

:
:
 $d_n=\{n \text{ data}\}$ the mathematically representation of the above data in data centers are as follows:

$$f(d_2)=\sum_{i=1}^n (y_i)$$

where $i=1, 2, 3,-----n$

$$f(d_3)=\sum_{i=1}^n (z_i)$$

where $i=1,2,3,-----n$

:
:

$$f(d_n)=\sum_{i=1}^n (m_i)$$

where $i=1, 2, 3,-----n$

Parameters	Definition
D	Data Centre domain
d	Different data center
x, y, z	Data at different data centre
cq	Cloud query
m	Data in n data center

Table-1

We define the data center data center $d_1,d_2,d_3,-----d_n$ are involved in cloud query, $f(cq)$ is computed as

$$\frac{[f(d_1)x(1 - f(d_1))] + [f(d_2)xf(1 - f(d_2))] - - [(f(d_n)x(1 - f(d_n))]}{f(d_n)} \text{ or}$$

$$f(cq)=\frac{\sum_{i=1}^n x_i \times \left(1 - \sum_{i=1}^n x_i\right) \oplus \sum_{i=1}^n y_i \times \left(1 - \sum_{i=1}^n y_i\right) \oplus \dots \sum_{i=1}^n m_i \times \left(1 - \sum_{i=1}^n m_i\right)}{\sum_{i=1}^n m_i}$$

where $\sum_{i=1}^n x_i$ is probability of data x to be found in data

center one and $\left(1 - \sum_{i=1}^n x_i\right)$ is not found. Similarly for

available of data m in data center n probability is

$$\sum_{i=1}^n m_i \text{ and for not found is } \left(1 - \sum_{i=1}^n m_i\right).$$

IV. Cloud Query Engine

The movement of cloud data is shown in the following diagram of cloud query language. The actual paths for data movement are shown with straight line and virtual movement of data is shown by dotted line in cloud database query engine. The cloud database query engine is shown in figure-2. Mohammad Farhan Husain et. al.[13] defined the query plan generation problem, and they show the best generating query plan for the ideal model and as well as for practical mode. This cloud query engine work for secured cloud as well as unsecured cloud. In the case secured cloud, it will take authentication from secured cloud for manipulation of data in cloud. Christophe Gueret et. al. [21] have discussed the key-value stores a robust query engine for approximate answers. They have test the scalability of the approach scaling to over 3 billion triples for complex queries. Their query approach does not rely on any low-level primitives. But their query approach can be used with any, possibly already deployed, indexing infrastructure. Minpeng Zhu and Tore Risch [22] have developed a system called BigIntegrator to enable general queries used to combine data in cloud-based data stores with relational databases. They have also presented that the design and working principle of the system. They have explained the scenario of querying data from both kinds of data sources with the help of an illustration. This system is general and extensible to integrate data from various types of data sources. Qiming Chen and Meichun Hsu [23] they have implemented operation event pipe as special kind of

continuous query that is called event pipe query. The cloud query engine is shown in this paper will work for any cloud data spread over the internet.

response to data load in the cloud environment. The cloud query engine use to create, retrieval, updating and deleting different data elements from the Cloud of various data centers. To retrieve the data from cloud a matrix of data centre and data will be used. This matrix will be used to store the information about data in cloud. For example there are four data centers namely A, B, C and D are respectively. The data in data center A are a_1, a_2, a_3, a_4 . This can be defined mathematically as

$$A = \{ a_1, a_2, a_3, a_4 \}$$

Similarly

$$B = \{ b_1, b_2, b_3, b_4 \}$$

$$C = \{ c_1, c_2, c_3, c_4 \} \text{ and}$$

$$D = \{ d_1, d_2, d_3, d_4 \}$$

The matrix is defined for the above data is as follow:

	A	B	C	D
A	a_1	a_2	a_3	a_4
B	b_1	b_2	b_3	b_4
C	c_1	c_2	c_3	c_4
D	d_1	d_2	d_3	d_4

Suppose we want to retrieve the data b_2 , it can be search in $A \times A, A \times B, A \times C$ and $A \times D$, if desired data is found, send the data to the user else see the data in $B \times A, B \times B, B \times C$, and $B \times D$, continue this till the end of data centre.

