ISSN No. 0976-5697

Volume 5, No. 2, March 2014 (Special Issue)

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Study on Database Virtualization for Database as a Service (DBaaS)

Mrs. Sridevi Bonthu
Department of Computer Science and Engineering
SVECW Bhimavaram, India
sridevi.db@gmail.com

Ms. Sai Divya Manasa Thammiraju
Department of Computer Science and Engineering
SVECW Bhimavaram, India
divyamanasa6@gmail.com

Dr. Y S S R Murthy
Department of Computer Science and Engineering
SVECW Bhimavaram, India
hodcse@svecw.edu.in

Abstract: Advancements in operating system virtualization and storage virtualization technologies have enabled the effective consolidation of heterogeneous applications in a shared cloud infrastructure. Research challenges which are novel arising from this new shared environment include load balancing, resource isolation, workload estimation, live migration, machine replication, and an emergent need of automation to handle large scale operations with minimal manual intervention. Databases are at the core of most applications that are deployed in the cloud, database management systems (DBMSs) represent a very important technology component that needs to be virtualized in order to realize the benefits of virtualization from autonomic management of data-intensive applications in large scale data-centers. The goal of this paper is to survey the techniques used in providing elasticity in virtual machine systems, shared storage systems and survey on database research on multitenant architectures and elasticity primitives. Studying Database as a Service advances, together with study of related topics in OS and storage-level virtualization, as they are central for anyone that wants to operate in this area.

Keywords: virtualization, elasticity; live-migration; load-balancing; resource isolation; multitenant systems;

I. INTRODUCTION

Most of the now-a-days enterprises are faced with managing a huge number of databases. The databases are consolidated into a unified service can reduce the high costs associated with running dedicated servers and maintenance by multiple database administrators. This is referred to as Database-as-a-Service (*DBaaS*). Traditional DBMS technology has focused on fully utilizing a single server to meet the needs of a specific application. Building a Database-as-a-Service infrastructure requires several key challenges to be addressed to transform traditional DBMSs into a scalable, elastic and autonomic database platform[1]. The scale of applications hosted within a database platform calls for solutions to self-manage resource allocation, load-balancing, replication and other issues which are not supported by commercial DBMS vendors. Several technological advances have been produced by the systems and storage communities to face similar virtualization and consolidation needs.

Virtualization technology has been a key enabler in building consolidated infrastructure services, by allowing physical hardware to be multiplexed into many virtual machines (*VMs*). Consolidated services, such as Amazon's public Elastic compute cloud EC2 (Amazon EC2 is a web service that provides resizable compute capacity in the cloud), use virtualization to share resources of large commodity servers between many smaller compute instances, which often can operate without dedicated resources. A *shared storage abstraction* can be used to separate the physical storage of virtual disks from the virtual machine in virtualized environments. Shared storage enables mobility of VMs by requiring only a machine's memory state be migrated between hosts, as the larger persistent image is stored and mounted from an external network service. For virtualized computers and storage, the systems must respond to over-utilized servers by using automated resource isolation and workload placement to ensure applications receive adequate resources. A large part of success in cloud computing is driven by such advances in virtualization technology that have resulted in huge economies of scale in large-scale datacenter operations.

The database research community and key industrial players have embraced the idea of virtualization, and contributed several approaches to tackle this problem ranging from direct application of OS and shared-storage virtualization techniques (e.g., Amazon EC2's DBMS in a VM approach), reinterpretations of the same ideas in the DBMS space, to completely new and useful technical solutions. The common underlying intuitions are that: 1) databases represent an important enough subclass of applications to be tackled in a special way, and 2) databases offer a narrow and semantically rich API towards the applications, that can be leveraged by system builders to devise specialized optimization. This is a rapidly evolving field with advances both in academic and industrial context and therefore in this paper we try to identify the importance of *Database Virtualization*, and conclude by identifying research trends in this direction.

