Consolidating OLTP Workloads on Dell™ PowerEdge™ R720 12th generation Servers

B Balamurugan

Phani MV

Dell Database Solutions Engineering

March 2012

This document is for informational purposes only and may contain typographical errors and technical inaccuracies. The content is provided as is, without express or implied warranties of any kind.

© 2012 Dell Inc. All rights reserved. Dell and its affiliates cannot be responsible for errors or omissions in typography or photography. Dell, the Dell logo, and PowerEdge are trademarks of Dell Inc. Intel and Xeon are registered trademarks of Intel Corporation in the U.S. and other countries. Microsoft, Windows, and Windows Server are either trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell disclaims proprietary interest in the marks and names of others.

March 2012| Rev 1.0

Contents

Executive summary 5
Audience 5
Introduction 5
Summary of goals 6
PowerEdge R720 architecture 6
Test configuration 7
Performance metrics 8
Performance testing utility 8
Scale factor 8
Test methodology 9
Legacy PowerEdge 2950 III results 10
PowerEdge R720 results 13
Comparison analysis of the PowerEdge 2950 and PowerEdge R720 15
Consolidation factor 17
Summary 18

Tables

Table 1. Test configuration 7

Figures

Figure 1. AQRT comparison of legacy production environment for different test iterations 10
Figure 2. CPU utilization comparison of legacy production environment for different test iterations 11
Figure 3. TPS comparison of legacy production environment for different test iterations 12
Figure 4. TPS comparison of PowerEdge R720 test environment with different test iterations 13
Figure 5. AQRT comparison of PowerEdge R720 test environment with different test iterations 14
Figure 6. CPU time comparison of PowerEdge R720 test environment with different test iterations 14
Figure 7. TPS comparison between legacy production and PowerEdge R720 at legacy saturated userload 15
Figure 8. Response time comparison between legacy production and PowerEdge R720 test environment at legacy saturated userload 16
Figure 9. CPU time comparison between legacy production and PowerEdge R720 test environment at legacy saturated userload 16
Figure 10. PowerEdge R720 power efficiency: savings from consolidation 17

Executive summary

The Dell™ enterprise portfolio is evolving to incorporate better-performing, more energy-efficient, and more highly-available products. With the introduction of Dell's latest server product line, customers have an opportunity to improve their total cost of ownership by consolidating distributed legacy environments.

Dell strives to simplify IT infrastructure by consolidating legacy production environments to reduce data center complexity while still meeting customers' needs. The tools and procedures described in this white paper can help administrators test, compare, validate, and implement the latest hardware and database solution bundles. Dell established these procedures and guidelines based on lab experiments and database workload simulations performed by the Dell Database Solutions Engineering team. Using the tools and procedures described in this document, customers may not only select the appropriate database solution hardware and software stack, but also optimize the solution to help optimize total cost of ownership according to the database workloads they choose to run.

In short, the focus of whitepaper is:

1. To consolidate OLTP[1] workloads on PowerEdge R720 servers.
2. To optimize the TCO of Dell solutions.
3. To highlight the benefits of consolidation.

Audience

The intended audience of this white paper includes database administrators, IT managers, and system consultants.

Introduction

Server consolidation can be defined as maximizing the efficiency of server resources, thereby minimizing the infrastructure, associated power and cooling, rack footprint and licensing costs. It essentially solves a fundamental problem— called server sprawl—in which multiple, underutilized servers take up more space and consume more power resources than the workload requirement indicates.

OLTP database systems typically service hundreds or thousands of concurrent users. An example of this type of system could be a travel reservation system with large number of customers and agents performing online travel reservations, or checking available flights or flight schedules. The OLTP database transactions performed by these thousands of concurrent users get translated into tens of thousands of I/O requests to the backend storage subsystem. Depending on the nature of these OLTP transactions, the database host CPUs may only be efficiently utilized if the backend storage subsystem is configured with a sufficient number of disks to handle the large number of I/O requests. Otherwise the database host CPUs exhibit large IOWAIT times instead of doing useful work. In this scenario,

---

TPC-C is an on-line transaction processing benchmark.
[1] http://tpc.org/tpcc/default.asp

consolidating, upgrading or migrating to a faster database server, or scaling the number of CPUs or memory does not help. The correct approach is to appropriately scale the backend disk subsystem to handle the I/O requests first, and then to move to the next stage of CPU and memory sizing as discussed later in this white paper.

The objective of this white paper is to consolidate an Oracle database running OLTP workloads on a legacy 9th Generation Dell PowerEdge 2950 2U 2-socket server to the new 12th Generation Dell PowerEdge R720 2U 2-socket server.

Summary of goals

- To determine if a multi-node Oracle RAC cluster running on a legacy PE 2950 system can be replaced/consolidated with an Oracle RAC cluster consisting of fewer nodes on a PowerEdge R720 server.
- To find out the consolidation factor at various CPU usage levels.
- To highlight the power savings that can be realized by consolidation.

PowerEdge R720 architecture

The PowerEdge R720[2] is a new 12th generation rack server product. It is one of the best Data center optimized servers, designed with improved performance in areas such as virtualization, power consumption, systems management and usability.

The PowerEdge R720 is 2U rack mount server includes support for the Intel Xeon® Processor E5-2600 Family (Sandy Bridge-EP architecture). This family of 6/8 core processors includes an integrated memory controller (IMC) and integrated IO (IIO) on a single silicon die. This new server supports PCIe generation 3 slots for peripherals, two internal raid controller cards, and can host memory up to 768GB.

Networking support on the PowerEdge R720 includes a rack network daughter card (rNDC). Unlike traditional LOM, rNDC offers greater flexibility in terms of expanded features, network types, speed and vendors. Also rNDC offers easy upgrade path from 1G to 10G speed.

