



Database Performance Variations in Virtualized Cloud Environment

SUMANTH KUMAR. VV¹, RAJANI. M², SHANTHI. D³

¹Ph.D Scholar, Dept of MCA, Bharath University, Chennai, India.

²Director (Research), Bharath University, Chennai, India.

³Professor, Bharath University, Chennai, India.

Abstract: A common and powerful method for organizing data for computerization is the relational data model. Use of this model often results in a database with many tables, and a common question is why such a complex structure should be necessary. There are so many commercial database vendors are available in market as of today and Oracle and SQL Server are the leaders in this specific domain. Huge amount of resources are invested to manage and maintain databases which server hardware costs and license costs for production as well as test data environment. Cloud computing is gradually occupying the front stage and its core concepts appear to be sinking in all the fields including database applications. A Database ecosystem supported by cloud infrastructure would be a major boon for software companies who are involved in developing products as well as providing IT Enabled services, where they can scale up and scale down as per their requirement. As the machines hosting databases run at server, the maintenance and administration of databases becomes very easy and convenient. This cloud based database ecosystem will be very useful and it takes the DBA tasks to the next level. The performance of the Cloud based database found to be highly acceptable and almost similar the performance of databases runs under dedicated physical server ecosystem.

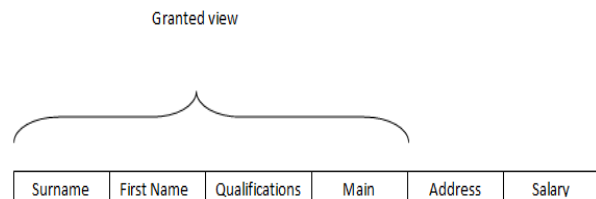
Keywords: E-Learning, Cloud Campus, Learning Ecosystem.

I. INTRODUCTION

Databases allow users to insert, update, delete and retrieve data from the database through a Data Manipulation Language (DML). Having a central repository for all data and data descriptions allows the DML to provide a general enquiry facility to this data, called a query language. Using a query language, directly or indirectly, enables new lines of enquiry to be constructed and satisfied quickly. The DBMS also provides a view mechanism that allows each user to have his or her own view of the database. The DDL is used to define a view that is a subset of the database. For example, a program to print a list of staff names, their qualifications and subjects that they teach would be granted a view of the database that included just these data items and excluded all others as shown in the following Figure. Thus enables secured and restricted view of datasets.

However for achieving all the required features implemented effectively by databases, it is imperative that database need to be well designed, configured and administered by a Data Base Administrator [DBA]. It is the DBA responsibility to see proper backups are enabling through automation so that when there are issues the data can be restored and recovered from

backups and online redo log files. Most of the administration of databases can be automated once we place the database in cloud and this can called as Cloud Database. A Cloud Database is a database that typically runs on a cloud computing platform, such as Amazon EC2, GoGrid, Salesforce and Rackspace. There are two common deployment models: users can run databases on the cloud independently, using a virtual machine image, or they can purchase access to a database service, maintained by a cloud database provider. Of the databases available on the cloud, some are SQL-based and some use a NoSQL data model.



Restricting an application's view of the database

Fig1.

There are two primary methods to run a database on the cloud:

1. Virtual machine Image: cloud platforms allow users to purchase virtual machine instances for a limited time. It is possible to run a database on these virtual machines. Users can either upload their own machine image with a database installed on it, or use ready-made machine images that already include an optimized installation of a database. For example, Oracle provides a ready-made machine image with an installation of Oracle Database 11g Enterprise Edition on Amazon EC2¹.

2. Database as a service (DBaaS): some cloud platforms offer options for using a database as a service, without physically launching a virtual machine instance for the database. In this configuration, application owners do not have to install and maintain the database on their own. Instead, the database service provider takes responsibility for installing and maintaining the database, and application owners pay according to their usage. For example, Amazon Web Services provides three database services as part of its cloud offering, SimpleDB, a NoSQL key-value store, Amazon Relational Database Service, an SQL-based database service with a MySQL interface, and DynamoDB. A third option is managed database hosting on the cloud, where the database is not offered as a service, but the cloud provider hosts the database and manages it on the application owner's behalf. For example, cloud provider Rackspace offers managed hosting for MySQL on dedicated² and cloud architectures³ and NoSQL databases via Object Rocket's managed MongoDB service⁴.