Based on the searching of data in the above matrix make the success and failure matrix as follow:

	A	B	C	D
A	0	0	0	0
B	0	1	0	0
C	0	0	0	0
D	0	0	0	0

Assume one (1) for data is found in the data centre and zero (0) for data is not found in the data centre.

If we desire to retrieve a_1, d_3, c_3 the success and failure matrix will be like as follow:

	A	B	C	D
A	1	0	0	0
B	0	1	0	0
C	0	0	1	0
D	0	0	1	0

The above matrix store the information of data is found or not in the cloud.

Similarly we can make the matrix for creation of cloud of data in data centers, Updating and deleting the cloud data in the data centers. Daniel J. Abadi [17] Data Management in the Cloud, They refer the applications for query a data store for use in business planning, problem

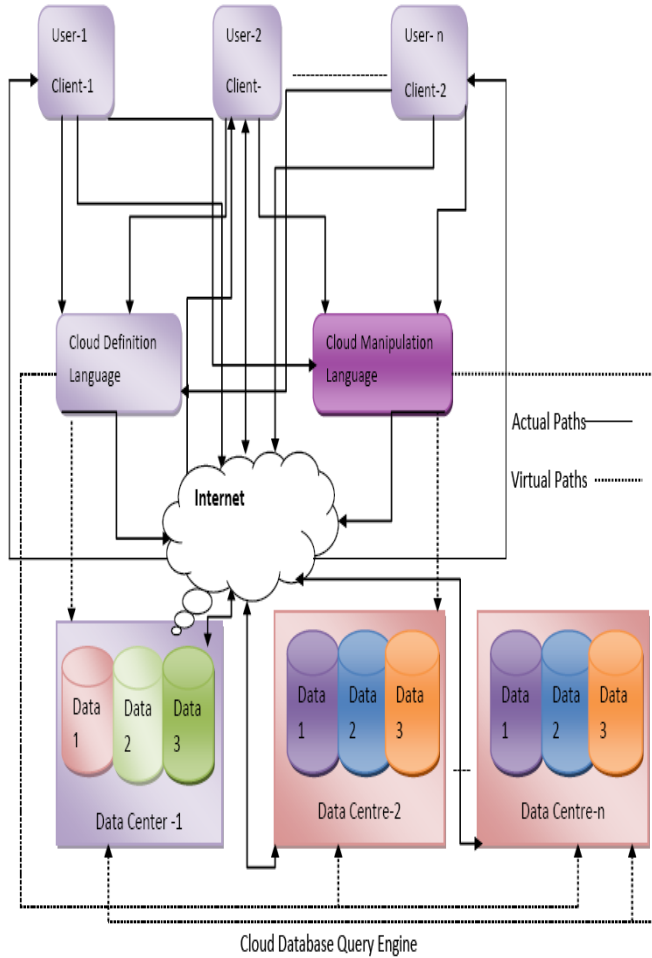


Figure-2

In this figure there are n data centre connected with internet. The data center mathematically represented as

$$\sum_{i=1}^n m_i \text{ where } m \text{ is data in } n \text{ data center and } i \text{ running from}$$

1 to n. The cloud definition language and cloud manipulation language are shown between internet and client (users). The cloud definition language and cloud manipulation language are also directly connected with data centers.

Brian D O'Connor et. al.[6] present the SeqWare Query Engine which has been created using modern cloud computing technologies and designed to support databasing information from thousands of genomes. Carlo Curino et. al. [10] state that Relational Cloud uses existing unmodified DBMS engines as the back-end query processing and storage nodes. They also describe that each back-end node runs a single database server. The set of back-end machines can change dynamically in

solving, and decision support and historical data along with data from multiple operational databases were involved in their analysis.

V. Conclusion

This paper shows how to make the query to access the cloud data, managed by cloud database management system. The data are distributed in cloud and the structured are complex so the management is very difficult. The cloud data are unstructured in nature. In this paper a cloud database query is designed for managing the cloud data in cloud database management system. The cloud data definition and cloud manipulation language concepts are described in the paper.