This server also offers options such as support for the latest Dell systems management software such as Life Cycle Controller 2.0 and iDRAC.

Dell PowerEdge Servers
[2] http://www.dell.com/us/enterprise/p/poweredge-rack-servers

Test configuration

Table 1 describes the complete software and hardware configuration that was used throughout testing on both the simulated legacy production environment and the 12G test environment.

Table 1. Test configuration

Component	Legacy Production Environment	12G R720 Test Environment
Systems	Two PowerEdge 2950 III,2U servers	One PowerEdge R720 2U 2 socket server
Processors	Two Intel Xeon X5460, 3.16 GHz quad core per node Cache: L2=2x4M per CPU	One Intel Xeon E5-2690, 2.90 GHz eight core Cache: L2=2MB L3=20MB
Memory	32 GB DDR2 per node (64 GB total)	64 GB DDR3
Internal disks	Two 73 GB 2.5" SAS per node	Two 73 GB 3.5" SAS
Network	Two Broadcom® NetXtreme II BCM5708 Gigabit Ethernet	Four Broadcom® NetXtreme II BCM5720 Gigabit Ethernet
External storage	Dell PowerVault MD3620f with 146GB SAS disks	Dell PowerVault MD3620f with 146GB SAS disks
HBA	One QLE2462 per node	One QLE2562
OS	Enterprise Linux® 4.6	Enterprise Linux 5.7
Oracle software	• Oracle 10g R2 10.2.0.4 • File System: ASM • Disk groups: DATABASE, DATA • sga_target = 2000M • pga_target = 1000M	• Oracle 11g R2 11.2.0.3.0 • File System: ASM • Disk groups: DATABASE, DATA • memory_target = 6000M
Workload	• Quest Benchmark Factory TPCC workload • Scale factor: 3000 • User connections: 200-5000	• Quest Benchmark Factory TPCC workload • Scale factor: 3000 • User connections: 200-5000

Performance metrics

During the testing process we collected the following 5 metrics for further study and analysis:

1. TPS (transaction per second)
2. AQRT (average query response time)
3. CPU utilization
4. Memory utilization
5. Power consumption.

For an OLTP environment the most commonly used metrics are transaction per second (TPS) and Average query response time (AQRT). Average query response time of an OLTP database environment may be described as the average time it takes for an OLTP transaction to complete and deliver the results of the transaction to the end user initiating that transaction. The average query response time is the most important factor when it comes to fulfilling end user requirements, and it establishes the performance criteria for an OLTP database. A 2 seconds response time metric was chosen as the basis for our Service Level Agreement (SLA) which was maintained throughout the testing.

Along with TPS and AQRT, CPU and memory utilization data was also collected during the testing. CPU utilization data is of help to know the saturation levels of database server and indicates how the server is scaling and delivering for increasing load. This data is useful to learn about performance bottlenecks from a server point of view.

Lastly, power consumption data helps to figure out the savings one can achieve by consolidating PowerEdge 2950 servers to PowerEdge R720.

Performance testing utility

We used Quest Software® Benchmark Factory®[3], a load-generating utility that simulates OLTP environment and transactions on a database for a given number of users. This tool can be used to create database tables, load data into the schema and stress test the database server. This utility provides options to simulate standard workloads such as TPC-C, TPC-H, and TPC-E. Based on their needs a user can select the type of workload to be simulated. We chose TPCC which simulates an OLTP workload. The TPCC workload provided by the Benchmark Factory schema simulates an order entry system consisting of multiple warehouses. After choosing the workload type, benchmark utility loads data based on scale factor.

Scale factor

Quest Benchmark Factory® provides an option to load data based on the need. For this the tool offers a unit called scale factor. When setting up a load test Benchmark Factory allows changing the scale factor. This unit decides how many rows of data have to be inserted into the schema. Ultimately the number of rows contributes to the overall database size. An increase in scale factor allows for

---

Benchmark Factory® for Databases is a database performance testing tool
[3] http://www.quest.com/benchmark-factory/

increases in the database size, allowing larger userloads to be used to place a greater stress on the system under test. Hence scale factor plays a major role in simulating workloads. For example, a scale factor of 3000 contributes to a database size of ~290GB. Throughout the testing process we have used 3000 scale factor.

Test methodology

Our consolidation study is a four-step process:

1. Find out the maximum performance delivered by the legacy PowerEdge 2950 III.
2. Find out the system behavior of the PowerEdge R720 by simulating a legacy saturated workload.
3. Find out the scalable performance of the PowerEdge R720.
4. Comparative analysis of performance between the legacy PowerEdge 2950 III and PowerEdge R720.

Initially the backend storage subsystem, consisting of a Dell PowerVault MD3620f storage array, was configured with ten 15K RPM 136GB disks in RAID 10 configuration. Once the data was populated, we started testing with 200 concurrent users and increased the user load to 5000 in increments of 200 users, randomly running transactions against the legacy database while making sure that the average query response time always stayed below 2 seconds.

