Hewlett Packard Enterprise

HPE Reference Architecture for Oracle 18c OLTP and OLAP workloads on HPE Superdome Flex and HPE 3PAR Storage

Contents

Executive summary

Businesses today demand faster transaction processing speeds, scalable capacity, consolidation and increased flexibility. Traditional architectures with separate transactional and analytical systems are complex, expensive to implement, and introduce data latency. These challenges can be addressed by combining online transaction processing (OLTP) with online analytical processing (OLAP). Doing so requires a large, scale-up system architecture that has the ability to meet the demands of the combined workloads.

The HPE Superdome Flex sets a new standard for scalability and expandability while ensuring flexibility for all transaction, analytical, and data warehouse workloads. HPE Superdome Flex coupled with HPE 3PAR StoreServ storage arrays is an ideal scale-up configuration. The ability to scale-up to 32 CPU sockets, 48 TB of memory, and virtually unlimited storage capacity means that this solution can meet the needs of the most demanding Oracle Database workloads. The large memory capacity available with this server allows taking advantage of Oracle In-Memory features, which can speed up analytical queries by orders of magnitude, while simultaneously processing transactions.

This Hewlett Packard Enterprise Reference Architecture demonstrates that the HPE Superdome Flex is capable of simultaneously handling OLTP and OLAP workloads with minimal impact on transaction rates and query response times.

In addition, the HPE Application Tuner Express (HPE-ATX) tool was utilized to provide optimal performance in a NUMA environment. Transaction throughput increased up to 67% when HPE-ATX was used to align Oracle processes with their data in memory and evenly spread them across NUMA nodes on a 16-processor HPE Superdome Flex configuration.

The testing featured in this Reference Architecture highlights capabilities, best practices, and optimal settings for Oracle transaction processing and analytics workloads running in a Red Hat® Enterprise Linux® environment on the HPE Superdome Flex with HPE 3PAR StoreServ storage.

Target audience: This Reference Architecture (RA) is designed for IT professionals who use, program, manage, or administer large databases that require high performance. Specifically, this information is intended for those who evaluate, recommend, or design new and existing IT high performance architectures. Additionally, CIOs may be interested in this document as an aid to guide their organizations in determining when to implement an Oracle OLTP environment alongside an Oracle In-Memory solution for their Oracle online analytical processing (OLAP) environments and the performance characteristics associated with those implementations.

Document purpose: The purpose of this document is to describe a Reference Architecture demonstrating the benefits of running Oracle Database 18c on an HPE Superdome Flex platform and HPE 3PAR StoreServ 9450 All Flash array.

This Reference Architecture describes solution testing completed in January 2019.

Introduction

In today's fully connected world, the exponential increase in data collection and management has never been higher. In order to keep up with this demand, businesses must constantly increase their database computing resources. Formerly, analytics was often limited to business data that provided only a historical snapshot of the business. Now tremendous growth comes from new real-time data sources such as IoT devices, social media, video, etc. Analyzing data sources with business intelligence tools to create integrated intelligence is now needed to keep pace with change, or to create a competitive advantage. In order to support these business intelligence capabilities in a near real-time fashion, a high-capacity OLTP system can help to feed these analytics.

Running a mix of transactional and analytic workloads on the same Oracle database can be game changing. It can eliminate or reduce some of the issues inherent in maintaining separate systems for each workload, including:

- Latency associated with data availability.
- ETL (Extract, Transfer and Load) processes to extract data from the OLTP environment, transport it to the OLAP environment, and load it into a data warehouse. In addition to the afore-mentioned latency, there can be challenges with transforming the data from the OLTP representation to the OLAP format.
- Using multiple systems for OLTP and OLAP may increase the number of Oracle licenses required.
- Each system must be configured to meet peak processing needs.