References

1. Robin Bloor, What is A Cloud Database? The Bloor Group, 2011.
2. Mike Hogan, Database Virtualization and the Cloud, Scale DB Inc., 2009
3. Rakesh Agrawal, et. al., The Claremont Report on Database Research, Resort in Berkeley, California, 2008.
4. Mike Hogan, Cloud Computing & Databases (How databases can meet the demands of cloud computing), Scale DB Inc., November 14, 2008.
5. Patrick Minnis et. al., A Global Cloud Database from VIRS and MODIS for CERES, SPIE 3rd International Asia-Pacific Environmental Remote Sensing Symposium 2002: Remote Sensing of the Atmosphere, Ocean, Environment, and Space Hangzhou, China October 23-27, 2002.
6. Brian D O'Connor,1 Barry Merriman,2 and Stanley F Nelson, SeqWare Query Engine: storing and searching sequence data in the cloud, BOSC) 2010, 9–10 July 2010, Boston, MA, USA
7. Cristina Basescu and Alexandra Carpen-Amarie, Managing Data Access on Clouds: A Generic Framework for Enforcing Security Policies, International Conference on Advanced Information Networking and Applications, IEEE, Computer Society, Washington, DC, USA, Page: 459-466, 2011.
8. L. Vaquero, L. Rodero-Merino, J. Caceres et al., “A break in the clouds: towards a cloud definition,” SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50–55, 2009.
9. David Chappell, Introducing Odata data Access for the web, the cloud, mobile devices, and more, Chappell & Associates, Sponsored by Microsoft Corporation, San Francisco, California, USA, 2011.
10. Carlo Curino, Evan P. C. Jones, Raluca Ada Popa, Nirmesh Malviya, Eugene Wu, Sam Madden, Hari Balakrishnan, Nickolai Zeldovich, Relational Cloud: A Database-as-a-Service for the Cloud, CIDR 2011, January 9-12, Asilomar, California, USA, 2011.
11. Y. Cao, C. Chen, F. Guo, D. Jiang, Y. Lin, B. C. Ooi, H. T. Vo, S. Wu, and Q. Xu, A cloud data storage system for supporting both oltp and olap. Technical Report, National University of Singapore, School of Computing. TRA8/10, 2010.
12. Gook-Pil Roh, Jong-Won Roh, Seung-Won Hwang, and Byoung-Kee Yi, Supporting Pattern-Matching Queries over Trajectories on Road Networks, IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 11, November 2011.
13. Mohammad Farhan Husain, James McGlothlin, Mohammad Mehedy Masud, Latifur R. Khan, and Bhawani Thuraisingham, Heuristics-Based Query Processing for Large RDF Graphs using Cloud Computing, IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 9, September 2011.
14. Qiuqing Lia, Hong Xu, and Ting Long Liu, Coordinating Transaction Model in CDBMS, The 7th International Conference on Digital Object Identifier, IEEE, Page(s): 421 - 425, Brazil, 2002.
15. Frank Munz, Oracle on Amazon Web Services (AWS):Middleware and Cloud Computing, Munz & more publishing, January 2011/First Edition.
16. Daniel J. Abadi. Data Management in the Cloud: Limitations and Opportunities, USA, IEEE 2009.
17. Andy Grant, Toad for Cloud Databases - Community Edition , Toad World, Created on: Nov 23, 2010.
18. <http://www.quest.com/toad-for-cloud-databases>.
19. Ken North, Cloud Databases: Connectivity and Platform Options, Dr. Dobb's the world of software development, August 16, 2011.
20. Snia, Cloud Data Management Interface, Snia Advancing Storage & Information Technology, April 12,2010, http://www.snia.org/sites/default/files/CDMI_SNIA_Architecture_v1.0.pdf
21. Christophe Guéret, Spyros Kotoulas, Paul Groth, TripleCloud: An infrastructure for exploratory querying over Web-scale RDF data, 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, IEEE, 2011.
22. Minpeng Zhu and Tore Risch, Querying Combined Cloud-Based and Relational Databases,2011 International Conference on Cloud and Service computing, Hong Kong,china,pp.330-335, IEEE, 2011.
23. Qiming Chen and Meichun Hsu. Continuous Access to Cloud Event Services with Event Pipe Queries, Lecture Notes in Computer Science (LNCS), 7045, pp-525-541, Springer, Verlag Berlin Hiedelberg, 2011.