The test methodology used was as follows:

1. To simulate the legacy production environment, we configured a two node Oracle 10g R2 RAC cluster comprising of a PowerEdge 2950 III with quad-core, dual socket 3.16 GHz CPU, connected to a Dell PowerVault MD3620f configured with a 100 GB LUN for the database SYSTEM and a 400 GB LUN for DATA ASM disk groups and a 2 GB LUN for the voting and Oracle Cluster Registry (OCR) partitions.
2. Using the Quest Benchmark Factory TPCC workload profile, we populated the data with a scale factor of 3000 into the simulated legacy server production environment.
3. To find out the saturation point of the legacy production environment, we started the first test iteration with 10 disks for the DATA ASM disk group and 200 userload. The user load was then increased in 200 user increments while constantly monitoring the average query response time. Once the average query response time crossed 2 seconds, the test was stopped.
4. To determine host performance limiting factor, the host CPU time analysis was carried out for all test iterations. For experimental purposes we have considered 90% or more CPU utilization as the performance limiting factor point. If CPU utilization is below this limiting factor we tried to find if there was any saturation bottleneck from storage.
5. Once the backend spindles were saturated, they started exhibiting large I/O latency which resulted in large IOWAIT at the host CPU and a large average query response time. To reduce the IOWAIT at the host CPU, the numbers of spindles were increased by 10 disks for the DATA ASM disk group for the next iteration performed. The above methodology was continued until the host CPU was optimally utilized with a smaller IOWAIT time. At the same time we monitored whether it could support a higher user load compared to the earlier iteration.

6. To simulate our test environment, we configured an Oracle 11g R2 single node RAC comprised of a PowerEdge R720 server populated with one 8 core Socket. Using the Quest Benchmark Factory, we populated the test data with the same TPCC scale factor used for the legacy production environment.
7. The test iterations similar to the legacy production environment were carried out on the R720 test environment to do a comparative study.

Legacy PowerEdge 2950 III results

As explained in the test methodology, we started testing the legacy environment with 10 disks. Based on the initial results, to find out the saturation point we performed two more iterations with 20 and 30 disks. Figure 1, 2 and 3 below shows a comparison in AQRT, CPU utilization and TPS of the legacy production environment for different test configurations with 10, 20 and 30 disks.

Figure 1. AQRT comparison of legacy production environment for different test iterations

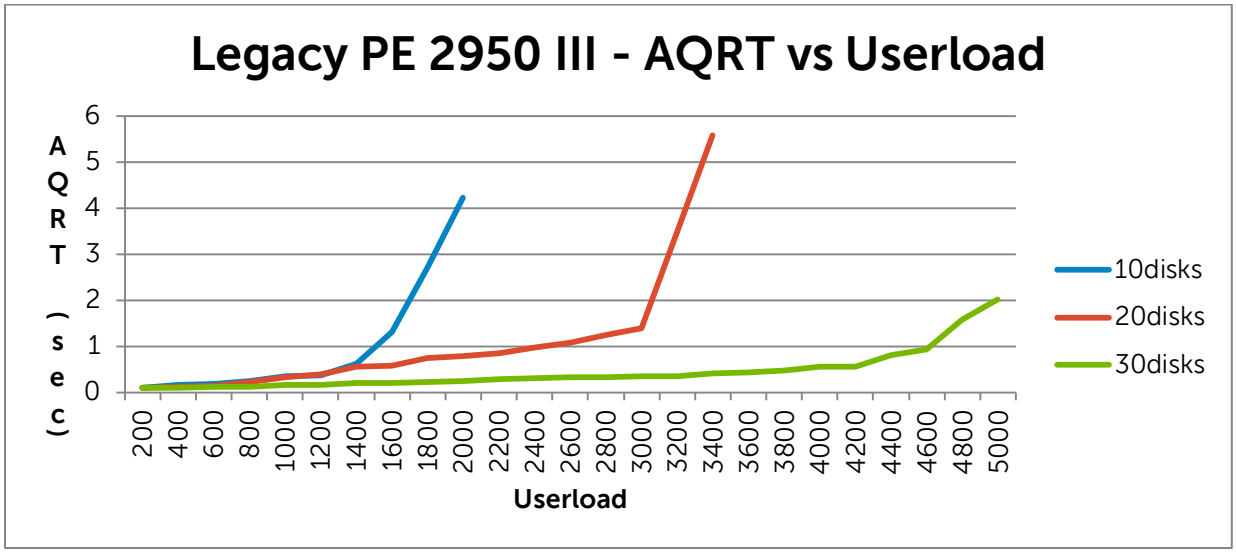
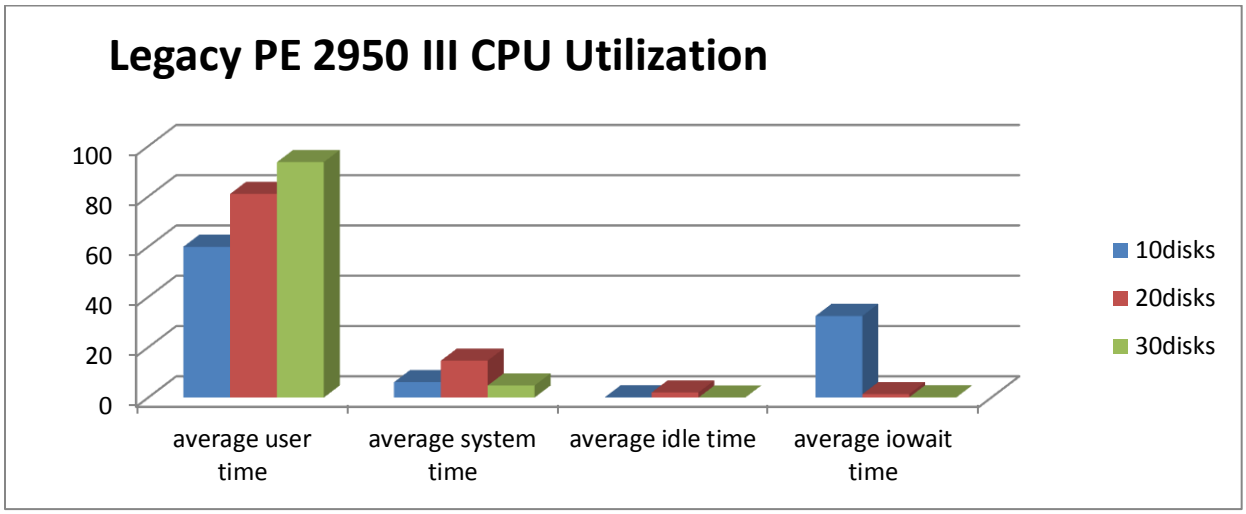