II. METHODOLOGY AND ARCHITECTURE

Our aim is to compare the performance of the databases in traditional versus and cloud environment. For this we have considered Virtual Machine Images and installed Oracle database 10g version release 2 under a private cloud environment using VMWare trail version software. The idea behind selecting private cloud environment is to have a complete control for allowing effective evaluation of the performance of the databases. NAARM Cloud campus is developed by implementing Vmware⁴ vSphere RC5 environment for optimal utilization of existing resources. VMware Inc is generous enough to provide vSphere software on trail basis for 60 days for the purpose of evaluation. A control node has been established to provide a convenient single point of control to the data center. Control node unifies the resources from individual hosts to be shared among virtual machines [VMs] in the entire data center. The vCenter Server allows migrating running virtual machines between host servers so that one can perform hardware maintenance with minimal downtime. It also features allow us to balance machine workloads across hosts and manage virtual machines for high availability and disaster recovery.

A. Software Used

Host Operating system: VMware ESXi 5.0).
Server management Node: Windows2003 R2 Enterprise edition.

Cloud Management node Software: VMware vCenter Server 5 (Trial Version).

Cloud infrastructure: VMware vSphere5 (Trial Version)
VMware View 5 (Trial Version)

Database: Oracle 10g EE release 2 version 10.2.0.1

B. Hardware Used:

We have used our existing infrastructure of two Sun Blade x6270 M2 blade servers for implementing the cloud. The following figure-1 depicts the design of Hybrid Cloud of NAARM.

System Global Area: 1.7GB

Programmable Global Area: 555MB

SAN Storage

EMC CLARiON CX4 iSCSI SAN;

Flare operating system revision: 03.24.040.5.011

III. RESULTS AND DISCUSSION

Administering and managing a database is a business critical process, and also a complicated one. Database Administrators (DBAs) spend approximately 85% of their time on routine, repetitive tasks such as database installation, user administration, patching, running backups and health checks. Cloud tools focused around automation and self-management of these mundane tasks are key in making database administration less challenging.

A. Preparing a Cent OS Machine with Oracle 10g Enterprise server installed

1. Database Administrator

There are many web-based interfaces on the market that delivers full lifecycle management of critical database administration tasks in the cloud. It utilizes enterprise class software to automate tasks that previously took hours to perform to be complete in minutes with a click of a button. This automation platforms frees up DBA resources so they are able to focus on the most valuable and demanding activities. On average DBAs can cost between \$125K - \$150K per year, and typically manage 21 databases, giving us an average cost of \$7,000 per year per database. With these cloud based web interfaces, administrators are able to execute operations simultaneously against multiple databases allowing them to manage more databases while drastically increasing productivity. The "Patching & Upgrade" service allows user to patch multiple databases at a time, as efficiently as possible. It typically takes up to 1 hour to apply a patch to a single database. Patching 10 databases would require 10 man hours of work. Cloud allows us to simultaneously patch all 10 databases with a click of a button, resulting in only 10 minutes of work.

Database Performance Variations in Virtualized Cloud Environment

The “Backup & Restore” service allows user to backup multiple databases with no server logins required. Detailed backup history logs are available to ensure all of your databases are being backed up consistently.

Process: We need to insert Cent OS in the CD Drive of physical Computer. We may have to change the boot options, so that the system boots from Compact Disc. While installing Cent OS, we have to provide IP address and may have to select the packages that need to be installed inside the operating system. We select “Everything” for uniformity, in all cases.



Fig2.

The time taken for completing the installation of Cent OS & Oracle 10g Server in a medium instance is as follows: (Physical Machine)

TABLE I

S. No	Installing Cent OS	Preparatory steps for oracle installation	Installing Oracle 10g
1.	28 minutes	15 minutes	10 minutes
2.	26 minutes	18 minutes	9 minutes
3.	30 minutes	14 minutes	9 minutes
Mean	28 minutes	15.66 minutes	9.33 minutes

Therefore total time taken: 28+15.66+9.33=53 minutes.

Whereas the average time taken for creating a virtual machine by cloning a “Cent OS + Oracle 10g template” is 50 minutes. The following figure shows progress during the cloning process. Even though the time difference between both of the above described processes is very less, it is highly advised to go for cloning process as the template is proven & devoid of any errors. Also, the time spent by administrator is

negligible as he just has to start the cloning process. & no intervention is required.



Fig3.

Database applications running on individual physical servers represent a large consolidation opportunity. However enterprises considering such consolidation want guidance as to how well databases scale using virtualization.

