Cloud Database Management System Architecture

Mansaf Alam and Kashish Ara Shakil

Abstract*— One of the major challenges for the IT companies today is how to manage large growing volumes of data and how to produce a quality driven software product ensuring optimal utilization of resources with minimum cost. The database management system is a software system i.e. a set of programs that provides its users with processes for defining, building, manipulating and sharing databases amongst the users and applications. A cloud data base management system is a database management system for management of cloud data and provides delivery of computing as services rather than as product. In this paper we have proposed an architecture for management of data in cloud termed as "Cloud Database Management System Architecture". The cloud database management system provides an approach for management of cloud data. The cloud data are spread over the internet and are stored to a remote server managed by a third party. Hence, the cloud data management is a major issue which needs to be catered to. A well-defined architecture is thus required to manage the cloud data, available at a remote location. In this work an architectural model for cloud database management system has been developed .This architecture is based on the three schema architecture for data base management system and three level object oriented database management system architecture.

Keywords*— Three Schema architecture, cloud system architecture, green computing, three level object oriented database architecture

I. Introduction

The major concern for any organization today is how to manage the ever increasing huge volumes of data and to deliver more and more enhanced services with reduced cost. The need of the hour is management of these huge volumes of data along with ensuring scalability, availability and reliability .This concern is what acted as the stepping stone for development of cloud.

Cloud computing enables IT resources such as the developers to concentrate more on the core issues like development of product, rather than worrying about secondary issues like availability of servers, storage space etc. It enables cloud users to use unlimited computing powers by renting more and more resources via cloud computing.

Mansaf Alam
Jamia Millia Islamia
India

Kashish Ara Shakil
Jamia Millia Islamia
India

Cloud computing is a very promising technology for the future. Cloud computing can cause a major cut down in the marketing time by taking charge of provisioning of resources such as servers, hardware or any other computing resources. It can lead to cost reductions as it employs the use of pay per use and also provides a better utilization of resources. Thus, cloud Computing is also forecasted as a Green computing technology. Cloud computing also promises (virtually) an infinite scalability along with flexibility [3].The probable benefits of cloud computing are astounding. Therefore, in order to attain these benefits it is required that each aspect of cloud platform must support the key design principles of cloud model. Some of the key design principles are dynamic scalability, availability, ability to allocate and reallocate resources. But, a majority of database servers are not able to meet these requisites [11]. Hence, requirements for a database in cloud are very different from traditional databases, as the cloud environment is very unpredictable [9] .Every database in cloud must be highly available and reliable. Attainment of scalability is also a very complex process. Also, it is very difficult to maintain distributed multiple copies of database at different locations. Thus database in cloud needs to be accessed and managed in a distinguished manner i.e. a proper framework for access and management of cloud data are essential.

In this paper we have proposed an architecture for cloud database management which aims at addressing these requirements of cloud databases.

II. Related Work

Our proposed approach "Cloud database management system architecture", bears a close resemblance to the three schema architecture for database management [8] and Three Level Object Oriented Database Architecture Based on Virtual Updatable Views [7].

A. *Three Schema Architecture*

The Three Schema architecture is a representative architecture of database systems. It provides support for multiple user views and program data independence with the aim to separate the user applications and physical database. In this architecture three levels are defined Fig. 1 [6]

1) *Internal level*: this level is composed of the internal schema, which describes the actual physical storage of data.

2) *Conceptual level*: this level is composed of the conceptual schema and describes the structure of whole database for a community of users. It hides details of data storage at physical level and concentrates on describing entities, relationships, user

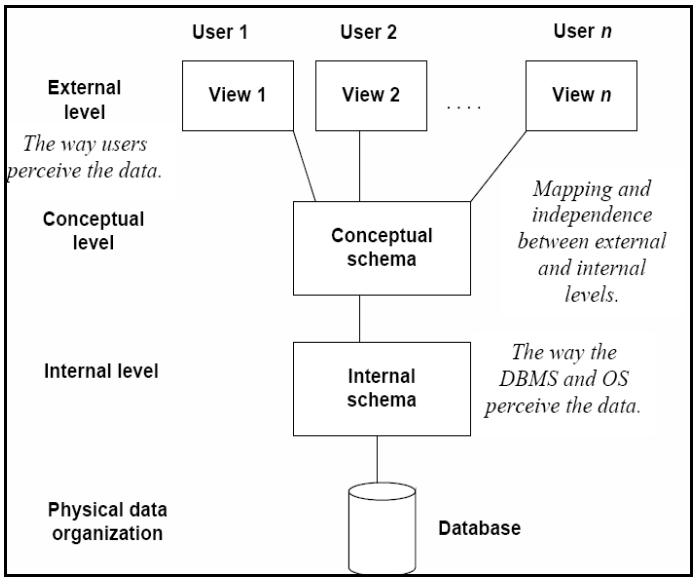

Figure 1. Three Schema architecture

operations and other constraints.