Figure 2. CPU utilization comparison of legacy production environment for different test iterations



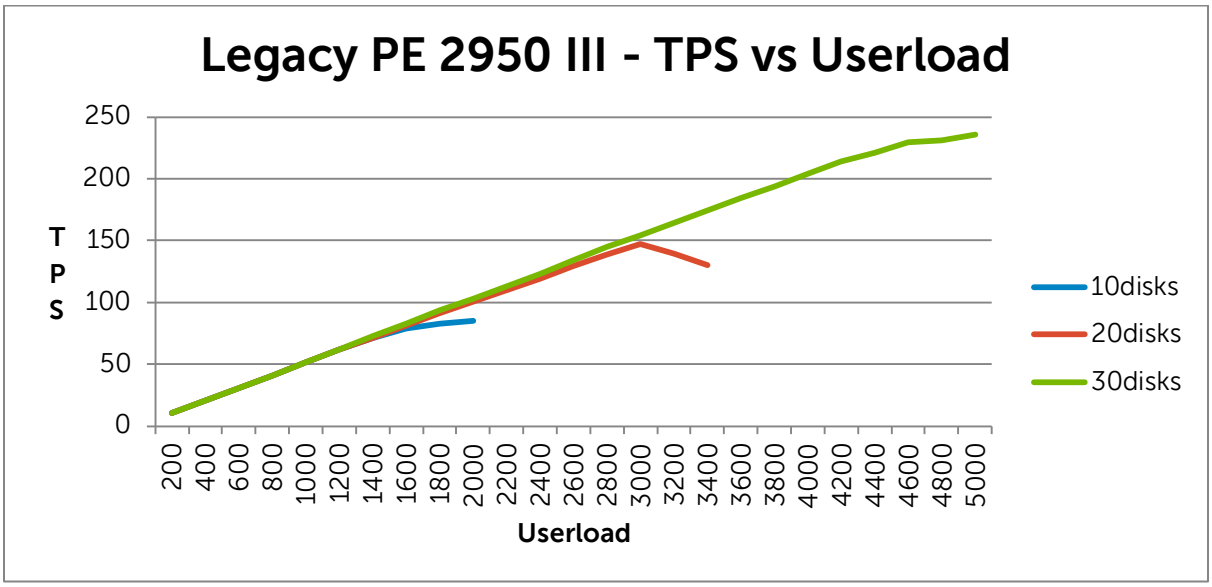
From Figure 1 and 2, it is observed that the legacy production environment performed better when the numbers of disks are increased to 20 and 30. It is also clear in the first test iteration with 10 disks that the AQRT crossed 2 seconds SLA after 1600 userload and corresponding CPU utilization was ~59%. At the same time we can also see that there is a huge IOWAIT of 30%, which means that storage is a bottleneck.

For better CPU utilization and to remove this storage bottleneck we increased the number of drives to 20. With 20 drives we performed the second test iteration. it is evident that at 3000 userload the system crossed 2 sec SLA, at the same point the CPU utilization is ~81% which is still less than our saturation point.

We then went ahead and increased the drives to 30 and performed a third test iteration. The results reveal that the system has hit CPU utilization of ~93% which can be treated as the saturation point. In addition, increasing the number of drives will not help because IOWAIT is negligible. At the same time, looking at the AQRT graph of 30 disks test iteration it is clear that the SLA is less than 2 sec.

At this moment we concluded that we saturated the legacy PowerEdge 2950 III, and the max userload supported was 4800 with 93% CPU utilization.

Figure 3. TPS comparison of legacy production environment for different test iterations



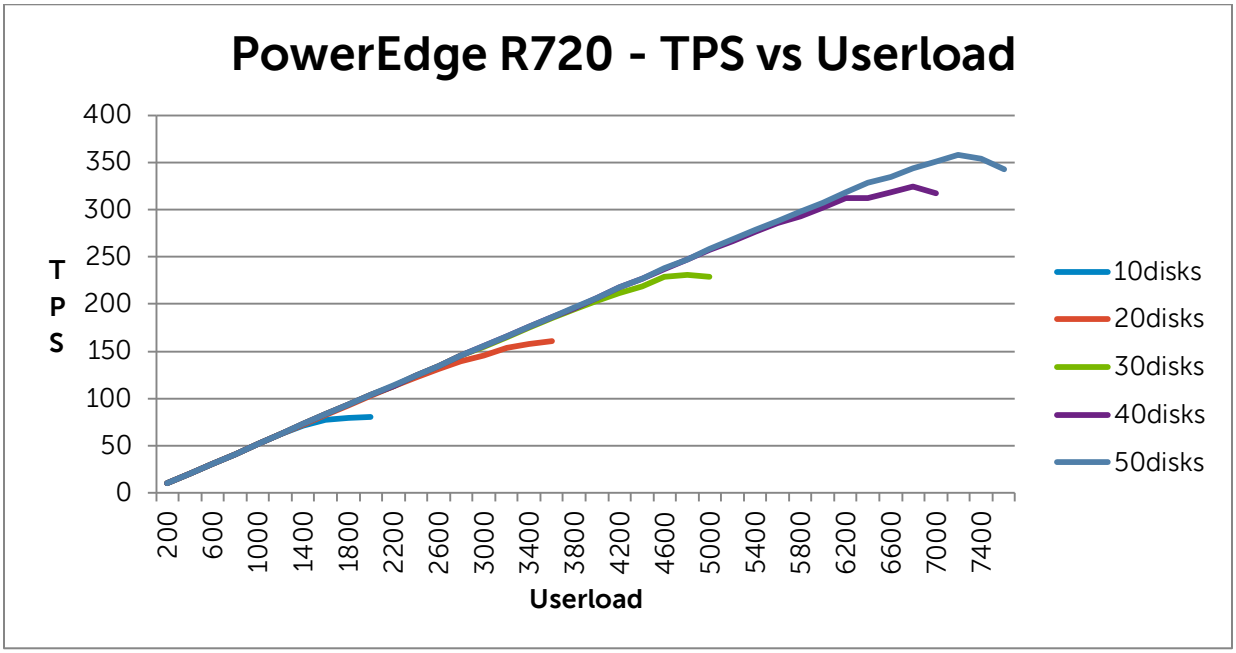
As mentioned earlier, the TPS metric plays a big role in OLTP workloads. During the test iterations while capturing AQRT & CPU utilization, we also captured TPS data. The TPS graph helps to determine whether the system is scaling to deliver more TPS with increasing userload.

