International Journal for Innovative Engineering and Management Research

A Peer Revieved Open Access International Journal

www.ijiemr.org

IJIEMR Transactions, online available on 06th Jan 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=ISSUE-1

DOI: 10.48047/IJIEMR/V12/ISSUE 01/31

Title **Augment Economics of Elastic Data Relational Data Warehousing With Cloud**

Volume 12, Issue 1, Pages: 327-334

Paper Authors

M Ravi, S Divya

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Augment Economics of Elastic Data Relational Data Warehousing With Cloud

M Ravi[1], S Divya[2]
[1]Assistant Professor, Dept. of Artificial Intelligence, VJIT, Hyderabad, INDIA
[2]Assistant Professor, Dept. of Computer Science and Engineering, VJIT, Hyderabad, INDIA

Abstract
Cloud computing has been the most adoptable technology in the trend and deployment of data warehousing systems as a service has also moved into cloud environment. Significant advantages of software-as-a-service, which is often on cloud infrastructures, include reduced operational costs, faster development times and risks, and a flexible pay-as-you-go deployment strategy that allows elasticity in the face of unpredictable workload patterns. Our goal in this project is to determine whether business intelligence applications, and specifically data warehousing settings, may profit from this development. For this purpose, we first provide a functional overview of state-of-the-art data warehousing techniques as well as cloud infrastructures, and a review of existing efforts to combine the two along with security for Economics of Elastic Data [4]. With Challenges of Implementation in Data Mining, we concentrate on a significant technical challenge for the underlying cloud database architectures: how can the elasticity of data warehouses be supported by current cloud infrastructures, which frequently have limited inter-node storage and network bandwidth.

Introduction
Significant advantages of software-as-a-service, which is often on cloud infrastructures, include reduced operational costs, faster development times and risks, and a flexible pay-as-you-go deployment strategy that allows elasticity in the face of unpredictable workload patterns. Our goal in this project is to determine whether business intelligence applications, and specifically data warehousing settings, may profit from this development. In order to achieve this, we first give a functional overview of modern data warehousing methods, cloud infrastructures, and an analysis of attempts to merge the two. Throughout, we concentrate on a significant technical difficulty for the underlying cloud

database architectures: how to provide data warehouse elasticity with the present cloud infrastructures, which frequently have limitations in terms of inter-node storage and network capacity. Throughout, we concentrate on a significant technical difficulty for the underlying cloud database architectures: how to provide data warehouse elasticity with the present cloud infrastructures, which frequently have limitations in terms of inter-node storage and network capacity. To inform decision makers within an organisation, data in a data warehouse can be used directly through a query language, indirectly through data intelligence tools (like data mining), or even completely automatically as part of an automated decision-making instance (for example to steer customer interactions on a website). A decision-maker can browse the data in a data warehouse by using OLAP (online analytical processing) tools, which are supported by some data warehouses. For instance, data warehousing systems can be used to generate precise statistics or find hidden trends in the data.

Data warehousing solutions are in high demand because of the benefits they may provide enterprises in terms of business intelligence. But there are a number of things that make creating and sustaining a data warehouse system difficult:

- A data warehouse's setup can take a lot of time.
- Over-provisioning (over-estimating the system's requirements to maintain a particular service level at times of peak workloads) can result in expensive expenses.
- It's possible that organisations lack the knowledge necessary to set up and maintain a data warehouse.
- System overloads, crashes, and downtime can all have a variety of negative effects on a company.

There is a potential solution for these issues: cloud computing.

The phrase "cloud computing" describes software or other computational resources that are made accessible through a computer network (e.g. the internet). Through the provision of software as a service (SaaS) to customers, several businesses are attempting to make use of cloud computing. Customers can now use the internet to access the application online. Google Apps is a prime example of SaaS. Other instances include the several suppliers of hosted email applications, such as SkyInsight [2] and GoodData [3]. Computational power can also be made available as a cloud service in place of software, allowing users to run their own apps on the platform (PaaS, or Platform-as-a-Service). The Azure services from Microsoft [6], the services from Rackspace [5], and the S3 and EC2 services from Amazon [1] are a few examples of these services. Businesses who use these services to host their applications gain a number of advantages, such as:

- The organisation itself no longer requires the skills required to create and manage the systems hosting these apps.
- Over-provisioning can be avoided since an organisation no longer needs to estimate how much processing power an application will require; in the cloud, the appropriate amount of resources can be delivered to an application without the customer's involvement. Pay-per-use (PaaS) services frequently let the user only pay for the processing power that is really consumed, saving the user money.

Cloud computing is ushering in a new era of analytic data management for business intelligence (BI) by enabling organizations to analyze terabytes of data faster and more economically than ever before. The key change: cloud database software is provisioned within minutes, without data center overhead, and it's licensed on an on-demand basis.