3) *External level*: This level provides an external view to end users i.e. it provides the end users with that part of database in which they are interested and hides other low level details

The above three schema architecture is well suited to the needs of a relational database environment but not for a cloud environment as a cloud environment requires access by many kinds of users having different service requirements therefore, it requires a greater level of customization. Also cloud environments have stringent security and privacy requirements which cannot be fulfilled with this architecture.

B. *Three-Level Object-Oriented Database Architecture*

Three level object oriented database architecture [7] is architecture for object oriented database access and management, is based on updatable views which provides a mapping of stored objects onto virtual objects. In this architecture the middle layer is referred as a DBMS managed middle layer. Its features are that it is transparent, provides users with the ease of management and modifications .This architecture defines the following user roles and responsibilities.Fig.2

Database programmer is responsible for creating internal and conceptual schema of the data based upon previously created design, as per the business requirements.

Database administrator is responsible for defining external schema for particular users. She/he creates updateable views which are built upon data store.

Application programmer is a database user, who uses the database and is well versed with interfaces of views provided by the database administrator.

Thus, the above three participants work together to provide management of user privileges along with directing database application development. Hence the Three level object oriented database architecture is well suited for object oriented database.

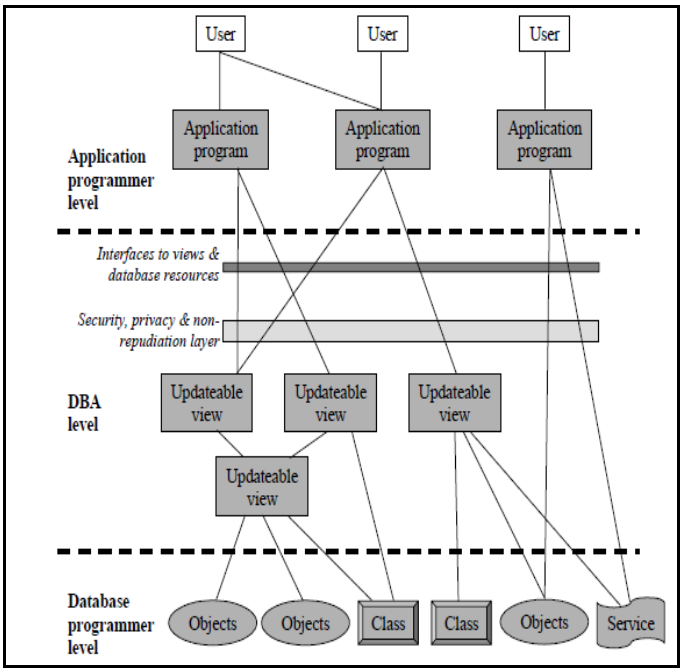

Figure 2.A conceptual view of three levels Object Oriented Architecture

Since we know that the computing environment for a cloud based system is very different from traditional database environments and for object-oriented and XML-oriented environments with emphasis on user data privacy, security, scalability elasticity, availability of resources etc. Different cloud clients may need different types of customized services and access controls as per their needs and requirements. Thus, in order to overcome the limitations of applicability of three schema architecture for database and Three-Level Object-Oriented Database Architecture we have proposed an architecture for cloud data access which has physical, logical and technical advantages over the existing architecture for database access.

Following are some of the major issues catered in our proposed approach:

- Customization: Since cloud data are to be used by a variety of different clients with different resource requirements. Each user of a cloud will have resources available via cloud as per their requirements.
- Conceptual modeling: As a cloud data will be used by clients with different types of requirements. Therefore each user will have a different view of cloud data.
- Information hiding: A cloud user should be constrained to use only the part of service which he/she has requested.

- Security, privacy and cloud data access: cloud data base administrators/servers should have flexible facilities to grant and access resources to the users according to the access rights granted to them.
- Orchestration/choreography of services: it's extremely important in cloud computing as it is based on scaling and pay per use, connecting and automating of work flows and hence cloud services must be orchestrated at every layer.

