ISSN 23499842(Online), Volume 1, Special Issue 1(NCFCSIT 15), March 2015
International Journal of Innovative Trends and Emerging Technologies

DYNAMIC RESOURCE ORGANIZATION AND QUERY OPTIMIZATION STRATEGIES FOR CLOUD DATABASES

NIRAJA R JAIN[1], DR.B.RAGHU[2],
[1]Research Scholar, Bharath University,Chennai
[2]Professor & Dean, Sri Ramanujar Engineering College, Chennai

Abstract*:* **Cloud Computing technology is changing the way data is stored and accessed over the network. It is imperative to provide security to these data access as protect the privacy of the data owner. Data confidentiality has to be balanced with query processing functions and performance. Most cloud platforms are not designed for structured data management and hence fail to support SQL queries. How to improve query efficiency in cloud data management system especially query on structured data has become an important problem. This research proposal projects to propose data privacy preservation algorithms for cloud data. It also will try to analyse effective utilization of Codd's rules for query optimization. It plans to use dataflow optimization techniques and propose need of data descriptor based algorithms for improving query processing efficiency in cloud.**

Introduction:

Cloud Computing embraces concepts such as *software as a service(SaaS)* and *platform as a service(PaaS),* which incorporate services for workflow facilities for application design and development, deployment and hosting services, data integration, and network-based access to and management of software. Customers of clouds, much like customers of utility companies, can subscribe to different services at different service levels to guarantee the desired quality of service.

The clouds can also be defined as the one that provide on demand resources or services over the Internet, usually at the scale and with the reliability of a data centre. There are quite a few different types of clouds but with no standard way of characterizing the different types of clouds. The types of clouds are to categorize into the architecture model, computing model, management model and payment model.

Though not everyone agrees on the exact definition of cloud computing, most agree the vision encompasses: i) a general shift of computer processing, storage and software delivery away from the desktop and local servers, ii) across the network, iii) into next generation data centres hosted by large infrastructure companies (such as Amazon, Google, Yahoo, Microsoft or Sun [4]).

Cloud computing is an increasingly popular paradigm for accessing computing resources. Following are some of the major features of systems implementing Cloud with data centre, virtualization and WEB2 interfaces.

- Efficiency.
- Fault Tolerance.
- Ability to run in a heterogeneous environment.
- Ability to operate on encrypted data.
- Ability to interface with business intelligence products.

Recent advances in data centre, Virtualization, Cloud and WEB2 technologies are challenging and demand more security and query optimization to support scalable Cloud. For example, fault tolerance feature represents cloning of the failing server resulting in data mirroring and/or replication. Mapping the database query to static server may personalize the database connectivity. This can induce redundancy in the cloud.

A popular class of computing clouds is Database Infrastructure as a Service (IaaS) clouds exemplified by Amazon’s Elastic Computing Cloud (EC2). In these clouds, users are given access to virtual machines (VM) on which they can install and run arbitrary software including database systems. Users can also deploy database appliances on these clouds which are virtual machines with pre-installed pre-configured database systems. Deploying database appliances on IaaS clouds and performance tuning and optimization in this environment introduce some interesting research challenges that include data security, database resource deployment on VM cloud optimization verses query optimization, applicability of Codd’s query optimization.

Thrust areas in existing systems:

The multidimensional growth in computing systems and technologies have resulted in advanced scalable, portable and large scale integrated systems and technologies. Data centers, Virtualization, Cloud and WEB2 technologies are frontiers of such growth. Scalability and flexibility are two important dimensions of these technologies [1]. Cloud computing represents a shift away from computing as a product that is purchased, or compute as a service that is delivered to consumer over the internet from large scale data centres or cloud. In 1970 Codd E.F of IBM Research Lab, introduced the rules for query optimization and relational database [2]. This resulted in an efficient, optimized query processing. Cloud computing represents an important step towards realizing McCarthy's dream that all aspects of computation may someday be organized as a public utility service. The data centres used to create cloud services represents a significant investment in capital outlay and ongoing cost. Both public and private cloud platforms are looking to deliver the benefits of cloud compuing to their customers. Whether it is a public or private cloud, database is the critical part of that platform. Therefore it is imperative that your cloud database be compatible with cloud computing.

Cloud computing essentially provides services that share computational resources to execute potentially diverse requests on behalf of users who may have widely differing expectations. The following paragraph describes cloud data services in the Claremont Report on Database Research [6].

Early cloud data services offer an API that is much more restricted than that of traditional database systems, with a minimalist query language and limited consistency guarantees. This pushes more programming burden on developers, but allows cloud providers to build more predictable services, and to offer service level agreements that would be hard to provide for a full-function SQL data service. More work and experience will be needed on several fronts to explore the continuum between today's early cloud data services and more full-functioned but probably less predictable alternatives.

Cloud services make easier for users to access their personal information from databases and make it available to services distributed across Internet getting exposed to challenges of privacy and security. Users have typically to establish their identity each time they use a new cloud service, usually by filling out an online form and providing sensitive personal information (e.g., name, home address, credit card number, phone number, etc.). This leaves a trail of personal information that, if not properly protected, may be misused.

Rationale and Significance of Research Problem:

Cloud services make easier for users to access their personal information from databases and make it available to services distributed across Internet resulting in a threat to privacy and security of data. Users need to establish their identity each time they use a new cloud service usually by providing some sensitive personal information. This leaves a trail of personal information that if not properly protected may be misused.