B. Performing load test

The test described in this work were conducted by pushing load by executing export & import processes which are typically useful for data refresh & taking backups. We will use export job at full database level, schema level and table level to check the performance of database in physical and virtual machine. We wanted to demonstrate the CPU scale-out capabilities of ESX with each virtual machine running at 100 percent CPU utilization. To accomplish this, the Oracle database application had to read with near zero latency the data required to process user queries. Therefore, we selected the 1GB database size. This allowed the entire database to be cached in the Oracle System Global Area (SGA), thus avoiding any I/O latency arising from reads and writes to the physical disks. The server used was a Sun Blade x6270 M2 server with 16 CPU cores and 256GB of memory running VMware ESX 3.5. Between one and eight workload virtual machines, each with two virtual CPUs and 4GB of memory, were running 64-bit Red Hat Enterprise Linux 4, Update 4 and Oracle 10g R2 (10.2.0.1). Load was generated by the Oracle 10g R2 32-bit Windows client on a system natively running Windows Server 2003 Release 2 Enterprise Edition with SP2.

C. Export & Import utilities

The Export and Import utilities provide a simple way for us to transfer data objects between Oracle databases, even if they reside on platforms with different hardware and software configurations. When we run Export against an Oracle database, objects (such as tables) are extracted, followed by their related objects (such as indexes, comments, and grants), if any. The extracted data is written to an export dump file. The Import utility reads the object definitions and table data from the dump file. An export file is an Oracle binary-format dump file that is typically located on disk or tape. The dump files can be transferred using FTP or physically

transported (in the case of tape) to a different site. The files can then be used with the Import utility to transfer data between databases that are on systems not connected through a network. The files can also be used as backups in addition to normal backup procedures. Export dump files can only be read by the Oracle Import utility. The version of the Import utility cannot be earlier than the version of the Export utility used to create the dump file.

We used the performance of a single virtual machine configured with two virtual processors as the reference score. We then compared the performance of additional identically configured virtual machines to this reference score as they were added to the ESX system. We ensured that the virtual CPUs of each of these virtual machines were fully saturated during the tests. Once they reached a steady state, we collected the transactions per minute and response times of the workload in each virtual machine as well as the average CPU utilization across all physical CPUs during a test cycle. This allows us to show variations in transactions per minute (TPM) from the baseline performance as fractions of the reference score when additional load is added to the ESX system. These fractions are plotted as the normalized values shown in the following figures.

C. Scaling Performance

The following figure shows the normalized aggregate transactions per minute of all the virtual machines powered on during a test cycle, while the following figures show the individual performance of virtual machines during a test cycle. As seen from the following figure the aggregate TPM scales in a nearly linear fashion as virtual machines are added. The scaling tapers off only as the last virtual machines are added, because at that stage the resources (especially CPU) are nearing saturation.

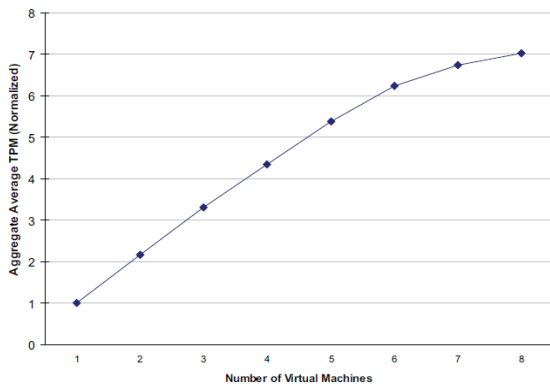


Fig4. Aggregate average transactions per minute.

As shown in the following figure, the normalized values of the number of transactions per minute performed by each workload virtual machine as additional workload virtual machines are powered on compare favorably to the baseline performance.

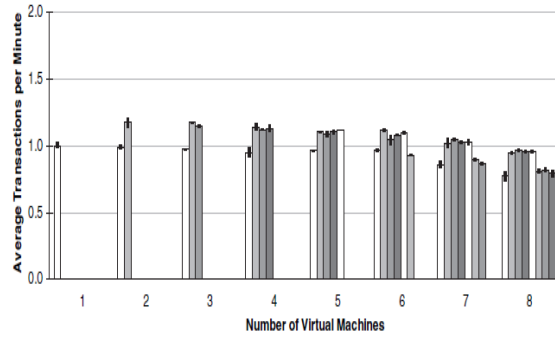


Fig5. Average transaction per minute

In the following figure, shows that as the number of virtual machines (and the corresponding load) on the ESX system increases, the database query response time shows only minimal degradation. With seven virtual machines, the response time of the lowest performing virtual machine is only 15 percent above the single virtual machine baseline, and even with eight virtual machines (and full CPU commitment), the response time does not increase by more than 21 percent. When it takes longer to process a transaction, the number of transactions per minute is lower. Therefore, virtual machines with higher transactions per minute in above figure correspond to virtual machines with lower response time in following figure.