From Figure 3 we can notice that the server scaled up and delivered more TPS. It is evident that increase in storage spindles complemented system performance, thus resulting in more userload support and TPS.

PowerEdge R720 results

During the consolidation study we performed five test iterations on the PowerEdge R720 with 10, 20, 30, 40 and 50 disks in the storage. Figures 4, 5 and 6 below show a comparison of TPS, AQRT and CPU utilization of the PowerEdge R720 test environment.

Figure 4. TPS comparison of PowerEdge R720 test environment with different test iterations



As seen in Figure 4, the PowerEdge R720 showed an improvement with every iteration (increased spindle count) in terms of TPS delivered and userload supported. The maximum userload supported with 10 disks was 1600. The userload supported increased with the increase of storage disks. The maximum userload supported for 20, 30, 40 and 50 disks are 3600, 4800, 6800 and 7400 respectively. Figures 5 and 6 show more about the server performance and the need of increasing spindle count with each test iteration.

Figure 5. AQRT comparison of PowerEdge R720 test environment with different test iterations

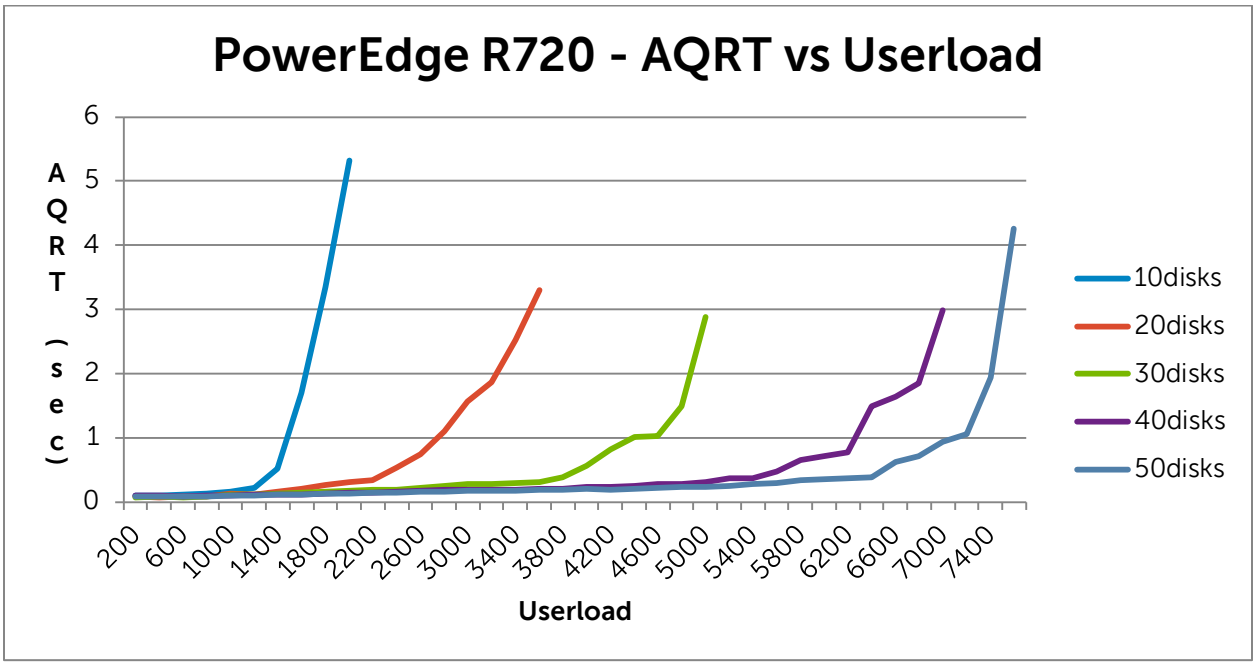
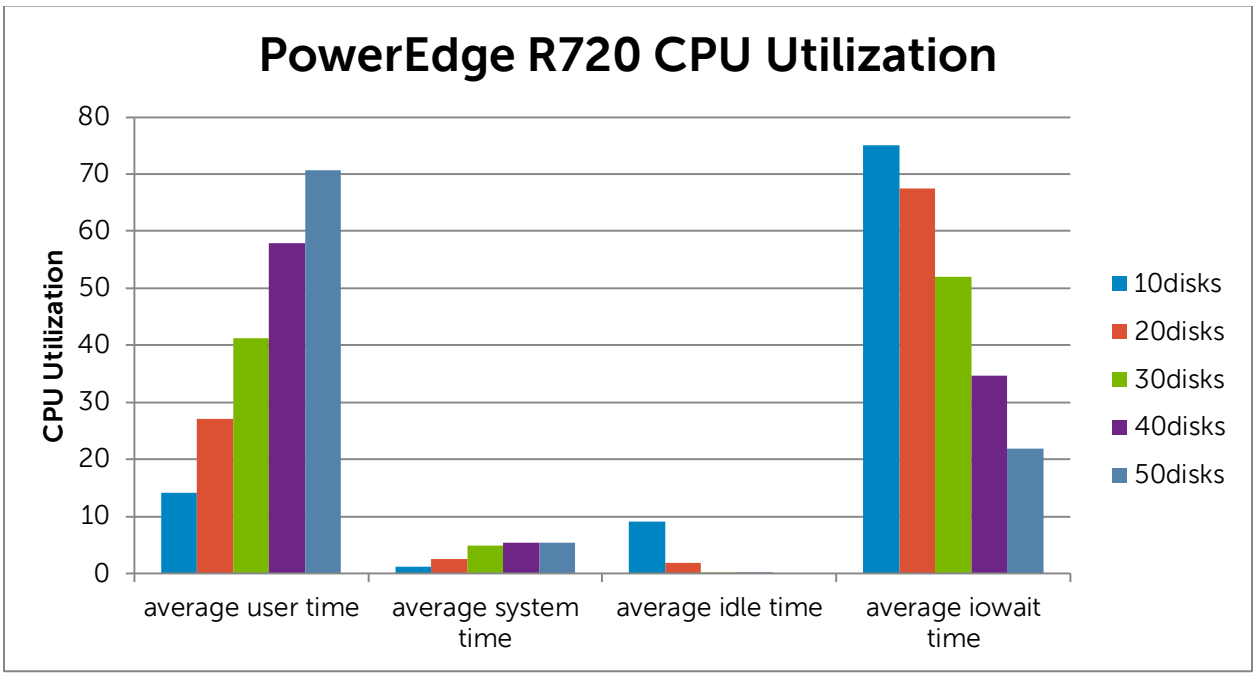