Capability of acquiring and releasing the resources on-demand for internet applications needs to consider the facts such as handling the demands without compromising the QoS requirements, periodically predicting future demands to determine resources required, and accordingly allocating the resources. Two approaches to deal with this problem could be either Proactive – allocate resources before they are needed or Reactive – react to immediate demand fluctuations.

Here is a need to develop methods for effectively querying and deriving insight from ensuing sea of heterogeneous data. A specific problem is to answer the keyword queries over large collections of heterogeneous data sources. Developing index structure to support quering hybrid data is difficult.
Emergence of WEB 2.0 creates potential for new kind of data management scenarios in which users join ad-hoc communities to create, collaborate, curate and discuss data online.

Proposed Methodology:

Imagine that a client have developed an innovative web-based search service that he or she would like to offer to the world on WEB2. Cloud Computing enables you to host this service remotely and deal with scale variability: as the business grows or shrinks, one can acquire or release Cloud resources easily and relatively inexpensive. On the other hand, implementation and maintenance of data services that are scalable and adaptable to such dynamic conditions becomes a challenge. This is especially the case for data services that are compositions of other, possibly third-party services (e.g., Google Search or Yahoo Image Search), where the former become data processing graphs that use the latter as building blocks (nodes) and invoke them during their execution.

Running services under various quality-of-service (QoS) constraints that different customers may desire adds further complications. Handcrafting data processing graphs that implement such services correctly, make optimal use of the resources available, and satisfy all QoS and other constraints is a daunting task. Automatic dataflow optimization and execution are critical for data services to be scalable and adaptable to the Cloud environment. This is an analogy to query optimization and execution in traditional databases but with the following differences:

The component services may represent arbitrary operations on data with unknown semantics, algebraic properties, and performance characteristics, and are not restricted to come from a well-known fixed set of operators (e.g., those of relational algebra); optimality may be subject to QoS or other constraints and may be based on multiple diverse relevant criteria, e.g., monetary cost of resources, staleness of data, etc., and not just solely on performance; the resources available for the execution of a data processing graph are flexible and reservable on demand and are not fixed a-priori. These differences make data flow optimization essentially a new challenging problem; they also generate the need for run-time mechanisms that are not usually available.

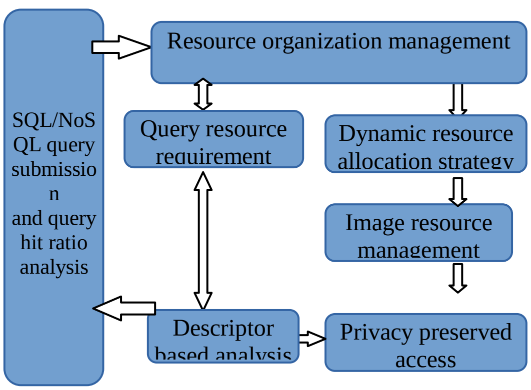

Figure: Introduction of Descriptor based analysis and classification algorithm to preserve privacy and help in dynamic resource allocation strategy

Expected contributions of research:

1. Security of the data in cloud is very important task.
2. Deployment of database on VM.
3. Optimization of recent WEB2 based query processing in cloud environment.
4. Analysis of cloud efficiency verses queries optimization efficiency.

Plan for evaluating work:

- To develop algorithm for data coding/decoding mechanism to assess the security of data in Cloud database system.
- To research the effectiveness of Codd's query Optimization in Cloud enabled Database to establish algorithmic tradeoff between cloud efficiency and query optimization efficiency.
- To develop Database resource deployment Algorithm on VM resource of cloud.

Conclusion:

The research proposal concentrates on the efforts to be taken in the following areas as is being underlined from the study of existing cloud based technologies:

i) Need to propose new resource organization strategy using data flow optimization techniques
ii) Need to propose algorithms for data access privacy preserving
iii) Need to propose a cloud query processing technique that conforms to the QoS.

With the introduction of descriptor based analysis and classification algorithm for parsing the query will eventually help in optimized dynamic resource allocation improving the query performance in cloud databases.

References:

1] Codd, Code optimization rules and relational database, Communications of the ACM Volume 13 , Issue 6 (June 1970) Pages: 377 – 387, ISSN:0001-0782
2] Daniel J. Abadi, Data Management in the Cloud: Limitations and Opportunities, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2009
3] Amazon. Amazon Elastic Compute Cloud (amazon ec2). ams.amazon.com/ec2, 2008.
4] Amazon. Amazon S3 Services Queue Service.Retrieved from http://aws.amazon.com/sqs, 2008.
5] Amazon. Amazon Web Services Developer Connection.Retrieved from http://aws.amazon.com, 2008.
6] R. Agrawal et al. The claremont report on database research. ACM SIGMOD Record, 37(3):9–19, 2008.
7] Rebecca Herald, What is the difference between security and privacy? CSI July 2002 Alert Newsletter
8] Hitoshi Motsumoto, Yutaka Ezaki, Dynamic resource management in cloud environment, FUJITSU Sci& Tech J, Vol 47, No 3, pg 270-276, July2011.
9] Sabrina De Capitani di Vimercati et al. Managing and accessing data in the cloud: Privacy risks and approaches, IEEE 2012
10] Shiyuan Wang et al. Compresensive framework for secure query processing on relational data in cloud, VLDB workshop on secure data management Nov 2010, doc 2010-25, University of California, Santa Barbara