II. DATABASE VIRTUALIZATION

Virtualization is the creation of a virtual version of something (operating system, storage devices, network, etc.) that can be deployed and managed in a more fine-grained manner than the physical item itself. For example, a single physical server (or machine) can be sliced into various virtual servers (virtual machines or VMs), each embodying various resources (memory, disk, CPU cores, etc.)[8]

Database virtualization means different things to different people. Database virtualization is simply from running the database executable in a virtual machine, or using virtualized storage, to a fully virtualized elastic database cluster composed of modular compute and storage components that are assembled on the fly to accommodate our database needs.

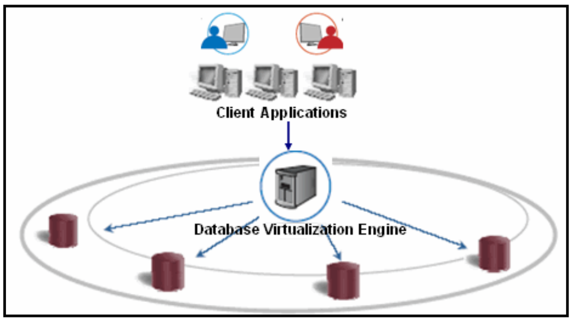

Figure: 1

Virtualization is abhorrence to traditional SQL-based OLTP databases. These databases tightly integrate the compute, caching and storage in order to: (a) optimize the performance (b) coordinate locking to ensure that the database remains consistent and (c) provide “copies” for fail-over or high-availability. Building a fully virtualized database management system (DBMS) is a huge undertaking. However, deploying and using a fully virtualized DBMS is actually quite easy, since it eliminates a collection of deployment challenges that are required to scale-put single instance DBMS. In database virtualization, the idea is to separate the logical (database functions) from the physical (the server). So the biggest challenge in virtualizing is to maintain the competitive performance while the databases are to abstract logical processes away from the physical hardware.

III. DBAAS: DATABASE AS A SERVICE

Cloud Database or Database-as-a-Service is one of the most demanded services in the cloud computing market. In short, a cloud database is a database that is specifically built for and runs on a virtualized cloud computing platform. There are two different models present to run the databases on the cloud independently in the industry now a days. They are (1)By using a virtual machine and (2) By paying for the Database-as-a-Service (DBaaS)

Figure: 2

These cloud database services allows customers to take benefits like the ability to pay solely by the provided storage capacity. It enables moving of databases from public to private cloud or from one provider to another and offers **on-demand scalability and high availability (HA).** Transporting the database to the cloud is very effective for companies who plans to deploy their applications broadly to the cloud as a part of **Software as a Service (SaaS)**. The ability to transport the database easily is one of the reasons the cloud DBaaS is much demanded.

In addition to this, a Database-as-a-Service is a service where clients pays per use and get access to the cloud database service on-demand. This kind of service is much affordable and attractive for clients, since the costs related to the hardware; software that runs and maintenance is bear by the service provider. Following are the outlines of the subjects relate to challenges in building a Database-as-a-Service infrastructure.

A. Process Elasticity:

Virtualization is a key component in providing cloud offerings, with an effective ability to consolidate operating systems into a single machine. Despite the presence of a controller (hypervisor) to share a system’s limited physical resources, co-located Virtual Machine instances do compete for the same limited resource capacity. Two techniques are often utilized, to mitigate the impact of resource starvation. Firstly, load-balancing the placement of instances can remove resource bottlenecks, but requires the ability to migrate a Virtual Machine instance between machines[2]. Secondly, resource allocation or isolation can be enabled by the hypervisor to ensure fair resource access between Virtual Machines [3]. Both solutions require the ability to characterize a workload and predict the impact of co-location.

B. Storage Elasticity:

Another important challenge in building database as a service is efficient, elastic sharing, and access to storage resources. Significant industrial development has been focused in providing an automated, decoupled, storage infrastructure which controls the placement and load-balancing of IO workloads across a set of storage devices. These systems enable a VM to mount a virtual disk, which is hosted on one or more servers within the same network. The decoupling of storage from computation enables lightweight elasticity, consolidation, availability, and fault-tolerance.

By managing virtual disks for tens to hundreds of data stores, these systems highlight the need for automated solutions in placement and load-balancing. To manage these issues, a system must make decisions based on workload characterization [4], device models, and analytical or sampling-based formulations [5].