M- Commerce

Mobile Commerce is the ability to access the internet "on the move" using your mobile phone, laptop, PDA or any other digital device. M-Commerce provides three main services:

1. Information based services such as instant messaging, e-mail or even searching for sports results.
2. Financial based services such as purchasing apps, games, music or even searching for the price of something then buying the product in a physical store.
3. Personalized based services that anticipate what the customer wants based on that person's location or data profile. For example, coupons for nearby restaurants

In 2004 there were 1.5 billion mobile subscribers compared to 5.3 billion users today which is over 77% of the world population. The trend in M-Commerce is

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

increasing dramatically and is becoming an essential asset with many companies.

The Graphic below highlights some of the key features of M-Commerce.

The Future of M-Commerce

- M-Commerce for educational purpose
- Junior Certificate/ Leaving Certificate
- Record answers over the phone
- Students more focused and concentrated
- Reduces inventory and labor costs
- Answers recorded and sent instantly for correcting

Objective(s) and Scope:

The purpose is to respond to the following essential queries:

- ➢ Is it possible to combine both Relational elastic data warehouse and Cloud along with security
- ➢ What might a cloud-based relational elastic data warehouse look like?

Description of the research work:

The fundamental design of a data warehouse

Data warehousing can be accomplished utilising a variety of for-profit and nonprofit systems. For a data warehouse to function as intended, a number of components are always required. The following figure depicts these elements.

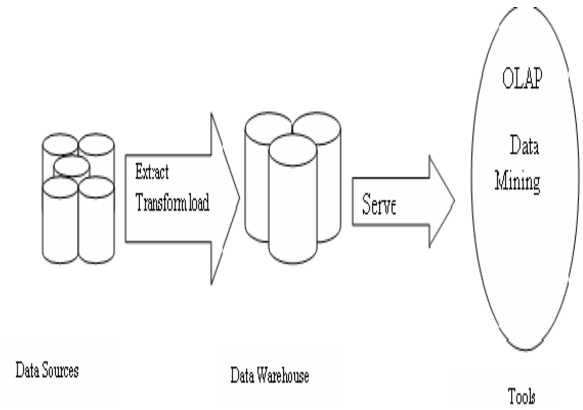

Figure 1: Depicts a data warehouse's general structure.

The data sources, which are primarily OLTP systems, are where a data warehouse gets its information from. All of the data from a company or organisation can be gathered in the data warehouse component. Before entering the data warehouse component, data might be cleaned. This means that for the data warehouse system to function as intended,

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

anomalies like inconsistencies and duplication must be eliminated from the data. It can be helpful if data is supplied in a specified consistent format so that the data warehouse can reason with it coming from potentially many various sources, which means that data will typically have

to be transformed before it enters the data warehouse. There is also the option of performing minimal or no cleaning or modification at all, however this can make data querying challenging [10].

Relational data warehouse design

A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing. It usually contains historical data derived from transaction data, but it can include data from other sources.

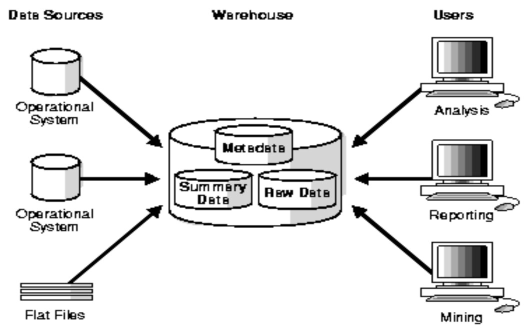

Figure 2: Relational data warehouse design

Cloud Computing: Guiding Ideas and Actual Use

In this article, we will provide a general overview of cloud computing to assist us decide whether or not data warehousing solutions can be modified to work in a cloud environment. This section will also examine the various cloud computing resource providers in terms of the performance and cost-based features of their services.

Why use cloud computing?

The cloud is a viable alternative to traditional computing methods for a number of reasons. Here are a few of these causes:

- Scalability to infinity, the illusion of boundless computer resources exists with cloud computing [7]. If a customer needs extra resources, he or she can rent them, and the consumer will nearly immediately have access to more capabilities.
- Deployment speed. When compared to internal deployment, cloud providers' full-service offerings can speed up deployment.
- Elasticity. A pay-per-use payment model is typically used in cloud computing [7], which means that you only pay for the services you actually utilise. With this methodology, there is no chance of incurring starting costs or over-provisioning charges while also

avoiding the risk of under-provisioning and possibly missing service levels.

Reliability. A cloud provider might theoretically attain high reliability.

- Making backups is one way to do this, but there are other ways as well, like having numerous data centres (allowing for handling for example power outages). This is dubious, though, as there have already been large service interruptions in cloud computing [7].
- Reduced costs. Costs can be decreased at the cloud provider due to economies of scale (things tend to get cheaper as scale increases), which may enable consumers to lower costs as well. When use is low, expenses can be decreased thanks to elasticity because the client uses less resource and pays less as a result. Customers may be able to save money by not needing to maintain specific expertise in-house because it can be centrally located at the cloud provider.