In addition to running multiple workloads on a single server, the ability to scale up capacity as requirements increase offers the following advantages over adding more servers (scaling-out):

- **Oracle database consolidation** – Over time, as Oracle database systems are deployed for departmental applications, and projects, many companies find that they have an abundance of under-utilized systems that need to be maintained. With the ability to consolidate many systems into instances running on a scale-up platform, management and infrastructure costs can be reduced significantly.
- **Legacy applications** – Many older legacy applications simply do not have the out-of-the-box capability to run on a scale-out platform. Often the cornerstone of many IT infrastructures, these applications may require additional costly middleware applications or extensive-rewrites. Moving to a scale-up platform allows the application to scale without modifications.
- **Resource-demanding applications** – Applications such as OLTP require real-time processing ability. These capabilities in turn require large amounts of CPU, memory, and storage resources. Performing these operations on a single platform avoids the overhead of aggregating data across multiple systems/storage.

For mission-critical workloads the HPE Superdome Flex provides the ease of scale-up combined with the capacity required to run mixed workloads. Since it's easily scaled by simply adding more chassis, there is no migration to a new server, you simply add the new resources to the existing partition. Mission-critical resiliency is provided through end-to-end implementation of processor RAS features, redundancy of key system components, and advanced system software, to help ensure the server is up and running 24 x 7.[1] Additionally, in rare cases, where a problem happens that would typically bring down the entire server, the HPE Superdome Flex's modular design means that the problem , can be remedied by modifying the existing partition to exclude the problem chassis, and then the server can be brought up (with reduced resources), in order to continue operations.

Management complexities are reduced as up to all eight chassis (32 processors) can be managed as a single entity. Unlike scale-out clusters, the HPE Superdome Flex provides great performance with minimal tuning.

In order to scale up a system such as the HPE Superdome Flex, a storage system with similar, scale-up capability is required. The HPE 3PAR portfolio can scale to four controller nodes for the midrange products, and eight nodes for the high-end products. In addition, capacity can be scaled from a few terabytes to over 80 PB in a four-system federation with a common OS, feature set, and management.

With HPE Infosight predictive analysis technology, and the ability to group arrays together for management and aggregation, these arrays are perfectly suited for mission critical environments.

HPE Infosight not only monitors the environment for problems and potential hazards, it also proactively predicts problems before they occur, and in some cases can resolve these problems without intervention. Infosight can see across your entire infrastructure, giving you a view that you may have never had before, transforming your whole support model.

The testing highlighted in this Reference Architecture details the OLTP and OLAP capabilities of a single, 16-processor HPE Superdome Flex configuration. The hardware and software components of this solution are detailed below.

[1] For further details, see HPE Superdome Flex server architecture and RAS

Solution overview

This solution included the HPE Superdome Flex with HPE 3PAR StoreServ 9450 All Flash storage, running Oracle Database 18c. The HPE Application Tuner Express software was utilized to achieve maximum performance in a NUMA environment.

HPE Superdome Flex

The mission-critical HPE Superdome Flex platform is ideal for managing the exponentially growing data coming in and out of a business. The HPE Superdome Flex was designed with a modular approach to adapt to ever-increasing data management needs.

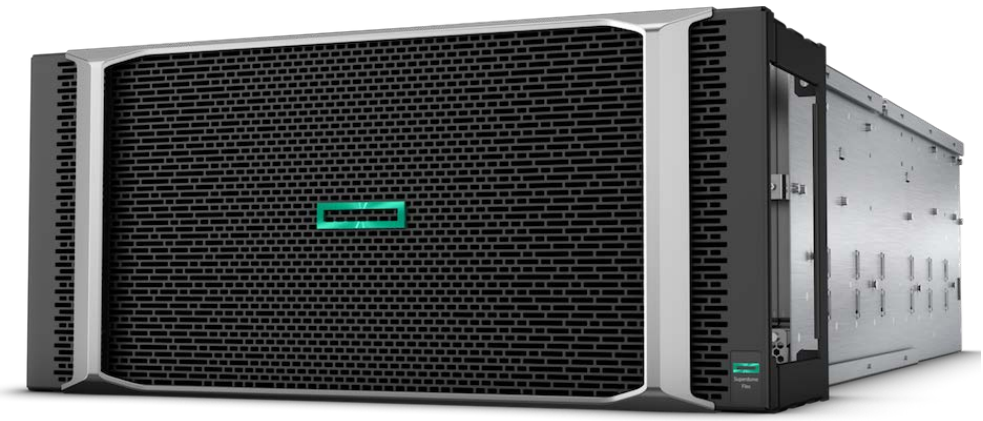

Figure 1. HPE Superdome Flex Chassis

With the HPE Superdome Flex, your Oracle database infrastructure can grow along with your data needs. Start out as small as a single 4-socket chassis and grow up to eight total chassis and 32 sockets.