III. Cloud Users

Cloud users play a very vital role in a cloud computing application. It's the cloud users whose data are ultimately managed by the cloud service providers. Mladen A. Vouk[4] has given a hierarchy of the different cloud users .The cloud user hierarchy consists of four users:

A. *Cloud Infrastructure developers:*

These are developers who are experts in specialized areas such as networks, computational hardware, storage, operating system imaging etc and are responsible for development, maintenance and administration of cloud framework. They are also responsible for hiding lower level details from its clients

B. *Service Authors:*

These are developers of base line images and services, which might be used directly or integrated into other services. This enables cloud users who want to use image creation tools and service management tools to concentrate on their development prospects rather than the details of cloud infrastructure.

C. *Service integration and provisioning:*

Service integration and provisioning experts are responsible for creation of composite solutions required by an end user. They create new customized services for the clients by making updating the existing services and images.

D. *End Users:*

They are the most important users of cloud services .They are the users who ultimately use the cloud services provides by the cloud providers. They require that the services made available to them must be reliable, secure, easy to use and scalable.

IV. Cloud Database

Database is an organized collection of data, and is the heart and soul of any information system. Cloud infrastructure consists of huge volumes of data which might be shared amongst multiple tenants .Thus, data management in particular is an essential aspect for storage in cloud. The data are distributed in cloud across multiple locations and might contain certain privilege and authentic information. Therefore it's very important to ensure that data consistency, scalability and security are maintained. In order to address these issues and several other critical issues regarding data, a data base management system for cloud data is imperative.

In cloud two primary DBMS architectures are used shared nothing and shared disk. Shared nothing is a distributed computing architecture in which each node is self-sufficient and is independent of any other node i.e. each node in shared nothing architecture has its own memory and disk storage and does not share it with any other node. There is no point of contention in the nodes.

Shared disk architecture is a computing architecture in which each node has its own memory but they share disk storage space. It actually partitions the data such that each database server processes and maintains its own piece of database [5].

In this paper we will first discuss shared nothing and shared disk architectures, their limitations and then propose our cloud data base management system architecture.

A. *Shared nothing*

In shared nothing architecture each node has a memory as well as storage space of its own. These nodes communicate to each other by message passing through an interconnected network. In a clustered system only one resource can be accessed at a time ,during failure resource ownership can be transferred to any other resource connected on the network[5].Shared nothing has a virtually infinite scaling capability as each of the node is isolated from other. This approach is suitable for applications with heavy data update requirements. Shared nothing does not provide inherent data consistency like shared disk as each node has its own set of memory therefore transactions must be configured in order to ensure consistency of data. Since, in shared nothing architecture each node has its own memory it has problem of load balancing as load of one server can't be shared with other servers.

B. *Shared Disk*

In shared disk architecture all the connected nodes share the same storage disk space. It's particularly suitable for applications in which distributing or partitioning of workload is very difficult. Shared disk is based on message or token passing each node communicates with every other node by passing messages.

Shared disk cannot scale up as well as shared nothing architecture but provides an advantage of dynamic load balancing which is the driving factor that accommodates temporal and evolutionary changes in usage patterns [5].Another advantage of shared disk is that it maintains consistency of data as there is only one copy of data shared by all the nodes.

Therefore we can conclude that both shared nothing and shared disk architectures can be used in cloud depending upon the requirements of the organizations, developers and customers. But it is necessary to have a separate architecture for cloud data base itself in order to overcome the limitations of shared disk and shared nothing architectures. We will now discuss our proposed approach "Cloud database management

architecture", which is based on three tier architecture of DBMS and three level object oriented database architecture.

v. Cloud Database Management System Architecture

Now, we will discuss our proposed approach for cloud database access and management which is based on three schema architecture and three level object oriented database architecture [7] and the different roles at each of the levels. It provides foundation for three level database architecture. Cloud database management system architecture represents data in three levels Data Center Level, Cloud Service provider level and Client Level: Fig. 3.