C. Database Elasticity:

A relational Database-as-a-Service provider can share resources of a single database server among multiple tenants. Multi-tenancy is an architecture in which a single instance of a software application serves multiple customers. Cost reduction for the cloud service provider is enabled by multi-tenancy, there by savings to the tenants.

Within the various multitenant systems, recent research has focused on bringing virtualization to database platforms. A critical issue in these systems is ensuring that each tenant has resources to serve well-formed requests within a certain time period, alternatively a service level object (SLO). As

CONFERENCE PAPER
Two day National Conference on Advanced Trends and Challenges in Computer Science and Applications
Organized by: Shree Vishnu Engineering College for Women, Bhimavaram A.P.
Schedule: 18-19 March 2014

with other consolidated systems, several approaches can ensure that SLOs are met, even in the presence of dynamic workload. Resource isolation can ensure that each tenant is guaranteed a fixed or minimal amount of underlying system resources [6], or intelligent tenant placement can ensure SLOs are met [7]. Both approaches require the ability to predict the impact of colocation, model workloads, and predict resource consumption. For solutions that rely on tenant placement to ensure SLOs, a migration primitive must be available to move tenants between servers.

D. Trends Pushing Database as a Service (DBaaS):

Following quotes correlate directly to the following enterprise trends that make the case for DBaaS:

a. ***Database Sprawl and Infrastructure Growth*** **–** Over time, companies created hundreds or thousands of copies of databases that are hardly used which sit idle on servers and cost millions to enterprise. With the growth of data, database infrastructure management has become complicated and introduced many risks.
b. ***Self Service Technology*** **–** Since IT became a department, it has been overworked and used. Work-groups (inside and outside IT) got tired of waiting and needed “IT on demand.” When the demand for infrastructure outstriped the ability for IT to deliver, people found answers by putting a server underneath someone’s desk to “serve themselves” or by going to an external service provider. With this trend, compliance, risk, and security have had a hard time keeping up.
c. ***Cost Savings from Virtualization*** **–** Virtualization has cut billions of dollars of costs from IT and has billions to go as more workloads get virtualized. Ironically, companies are seeing up to 40-50% cost reduction in infrastructure budgets, while database budgets continue to increase.
d. ***Data Driven Business Decisions*** **–** Every workgroup and department is using applications to improve business operations and companies want to better understand customer usage of the internet and apps. As we use data more to power our businesses through analytics, “real production data” becomes more and more important.
e. ***Cost-Effectively Scaling the Data Layer*** **–** While people are using applications and data more than ever, the traditional database/infrastructure model has been expensive to scale at an enterprise level of SLAs. For example, growth in data often leads to new and expensive orders for both hardware and software, and there is too much expense and complexity to make Tier 2 databases highly available.

IV. CLOUD DATABASE SERVICES

Cloud database technology is in high demand because these databases are attractive for small or large businesses as new applications can be deployed with little to no upfront investment or long-term commitment. Number of cloud database solutions exists today and with most services, the pricing and quality can vary greatly from one cloud database solution to another. Some of the popular cloud databases are listed below.