Figure 6. CPU time comparison of PowerEdge R720 test environment with different test iterations



Figures 5 and 6 show the results of AQRT and CPU time for 5 different iterations. From the above graphs it is observed that for the first iteration the max userload supported is 1600 and corresponding CPU time is 14% at 2sec SLA. On further analysis of CPU data it evident that there is a huge IOWAIT of ~70% which proves that storage is a bottleneck.

Now we increased the drives to 20 and performed the second test iteration. From the graph it can be found that 3400 userload is supported in this iteration with 27% CPU utilization. Still the IOWAIT time is 60%. Following our testing methodology we went ahead and performed 3 more iterations with 30, 40 and 50 disks. It is clearly evident that R720 improved in AQRT and performance with the increase of storage disks.

From the results we can say that a 7400 userload was supported with 70% CPU utilization and the system has some more power to do productive work by reducing IOWAIT.

Comparison analysis of the PowerEdge 2950 and PowerEdge R720

In the previous sections we have seen the results of performance study, separately for the PowerEdge 2950 and PowerEdge R720. Now we shall shed some light on the comparison analysis of PowerEdge 2950 versus PowerEdge R720. For the comparative study it will be appropriate to choose the point or test iteration where the legacy 2950 delivered maximum performance.

Earlier we found that the PowerEdge 2950 saturated at a maximum userload of 4800. Figure 7 and 8 below show the comparison of the test results for the PowerEdge R720 test environment and the legacy production environment, at the legacy saturated user load 4800.

Figure 7. TPS comparison between legacy production and PowerEdge R720 at legacy saturated userload