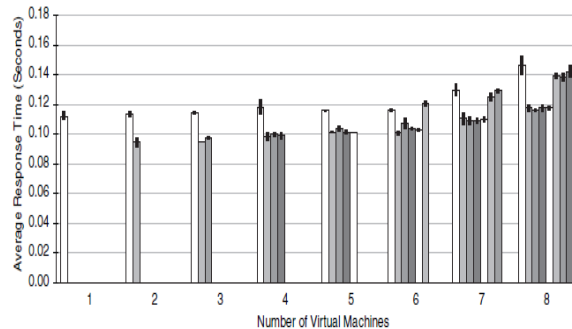


Fig6. Average response time

The performance of the virtual machines in these tests fell into two distinct groups, one group slightly higher, the other slightly lower. Each virtual machine’s performance corresponded to the NUMA node on which that virtual machine was running, a variance believed to be caused by the architecture. In addition, the eight virtual machine test fully committed the 16 physical CPUs. This distributed the overhead of processing interrupts for disk and network I/O across all CPUs instead of restricting them to free CPUs, as was done when fewer virtual machines were running. Our goal was to see how resource utilization scales as the number of virtual machines, and thus the overall load on the ESX host, increases. We also wanted to see the performance of the ESX scheduler when the CPU

Database Performance Variations in Virtualized Cloud Environment

resources were fully committed. The last two figures show that ESX efficiently allocates available resources to virtual machines as the demand increases. The following figure shows the average CPU utilization as seen by ESX. This includes the CPU cost of the virtual machines running the Oracle database as well as all virtualization overheads. The figure shows that as the number of virtual machines is increased there is a nearly linear scaling of CPU utilization on the ESX system, which indicates the efficiency of the ESX resource scheduler.

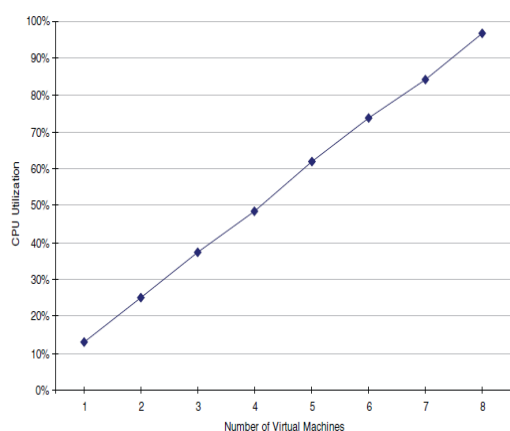


Fig7. CPU Utilization

IV. CONCLUSION

When considering the time taken for preparing Oracle database with Cent OS installed, Even though the time difference between both of the above described processes is very less, it is highly advised to go for cloning process as the template is proven & devoid of any errors. Also, the time spent by administrator is negligible as he just has to start the cloning process. & no intervention is required. It is well demonstrated that when running multiple virtual machines with Oracle database workloads on VMware ESX 3.5, the individual performance remains close to that of the Oracle database workload in a single virtual machine, while CPU utilization scales in a nearly linear fashion. Test was conducted with a 1GB DS2 database that was cached in Oracle's SGA in each virtual machine. The physical CPUs on the server were fully saturated and I/O latencies arising from read and writes to the physical disks were avoided. At saturation, CPU utilization was directly proportional to the number of virtual machines powered on. In the second phase we also observed that performance of the Oracle database in a virtual machine on ESX 3.5 was 94 percent of native. The tests confirm that Hypervisor is very efficient in scaling overall resource utilization with a very minimal effect on performance.

V. REFERENCES

- [1] Amazon Machine Images - Oracle Database 11g Release2 (11.2.0.1) Enterprise Edition - 64 Bit, Amazon Web Services, Retrieved 2011-11-9.
- [2] "MySQL Server Support at Rackspace", [www.rackspace.com Rackspace.com], Retrieved 2014-02-24.
- [3] "MySQL Databases on the Cloud at Rackspace", [www.rackspace.com Rackspace.com], Retrieved 2014-02-24.
- [4]"Managed MongoDB at Rackspace", [www.rackspace.com Rackspace.com], Retrieved 2014-02-24.
- [5] http://en.wikipedia.org/wiki/Cloud_database , Retrieved 2014-02-24.
- [6]http://en.wikipedia.org/wiki/Cloud_computing; accessed August 2011.
- [7]http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf; NIST of US Department of Commerce, accessed August 2011.
- [8] "Cloud Computing: Clash of the clouds". The Economist. 2009-10-15. Retrieved 2011-11-03.
- [9] www.oracle.com ; Retrieved 2011-11-03.
- [10] www.centos.org, Retrieved 2011-11-03.
- [11] Chris Kanaracus, "EnterpriseDB Adding New Cloud Option for PostgreSQL Database", http://www.pcworld.com/businesscenter/article/238775/enterprisedb_adding_new_cloud_option_for_postgresql_database.html, Retrieved on 9th NOV 2011.