Journal of Computer and System Sciences 81 (2015) 85–96

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Cloud BI: Future of business intelligence in the Cloud

Hussain Al-Aqrabi, Lu Liu *, Richard Hill, Nick Antonopoulos

School of Computing and Mathematics, University of Derby, Derby, Derbyshire, United Kingdom

ARTICLE INFO

Article history:
Received 31 January 2013
Received in revised form 15 October 2013
Accepted 22 January 2014
Available online 25 June 2014

Keywords:
Business intelligence
Online analytical processing
Cloud computing
Software-as-a-service
Database-as-a-service
Massively parallel systems

ABSTRACT

In self-hosted environments it was feared that business intelligence (BI) will eventually face a resource crunch situation due to the never ending expansion of data warehouses and the online analytical processing (OLAP) demands on the underlying networking. Cloud computing has instigated a new hope for future prospects of BI. However, how will BI be implemented on Cloud and how will the traffic and demand profile look like? This research attempts to answer these key questions in regards to taking BI to the Cloud. The Cloud hosting of BI has been demonstrated with the help of a simulation on OPNET comprising a Cloud model with multiple OLAP application servers applying parallel query loads on an array of servers hosting relational databases. The simulation results reflected that extensible parallel processing of database servers on the Cloud can efficiently process OLAP application demands on Cloud computing.

1. Introduction

Cloud computing has become one of the revolutionary technologies over recent years. Cloud computing is conceptualised in three forms – software-as-a-service (SaaS), platform-as-a-service (PaaS) and infrastructure-as-a-service (IaaS). The SaaS providers interface with the end users by virtue of provisioning of business application services similar to the ones that have been traditionally self-hosted by the corporate houses [1]. Cloud computing paradigm has emerged to bring large-scale computing, storage resources and data service resources together to build a VCE (virtual computing environment) [2]. Cloud computing users can discard the hassles of large-scale investments in hardware and software platforms, in upgrading them regularly and in expensive licenses of application software used to run business processes, related transactions and decision-support systems [3].

The Cloud is generally a multi-tenant computing environment; the multi-tenant Cloud solutions can optimise resource sharing while providing isolation solution at different levels required to the tenant [4].

This model has ensured better affordability of the best possible application systems thus supporting an increase in efficiency of businesses [5]. Resources are allocated to end users against service requests made by their end terminals, and resources are allocated by a service provisioning engine that verifies the eligibility of users from a separate schema object that holds multi-tenancy data about all Cloud users and groups. Once the eligibility is verified, the resources are reserved for the user through session bindings until the computing processes are in progress by the user terminal. The terminal is normally a virtualised client presented through a virtual server farm. However, there can be direct loading of resources as well (example, for data backup). A separate layer monitors the session usage and utilisation of resources

* Corresponding author.
E-mail address: l.liu@derby.ac.uk (L. Liu).

http://dx.doi.org/10.1016/j.jcss.2014.06.013
0022-0000/

such that the billing related information can be generated [6]. NIST (National Institute of Standards and Technology) is in the process of developing standard protocols for user connectivity to the Cloud through virtualisation interface, terminal emulation interface, thin client interface and Internet browser interface. As of now, there is no standard protocol for users' connectivity to Cloud hosted resources [7].

As a data-centric approach to business intelligence (BI), data acquisition is becoming easier to acquire and large data warehouses with 10–100s of terabytes of relational database management systems (RDBMS) are becoming increasingly common due to the popularity of interactive, web-based databases [33]. BI has been historically one of the most resource intensive applications. It comprises a number of data warehouses created by fetching decision-support data from organisation wide databases. The data warehouses are updated at frequent intervals through appropriate queries executed on the business processing and transactional databases. Online analytical processing (OLAP) is the user-end interface of BI that is designed to present multi-dimensional graphical reports to the end users. OLAP employs a technique called multidimensional analysis is mainly used to enable flexible interactive analysis of multidimensional data [32]. Whereas a relational database stores all data in the form of rows and columns, OLAP also employs data cubes formed as a result of multidimensional queries run on an array of data warehouses. Furthermore, an OLAP application fetches data from the data warehouses, organises them in highly complex multidimensional data cubes, and presents to the users through user defined and configured GUI dashboards [8]. BI and OLAP framework has a high business utility, because it helps in locating and eliminating or solving business process deficiencies, inefficient process steps and waste process steps. A BI and OLAP framework is expected to provide timely, accurate, organised and integrated information to business decision makers [8,9].