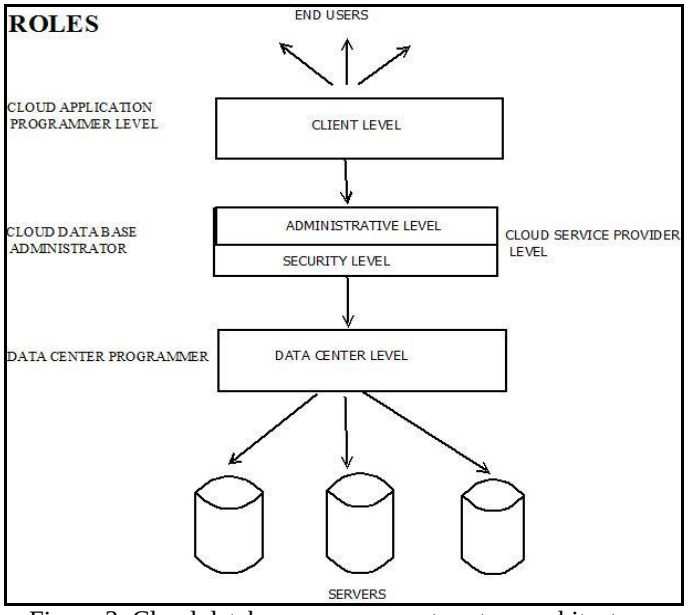

Figure 3. Cloud database management system architecture

A. *Data Center level*

This level represents the actual physical storage location of data in cloud and consists of several servers catering to the needs of cloud users Since we know in a cloud infrastructure data are stored at the data center .This level is a representative of a cloud database. In cloud computing data are stored as a virtualized pool of storage. Cloud providers operate these data centers as per the requirements of the clients. The data center operators provide the cloud users with a virtualized illusion of the resources according to their requirements, and expose them with storage pools where the customers can store their files and data. The cloud algebra [12] is used to manipulate the data available in data center.

B. *Cloud Service provider level*

At this level building and management of cloud applications is done. This is a middleware level and consists of several distributed servers catering to the needs of cloud users. It is these servers that are responsible for providing all the facilities promised by a cloud service provider to its users. It ensures the availability of cloud data at all times, provides facility of multitenancy, on demand self-service, elasticity and various other characteristics of a cloud. It adds the feature of data abstraction in a cloud .It hides the details of data storage at the datacenter level and makes the underlying software and database transparent to its client users. It provides customized cloud facilities as per the customer's requirements. It's this part which makes decisions regarding access rights of users etc.

The cloud service provider level can in turn be divided into two sublevels on the basis of functionalities at each of these levels. This level is subdivided as administrative level and security level:

1) Administrative level

At this level database instances are controlled using a service API. This API is available to the end users and enables database users to scale and maintain their database instances. For example, the Amazon Relational Database Service's service API allows creation and deletion of database instances. It also has equipped with the power of modifying resources available to database instances, creating a backup of data and also restoring the database instance in case of failure.

It is responsible for providing scalability of resources by adding more resources as per the user requirements. Scalability can either be provided automatically or users may have the privilege to scale up or down as per their requirements through an API.

High availability of database which is a commitment of cloud service providers is also taken care of at this level.

2) Security level

This level is responsible for providing security control to the cloud user's data and to assure continuous correctness of their stored data without any local copies. It ensures end to end data confidentiality, prevents data loss. User authentication, data encryption, intrusion detection are some of the measures adopted to ensure cloud security.

C. *Client level*

This level consists of cloud users or client computers. It is the most visible level to the clouds end users[10].At this level cloud users have the view of that part of cloud database that a particular user is interested while all the other details of cloud service provider level and data center level remain hidden from the end users. At this level, data center operators provide the cloud users with a virtualized illusion of the resources as per their requirements, and expose them with storage pools where the customers can store their files and data. End users of a cloud are client computers that access the cloud data. Client of a cloud infrastructure is not necessarily a computer but it can also be any other computing device such as web browser, mobile apps, mobile phone, tablet etc. End users make use of cloud computing infrastructure on the basis of pay per cost model thereby reducing the overall costs of developing their applications.

VI. Roles In Cloud Database Management System Architecture

The main aim of a database management system is that they provide a clear separation between a user application and actual physical data base along with ease of management and modifications. Therefore in order to ensure a systematic access of cloud data base for cloud users and for providing all the facilities of cloud database management system it is vital to define clearly the roles at each level in our cloud database management system along with the responsibilities .The following are the roles and responsibilities:

A. *Cloud Application Programmer*

Cloud Application Programmer is the actual database user, who uses the database as service provided to him or her by the cloud service provider and creates the database schema. The application programmer operates at the client level, accesses the cloud databases via an API provide by the client service provider as per his /her requirements. The programmer can perform tasks on the cloud database similar to a tasks performed on traditional databases like data retrieval, manipulation, modification, deletion and storage. However the security of this data is handled by the database administrator.