Putting it all together: Cloud Elastic Data Warehousing

We will now make assumptions regarding the relationship between the domains of data warehouse and cloud computing to determine whether or not there are bright prospects for data warehousing in the cloud. We'll achieve this by examining the previously cited grounds against going to the cloud: the slow speed of data transfer, subpar cloud performance, loss of control, and cost concerns. We will explicitly examine the available means of allowing flexibility. Following that, we will outline a set of functional requirements for cloud-based data warehousing solutions and talk about the challenges associated with meeting these objectives.

A New Analytic DBMS for the New Frontier

In order for these pioneering analytic cloud projects to succeed -- especially as data volumes grow—they will require a database architecture that is designed to function efficiently in elastic, hosted computing environments like the cloud.

A computing cloud, such as the Amazon EC2, is composed of thousands of commodity servers running multiple virtual machine instances (VMs) of the applications hosted in the cloud. As customer demand for those applications changes, new servers are added to the cloud or idled and new VMs are instantiated or terminated.

Cloud computing infrastructure differs dramatically from the infrastructure underlying most in-house data warehouses and data marts. There are no high-end servers with dozens of CPU cores, SANs, replicated systems, or proprietary data warehousing appliances available in the cloud. Therefore, a new DBMS software architecture is required to enable large volumes of data to be analyzed quickly and reliably on the cloud's commodity hardware. Recent DBMS innovations, such as those featured in the Vertica Analytic Database for the Cloud make this a reality today, and they include.

Automatic high availability

Node failures and node changes are possible in a cloud-based analytic database cluster. If the database has the appropriate built-in failover capabilities, these failures can be made transparent to the end user given the enormous number of processing components within a cloud. The best cloud databases will automatically replicate data across the cloud cluster's nodes, be able to continue operating in the event that one or more nodes fail ("k-safety"), and be able to automatically restore data on recovered nodes without DBA help. The ideal replication will be "active-active," allowing queries on the redundant data to boost performance.

Ultra-high performance

The capability of the cloud to quickly set up analytical applications is one of its game-changing benefits (without waiting for hardware procurement). However, the virtualized cloud environment and the Internet access rates can have an impact on performance. The advantage disappears if the analytical performance is poor. Fortunately, shared-nothing columnar databases like Vertica for the Cloud are created expressly for analytical workloads and have proven to significantly outperform conventional, row-oriented databases in terms of performance (as verified by industry experts, such as Gartner and Forrester, and by customer benchmarks). This increase in software performance, when combined with the hardware economies of scale made possible by the cloud environment, gives cloud analytics a new economic model and a competitive edge.

Aggressive compression

Since cloud costs are typically driven by charges for processor and disk storage utilization, aggressive data compression will result in very large cost savings. Row-oriented databases can achieve compression factors of about 30% to 50%.

Conclusion

Both the subject of data warehousing and the field of cloud computing have been examined throughout this study. The pros and cons of merging the two have been studied in order to perhaps offer elastic data warehousing that is both economical and effective via the cloud. We think that cloud-based data warehousing systems have a lot of potential because they offer the potential for elasticity, scalability, rapid deployment, reliability, and lower prices (due to e.g. elasticity). However, we think that because of present concerns with cloud performance, data transfer speed, and pricing, current cloud computing systems are not now instantly capable of executing large-scale data warehousing. In addition to meeting many of the functional requirements covered in this paper, a data warehousing system will need to be highly parallel and distributed in order to take use of the cloud's capabilities. Additionally, we think that data marts may have greater promise in the cloud in the near future than data warehousing systems because they are typically smaller due to their specialised nature, which allows them to fit in fewer or just one node. Distributing and parallelizing become less of a problem as a result. The choice to move data warehousing systems or data marts into the cloud will probably always involve security concerns.

More study is required to determine how to combine data warehousing and cloud computing. Our research on data warehousing in the cloud is primarily speculative, and benchmarking the cloud, for instance, may reveal fresh information on the viability and impossibility of deploying data warehousing systems in the modern cloud environment. Testing these recently introduced features may give us further information into the suitability of these systems in the cloud. Amazon supports IBM's DB2 and Oracle's database 11g. A universal design that can take advantage of the cloud's potential for data warehousing may also be introduced, and this may be of great utility.

References

[1] Amazon web services http://aws.amazon.com/

[2] SkyInsight web page http://www.ingres.com/products/cloud/sky insight

[3] GoodData general website http://www.gooddata.com/

[4] Mr. M.Ravi Mrs. S.Divya Economics of Elastic Data Warehousing with Cloud

[5] Rackspace http://www.rackspace.com/

[6] Microsoft Azure http://www.microsoft.com/windowsazure/

[7] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Konwinski, G. Lee, D.A. Pattersom, A. Rabkin, I. Stoica, M. Zaharia: Above the Clouds: A Berkeley View of Cloud Computing, 2009.

[8] J.Wijsen: Een kort overzicht van data warehousing en OLAP. 2006

[9] E.Rahm, H.H.Do: Data Cleaning: Problems and Current Approaches. IEEE Data Engineering Bulletin, 2000.