a. ***Amazon Relational database (RDS)*** **-** Amazon Relational Database Service (Amazon RDS) is a web service which makes easy to set up, scale and operate a relational database in the cloud. It provides cost-effective and resize the capacity while managing time-consuming database managing tasks and freeing user up to focus on main his/her applications and process. Amazon RDS provides user access to the capabilities of a just like Oracle, SQL Server, MySQL, or PostgreSQL database engines. That means the applications, code, and tools user already use today with his/her existing databases can be used with Amazon RDS. Amazon also provides a variety of auxiliary data management services, such as its newly named data warehouse named Red shift, as well as Data Pipeline, which helps users integrate data from multiple sources for easier management.
b. ***Amazon DynamoDB :*** Amazon DynamoDB is ideal for database applications that require very low latency and predictable performance at any scale but don’t need complex querying capabilities like joins or transactions. Amazon DynamoDB is a completely-managed NoSQL database service that provides high performance, expectable throughput and less cost. It is easy to set up, operate, and scale. With Amazon DynamoDB, you can start low, express the throughput and space user needed, and easily examine user capacity requirements on the fly. Amazon DynamoDB mechanically partitions data over a number of servers to meet your request capacity. In addition to that, DynamoDB automatically replicates your data synchronously across multiple Availability Zones within an AWS Region to ensure high-availability and data durability.Amazon DynamoDB uses a table based data model that doesn’t require a fixed schema and enables data access mainly through particular special keys. With that, the service goes to strongly consistent reads and natively supports Atomic Counters, allowing you to atomically increment or decrement numerical attributes with a single API call.
c. ***Google Cloud SQL:*** Google Cloud SQL is a MySQL database that lives in Google's cloud. It has all the abilities and functionalities of MySQL, with a few extra features and some unsupported features .Google Cloud SQL is easy to use, don’t need extra software installation or maintenance and is ideal for small to medium-sized applications. Google Cloud SQL presently available for Google App Engine applications that are written in different languages likeJava, Python, and so on.user can also access Google Cloud SQL by using MySQL Client, and other administration and reporting tools that work with MySQL databases.
d. ***Microsoft Azure:*** Windows Azure SQL Database, previously SQL Azure, is a completely managed relational database service that delivers flexible manageable, it includes built-in high availability, and offers expectable performance, and supports huge scale-out. Featured in different service tiers to meet basic and high-end needs, SQL Database supports organizations to rapidly build, extend, and scale relational applications in the cloud with familiar tools.
e. ***SAP:*** Enterprise software giant SAP is now playing in the cloud along with HANA, a platform built on in-memory technology. Its cloud database from HANA which complements the company's other on-premise

database tools, including Sybase, and also available in Amazon Web Services cloud. HANA includes other non-database apps ,along with business management tools and application development.

f. ***Xeround:*** Xeround (pronounced zeh-round) is a management tool for deploying easily scalable MySQL databases across a variety of cloud providers and platforms. It's software allows for high availability and scalability and it works across a variety of cloud providers which include AWS, Rackspace, Joyent and HP, as well as on OpenStack and Citrix platforms.

V. CONCLUSION

Database Virtualization, by virtue of the rich benefits it delivers, has been an unstoppable force. Cloud computing has accelerated the adoption wave. The success of server virtualization has led to virtualization of storage, networking and now databases.

At this point, all aspects of the typical cloud environment can be completely virtualized with the sole exception of SQL databases. In this paper we studied, DBaaS architecture provides significant benefits in terms of increased security; efficiency and service levels while also providing greater transparency and flexibility with costs.

VI. REFERENCES

[1]. C. Curino, E. Jones, R. Popa, N. Malviya, E. Wu, S. Madden, H. Balakrishnan, and N. Zeldovich. Relational Cloud: A Database Service for the Cloud (In CIDR, pages 235–240), 2011.

[2]. C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield. Live migration of virtual machines (In NSDI, pages 273–286), 2005.

[3]. C. A. Waldspurger. Memory resource management in vmware esx server (In OSDI, pages 181–194), 2002.

[4]. A. Gulati, C. Kumar, and I. Ahmad. Storage Workload Characterization and Consolidation in Virtualized Environments (In VPACT), 2009.

[5]. A. Gulati, G. Shanmugathan, I. Ahamad, C. waldspurger, and M. Uysal. Pesto: Online Storage Perfromance Management in Virtualized Datacenters (In SoCC, pages 19:1–19:14), 2011.

[6]. Z. Liu, H. Hacig¨um¨us, H. J. Moon, Y. Chi, and W.-P. Hsiung. Pmax: tenant placement in multitenant databases for profit maximization (In EDBT, pages 442–453), 2013.

[7]. V. Narasayya, S. Das, M. Syamala, B. Chandramouli, and S. Chaudhuri. Sqlvm: Performance isolation in multi-tenant relational database-as-a-service (In CIDR), 2013.

[8]. Database Virtualization – "The Next Wave of Virtualization "

URL:
http://www.scaledb.com/pdfs/DBMS_Virtualization.pdf