Despite of excellent business utility of BI and OLAP framework, many business owners were compelled to look for its alternative because of uncontrolled increase in computing and storage resource requirements in self hosted environments. At some stage, the cost of maintaining and upgrading the BI and OLAP framework becomes unjustified for a business [10]. However, the unique selling points of Cloud computing offer exactly what businesses need to successfully run BI and OLAP frameworks-unlimited resources, resource elasticity (resources on demand), moderate usage costs, high uptime and availability, high security, no hassles of upgrading and maintaining loads of servers and databases, and so on [1,5]. Hence, it is hereby argued that Cloud computing has the potential to offer a new lease of life on BI and OLAP framework. Moreover, it is also argued that Cloud computing can extend the power of BI and OLAP to small- and medium-scale businesses, which could not afford the framework in self-hosted IT infrastructures. However, it is important to establish a framework for implementing BI and OLAP on a Cloud computing platform.

The rest of the paper is organised as follows. Section 2 is a literature review on how BI and OLAP framework can be implemented on the Clouds and presents key benefits of Cloud computing for BI. Section 3 shows an approach for taking BI to the Cloud as well as key challenges in hosting BI on Cloud. Section 4 describes and explains in detail how BI and OLAP framework can be modelled on a Cloud and how it should behave in order to extend maximum utility to the businesses by virtue of an OPNET based simulation experiment. Section 5 presents the summary of the research results and analysis. Finally, the conclusions of this work with future directions are discussed in Section 6.

2. Literature review

2.1. A review of business intelligence and OLAP and their porting on Cloud computing

Business intelligence (BI) is employed for monitoring the performance of business processes through accurate presentation and analysis of multidimensional data taken from distributed transaction processing systems across the enterprise [23]. The analysts using OLAP-enabled dashboards and reporting systems prefer to map financial data with performance data (of people and processes) to identify inefficiencies and reduce them through strategic restructuring of the business processes and workflows [23]. As per [24], a BI system is made of seven layers: IT and related infrastructure, data acquisition, data integration, data storage, data organising, data analytics and data presentation.

The OLAP cubes form the data storage (partially, in the form of views) and data organising layers of a BI system. It sits over the data warehouse tables in the form of multidimensional views. A cube structure is made of a number of cross-referenced columns having data fetched from different processes and financial data tables over a period. The periods in a cube are shorter than the ones in the data warehouse that is tasked to store process related and financial data of longer periods (typically, five years or more). The OLAP queries are heavy duty search commands comprising data from multiple views at a time. The presentation layer makes the data sets visible in the form of multidimensional graphical screens [25,26].

A number of OLAP query types are used to generate the views and present them in the dashboards/presentation screens. Some of the popular OLAP queries include: slice and dice, pivoting queries, merge/split queries, rolling-up queries, and drilling down queries [26]. The OLAP cube can be visualised as a stack of two-dimensional matrix planes, whereby each matrix plane represents a relationship between two different dimensions [26]. Fig. 1 is a presentation of a multidimensional OLAP cube comprising two dimensional matrix planes.

It should be noted that these matrices are not independent of each other. All the attributes in these planes are nested with each other, and are linked with a primary key that controls all the relationships across multiple composite elements many-to-many relationships. For example, a product code can be viewed as a primary key that controls the relationships among all the attributes related to sales in a business [27].