

Chapter 5

Workload Management Systems for the Cloud Environment

Eman A. Maghawry

Ain Shams University, Egypt

Rasha M. Ismail

Ain Shams University, Egypt

Nagwa L. Badr

Ain Shams University, Egypt

Mohamed F. Tolba

Ain Shams University, Egypt

ABSTRACT

Workload Management is a performance management process in which an autonomic database management system on a cloud environment efficiently makes use of its virtual resources. Workload management for concurrent queries is one of the challenging aspects of executing queries over the cloud. The core problem is to manage any unpredictable overload with respect to varying resource capabilities and performances. This chapter proposes an efficient workload management system for controlling the queries execution over a cloud. The chapter presents architecture to improve the query response time. It handles the user's queries then selecting the suitable resources for executing these queries. Furthermore, managing the life cycle of virtual resources through responding to any load that occurs on the resources. This is done by dynamically rebalancing the queries distribution load across the resources in the cloud. The results show that applying this Workload Management System improves the query response time by 68%.

INTRODUCTION

Modern business software and services demand high availability and scalability. This has resulted in an increasing demand for large scale infrastructures which generally tend to be highly complicated and expensive for the organization requesting them. These higher resource and maintenance costs have given rise to a paradigm shift towards cloud computing. Its services are offered and maintained by various providers over the Internet. These service offerings range from software applications to virtualized platforms and infrastructures (Mittal, 2001). Cloud Computing is the combination of parallel and distributing computing paradigms. This distributed computing paradigm is driven by the economies of scale, in which a pool of virtualized and managed computing power, storage, platforms and services are delivered on demand to remote users over the Internet (Foster, Zhao, Raicu, & Lu, 2008). It is helping in

DOI: 10.4018/978-1-5225-2229-4.ch005

realizing the potential of large scale data intensive computing by providing effective scaling of resources (Mian, Martin, Brown, & Zhang, 2001). Also, a Cloud computing environment is widely employed in scientific, business and industrial applications to increase the use and delivery model involving the Internet to provide dynamic virtualized resources (Jeong & Park, 2012). It offers the vision of a virtually infinite pool of computing, storage and networking resources where applications can be scalable deployed (Hayes, 2008).

As data continues to grow, it enables the remote clients to store their data on a cloud storage environment with different clients' expectations over the internet. An increasing amount of data is stored in cloud repositories which provide high availability, accessibility and scalability (Duggan, Chi, Hacigumus, Zhu, & Cetintemel, 2013). As clouds are built over wide area networks, the use of large scale computer clusters are often built from low cost hardware and network equipment, where resources are allocated dynamically amongst users of the cluster (Yang, Li, Han, & Wang, 2013). Therefore the cloud storage environment has resulted in an increasing demand to co-ordinate access to the shared resources to improve the overall performance.

Users can purchase traditional data centers, sub-divide hardware into virtual machines or outsource all of their work to one of many of cloud providers (Duggan et al., 2013). As the number of users submitting the queries increase, that leads to increases in the load and traffic of the virtual resources. So it is essential to incorporate a mechanism to balance the load across these virtual resources (Somasundaram, Govindarajan, Rajagopalan, & Madhusudhana, 2012).

Workload management is the discipline of effectively managing, controlling and monitoring application workloads across computing systems (Niu, Martin, & Powley, 2009). Workload is as a set of requests that access and process data under some constraints. The data access performed by a query can vary from retrieval of a single record to the scan of an entire file or table. Since the load on data resources in a cloud can fluctuate rapidly among its multiple workloads, it is impossible for system administrators to manually adjust the system configurations in order to maintain the workloads objectives during their execution. Therefore, managing the query workload automatically in a cloud computing environment is a challenge to satisfy the cloud users. This is done by relocating resources through admission control in the presence of workload fluctuations (Niu, Martin, Powley, Horman, & Bird, 2006).

The workload produced by queries can change very quickly; consequently this can lead to decreasing the overall performance (e.g. query processing time) depending on the number and the type of requests made by remote users. In this case, a cloud service provider must manage the unpredictable workloads, through making decisions about which requests from which users are to be executed on which computational resources.

These decisions are based on the progress feedback of the workload or the behavior of the resources to recover any load imbalance that may occur. This can lead to improving the overall performance during the queries execution over the distributed resources. Adaptive query processing changes the way in which a query is being evaluated while the query is executing over the computational resources (Paton, Araújo, & Fernandes, 2012). Furthermore, the importance of managing the workload arises by the demand to revise resource allocation decisions dynamically.

The challenge in this chapter is how to provide a fast and efficient monitoring process to the queries executed over the distributed running resource. Furthermore, responding to any failure or load imbalance occurs during the queries execution. This is done by generating an assessment plan that redistributes the queries execution over the replicated resources.

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/workload-management-systems-for-the-cloud-environment/180941

Related Content

Towards a Model of Fuzzy Commonsense Reasoning

Xenia Naidenova (2010). *Machine Learning Methods for Commonsense Reasoning Processes: Interactive Models* (pp. 279-318).

www.irma-international.org/chapter/towards-model-fuzzy-commonsense-reasoning/38487

A Proposal of SDN Based Disaster-Aware Smart Routing for Highly-Available Information Storage Systems and Its Evaluation

Satoru Izumi, Misumi Hata, Hiroyuki Takahira, Mustafa Soylu, Asato Edo, Toru Abe and Takuo Suganuma (2017). *International Journal of Software Science and Computational Intelligence* (pp. 69-83).

www.irma-international.org/article/a-proposal-of-sdn-based-disaster-aware-smart-routing-for-highly-available-information-storage-systems-and-its-evaluation/175656

Improving the Performance of Neuro-Fuzzy Function Point Backfiring Model with Additional Environmental Factors

Justin Wong, Danny Ho and Luiz Fernando Capretz (2014). *Exploring Innovative and Successful Applications of Soft Computing* (pp. 260-280).

www.irma-international.org/chapter/improving-the-performance-of-neuro-fuzzy-function-point-backfiring-model-with-additional-environmental-factors/91884

Classification with Incomplete Data

Pedro J. García-Laencina, Juan Morales-Sánchez, Rafael Verdú-Monedero, Jorge Larrey-Ruiz, José-Luis Sancho-Gómez and Aníbal R. Figueiras-Vidal (2010). *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques* (pp. 147-175).

www.irma-international.org/chapter/classification-incomplete-data/36984

Sparse Based Image Classification With Bag-of-Visual-Words Representations

Yuanyuan Zuo and Bo Zhang (2011). *International Journal of Software Science and Computational Intelligence* (pp. 1-15).

www.irma-international.org/article/sparse-based-image-classification-bag/53159