B. *Cloud Database Administrator*

Cloud database administrator operates at the client service provider level of a cloud database management system. They are the driving force behind the entire cloud database management system as they are responsible for providing database as a service to the cloud users .They provide customized external schemata for clients as per their requirements i.e. providing cloud users with the part of database service they are interested in, hiding all the data center level information. They are the security inspectors and in charge of data authentication, granting privileges and access rights to users. Scalability and availability of cloud database instances is also handled by the cloud database administrators.

C. *Data Center Programmer*

Data center programmers/operators operate at the data center level. They are in charge of storage of cloud data across several servers and for ensuring integrity of data. They define the behavior of the stored cloud data. Thus, our proposed well defined cloud database management system architecture has the following advantages over traditional DBMS:

a) User control over provisioning and management of data. It enables users to manage and store data themselves.

b) Helps in storage of files in different format's as per the user requirements like text files, image files, xml files etc.

c) It provides cloud users to perform all the tasks which can be performed on a normal DBMS like data manipulation, data storage, data retrieval etc.

d) It is cheaper than the traditional dbms.

e) It is more accessible.

VII. Conclusion

Three Schema Architecture and Three-Level Object-Oriented Database Architecture are two well defined architectures for relational dbms and object oriented data respectively. In cloud, data are distributed over the cloud across several distributed servers and cloud database has certain special requirements like scalability, availability along with stringent security and user authentication needs. As, these requirements cannot be fulfilled with the existing architectures for relational and object oriented data thus we proposed our architecture for cloud data called Cloud Database Management System Architecture.

Our, architecture is based on ANSI/SPARC three schema architecture and three level object oriented architecture. We have defined cloud database to be organized into a hierarchy of three levels cloud data center level, cloud service provider level and client level. We have also identified roles defined at each level i.e. roles of Cloud Application Programmer, Cloud database administrator and Data center programmer. Several related issues like how the security measures are adopted in database and other issues like administration have not been fully voiced here and are potential topics of research.

For our future research work we have planned to work on the cloud service provider level and perform a deep study of administrative and security levels, and aims at addressing the issues with the existing clouds security and issues with administration as well.

References

[1] Gerard Conway and Edward Curry, "Managing Cloud Computing: A life cycle approach", 2nd International Conference on Cloud Computing and Services Science (CLOSER 2012), 2012, pp. 198-207.

[2] Ashraf Aboulnaga, Kenneth Salem, Ahmed A. Soror, Umar Farooq Minhas,Peter Kokosielis, Sunil Kamath "Deploying Database Appliances in the Cloud",Bulletin of the IEEE Computer Society Technical Committee on Data Engineering", 2009.

[3] Donald Kossmann, Tim Kraska, Simon Loesing, "An Evaluation of Alternative Architectures for Transaction Processing in the Cloud", SIGMOD'10, 2010.

[4] Mladen A. Vouk, "Cloud Computing – Issues ,Research and Implementations", Journal of Computing and Information Technology - CIT , 2008,pp 235–246.

[5] Sunguk Lee, "Shared-Nothing vs. Shared-Disk Cloud Database Architecture", International Journal of Energy, Information and Communications Vol. 2, Issue 4, 2011.

[6] http://www.dcs.vein.hu/CIR/cikkek/Database_Environment.pdf

[7] Piotr Habela1, Krzysztof Stencel, Kazimierz Subieta, "Three-Level Object- Oriented Database Architecture Based on Virtual Updateable Views1" ADVIS 2006,Fourth Biennial International Conference on Advances in Information Systems, Volume 4243, 2006, 2006,pp 80-89.

[8] D.C. Tsichritzis, A. Klug (eds.): The ANSI/X3/SPARC DBMS Framework: Report of the Study Group on Data Base Management Systems, Information Systems , 1978.

[9] Dave Rosenberg, "Are database in cloud really all that different?,CNET" , 2011.

[10] Lamia Youseff, Maria Butrico, Dilma Da Silva, "Towards a Unified cloud Computing", Grid Computing Environments Workshop, GCE '08, 2008.

[11] http://www.scaledb.com/pdfs/CloudComputingDaaS.pdf

[12] Mansaf alam, " Cloud algebra for cloud database management system", CCSEIT '12, ACM New York, NY, USA, 2012 pp 26-29.