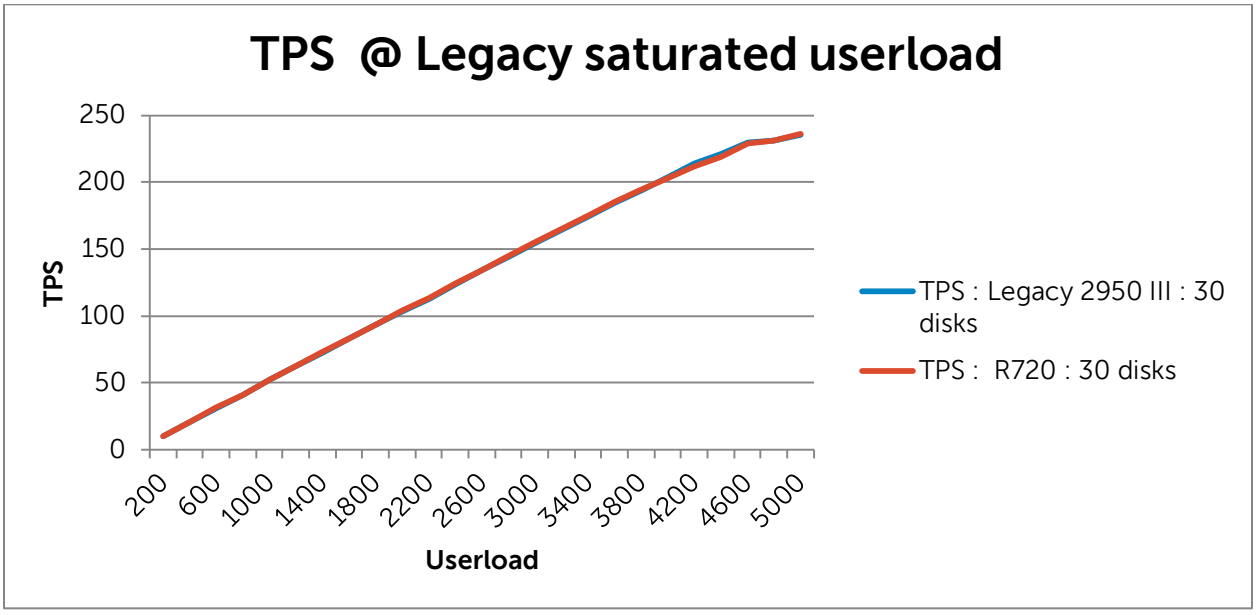
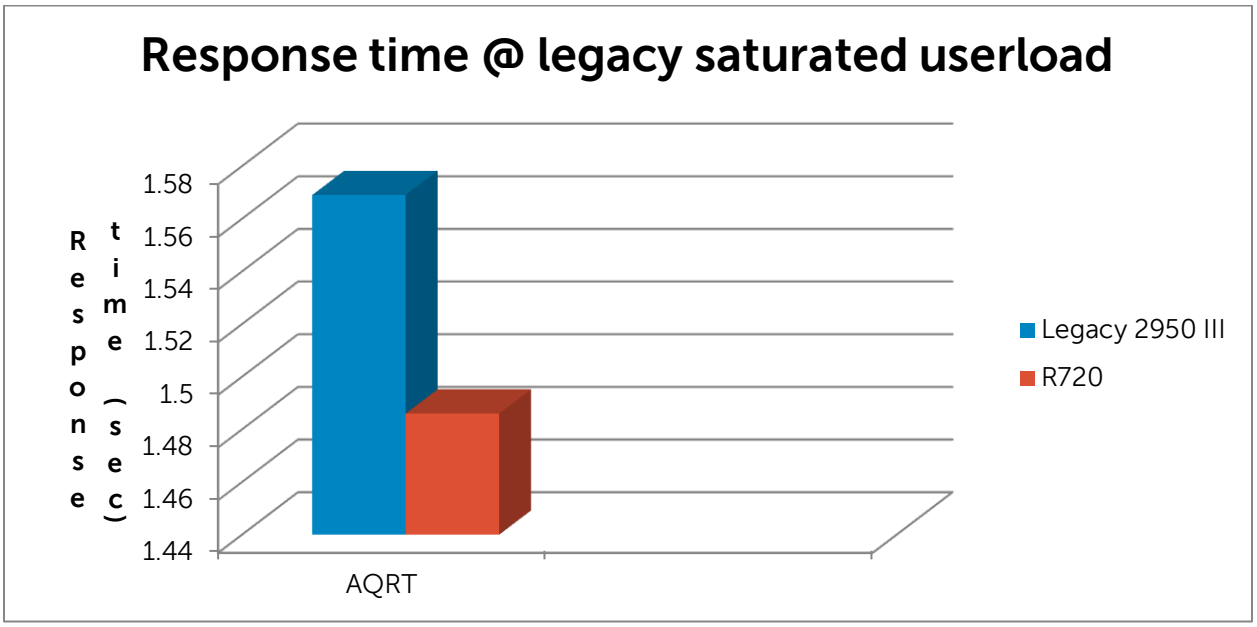
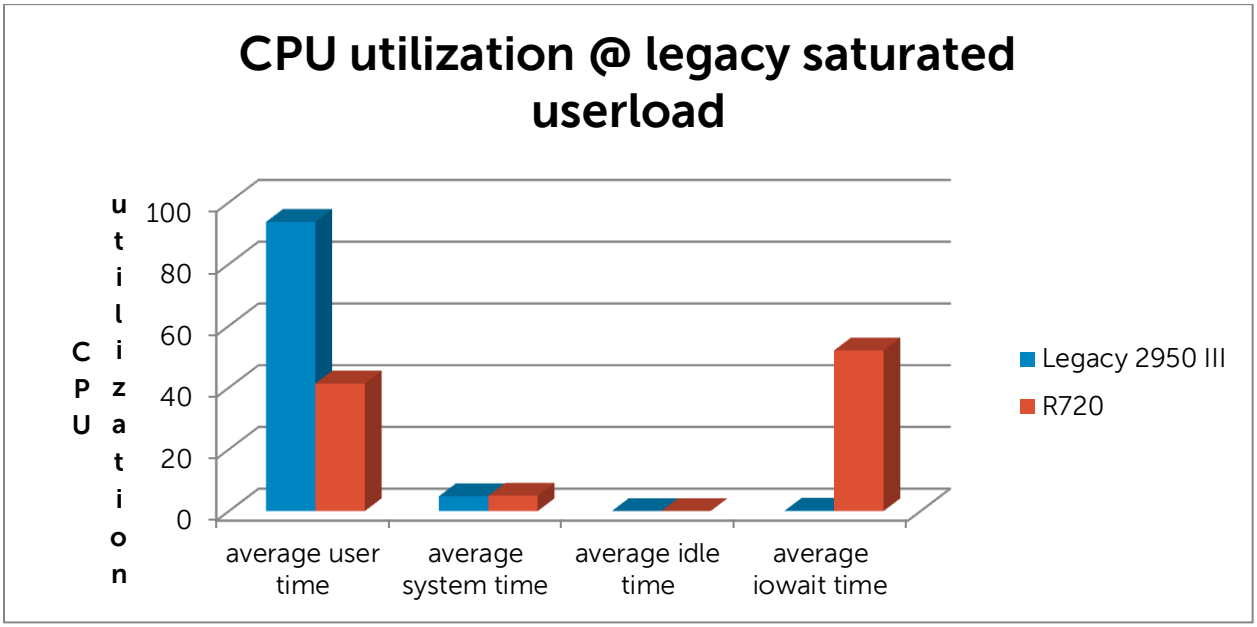


Figure 8. Response time comparison between legacy production and PowerEdge R720 test environment at legacy saturated userload



As seen in Figure 7, at the legacy saturated user load the TPS on both the environment is similar. From Figure 8 we can see that the response times of legacy 2950 and R720 servers are similar (only a difference of .08 sec) and below the 2 sec. SLA.

Figure 9. CPU time comparison between legacy production and PowerEdge R720 test environment at legacy saturated userload



On analysis of the CPU time in Figure 9, it is observed that the average CPU user time for the PowerEdge R720 test environment is about 42% whereas on the legacy PowerEdge 2950 environment it is about 90%.

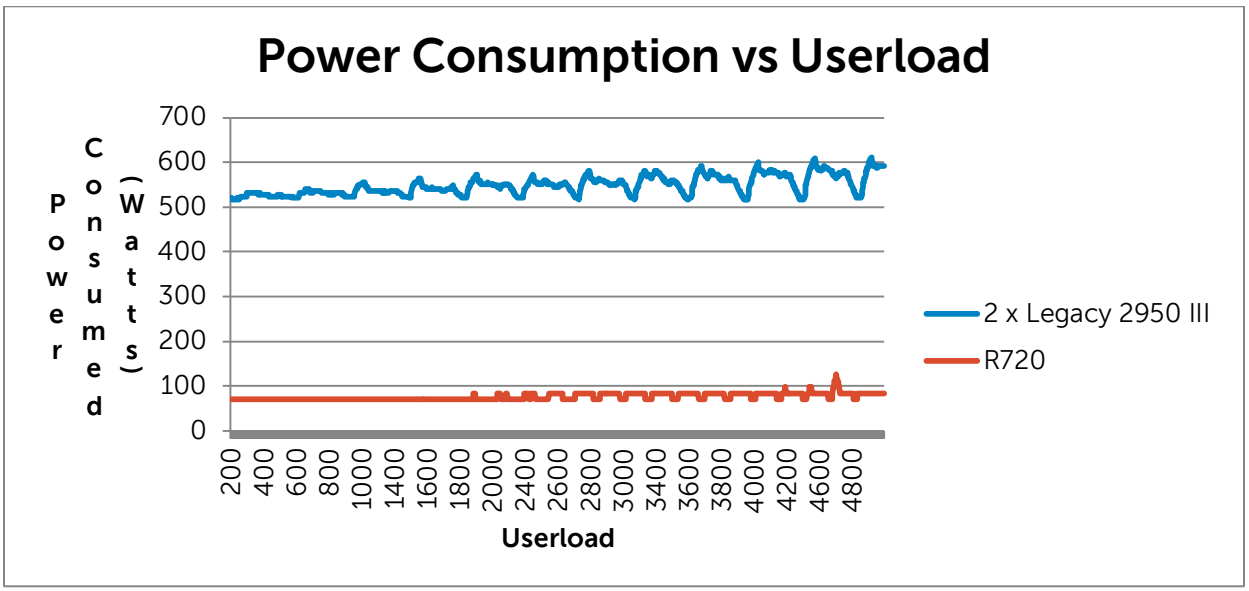
In addition, the CPU IOWAIT time on the PowerEdge R720 test environment is about 50% of the total CPU time whereas on the legacy production environment it is about 1%.

Consolidation factor

Based on the comparative analysis one can derive the following possible consolidating options:

1. **2:1 Consolidation at 42% Utilization.** A single-node Oracle 11G R2 RAC running on a PowerEdge R720 populated with 1 socket with 42% utilization was able to handle the OLTP workload of two nodes Oracle RAC running on 9th generation PowerEdge 2950 III servers.
2. **4:1 Consolidation at 90% Utilization.** If we utilize 90% of a R720 server, then it can consolidate the OLTP workload of 3-4 Oracle RAC nodes running on 9th generation PowerEdge 2950 III servers, provided that both the environments are configured with sufficient host memory and I/O disk subsystems.
3. **8:1 Consolidation at full Utilization.** Based on the preceding consolidation, a PowerEdge R720 populated with 2 sockets can consolidate the workload of 6 to 8 Oracle RAC nodes running on 9th generation PowerEdge 2950 III servers.

Figure 10. PowerEdge R720 power efficiency: savings from consolidation



The 12th generation servers are the most efficient ever in the PowerEdge lineup. Because of their special design, there will be huge power savings by consolidating from legacy servers to the PowerEdge R720. From Figure 10, we can clearly see that there is a huge difference in the power consumption between a PowerEdge R720 and a 2 node legacy 2950 production environment at various userload levels.

Let's look at the average power savings that can be achieved:

- Average power consumption by PowerEdge R720 - 80 watts
- Average power consumption by the legacy 2950 environment - 540 watts
- Power savings % = (540-80)*100 / 540 = 84%

Hence on an average 84% of power savings can be achieved.

Summary

Any server consolidation project must be preceded by a well-planned effort to predict the energy consumption and performance capacity of the new platform as compared to the legacy environment.

In this study, we showed that the new PowerEdge R720 is an ideal platform for consolidation from legacy PowerEdge 2950 environments. This is apt as the server supports the latest Intel Sandy Bridge processor architecture and comes with latest features like PCIe gen 3 supports, enhanced memory support, and flexible network card support. We also showed the performance gains and power savings that can be achieved with consolidation.

Customers running Oracle 9i or 10g RAC environments on legacy servers and storage can follow the guidelines and procedures outlined in this white paper to consolidate power-hungry RAC nodes into fewer, faster, more energy efficient nodes. The resulting legacy RAC node consolidation can also drive down Oracle licensing costs, resulting in savings that you can use to fund additional backend storage resources to improve average query response time, implement disaster recovery sites and additional RAC testbed sites for application development and testing. The reduced number of nodes does not compromise performance when paired with PowerEdge R720 servers. The result is less cluster overhead, simplified management, and positive movement toward an objective of simplifying IT and reducing complexity in data centers.

Specific TCO possibilities include:

1. 2:1 to 8:1 consolidation possibility
2. 84% Power Consumption Savings
3. License fee savings from server consolidations