- 48 DIMM slots per chassis
- Up to 16 I/O slots per chassis
- Support for 128GB memory DIMMs and Intel Xeon Scalable processors, including Platinum, Gold, and M versions.

The HPE Superdome Flex can be configured with an embedded management controller (for use with up to two chassis only) or an external rack management controller (RMC).

The RMC is the administrative node for the HPE Superdome Flex. The RMC includes one network port for administrative access to the system console. All administrative functions of the HPE Superdome Flex are performed through the management controller:

- Configuring system partitions
- Network configuration
- Booting, rebooting, and shutting down the system
- Viewing hardware resources, etc.

Computing resources on the HPE Superdome Flex are assigned using partitions. Partitions are currently aligned along the chassis boundaries within the HPE Superdome Flex environment. This greatly enhances the ease in which CPUs can be added and memory can be expanded to the installed operating system partition.

The solution highlighted for this Reference Architecture features a single four-chassis, 16-processor server. Additional chassis can be added at a later time by simply adding them to the existing server.

HPE 3PAR StoreServ 9450 Storage

HPE 3PAR StoreServ 9450 Storage is an enterprise-class, all-flash array that helps consolidate primary storage workloads without compromising performance, scalability, data services or resiliency. This 3PAR model is built for all-flash consolidation, delivering the performance, simplicity, and agility needed to support a hybrid IT environment. It can scale up to 6000 TiB of raw capacity, and is capable of over 2 million IOPS at sub-millisecond latency. These capabilities are complemented by enterprise-class, Tier-1 features and functionality. HPE 3PAR StoreServ All Flash is designed for 99.9999% availability with full hardware redundancy, supporting availability objectives for the most demanding environments. Enhanced storage capabilities provide continuous data access and the HPE 3PAR Priority Optimization software offers fine-grained QoS controls to ensure predictable service levels for all applications without physical partitioning of resources.

Figure 2. HPE 3PAR StoreServ 9450 storage

HPE Infosight for HPE 3PAR

In addition, HPE 3PAR customers can now benefit from HPE InfoSight for HPE 3PAR. HPE Infosight is an industry-leading predictive analytics platform that brings software-defined intelligence to the data center with the ability to predict and prevent infrastructure problems before they happen. The first release of HPE Infosight for HPE 3PAR provides the following capabilities:

- **Cross-stack analytics.** For HPE 3PAR customers running the latest release of the HPE 3PAR operating system,[2] HPE InfoSight offers the ability to resolve performance problems and pinpoint the root cause of issues between the storage and host virtual machines (VMs). It also provides visibility to locate "noisy neighbor" VMs.
- **Global visibility.** Through a new cloud portal that combines HPE Infosight and HPE StoreFront Remote, all current HPE 3PAR customers with systems that are remotely connected will see detailed performance trending, capacity predictions, health checks and best practice information across all of their HPE 3PAR arrays.
- **Foundation to enable predictive support.** Analytics and automation infrastructure are now in place that in the future will be used to detect anomalies, predict complex problems, and route cases directly to Level 3 support.

HPE 3PAR benefits for Oracle

Oracle customers can use HPE 3PAR StoreServ Storage to address their most significant challenges including:

- **Performance** – HPE 3PAR delivers high throughput and low latencies in multi-tenant, mixed-workload Oracle environments. With industry leading performance and sub-millisecond latencies, HPE 3PAR provides high transactions-per-second (TPS) and minimal wait times for Oracle OLTP workloads, using features such as Priority Optimization and Adaptive Flash Cache.[3]

[2] Requires HPE 3PAR OS version 3.3.1 GA or later and Service Processor version 5.0.3
[3] Adaptive Flash Cache utilizes SSDs as a cache for slower storage devices, and therefore is not needed in the all-flash storage arrays such as the HPE 3PAR StoreServ 9450.

- **Efficiency and Data Reduction** – In many Oracle environments, overprovisioning the primary database has become a matter of survival. As Oracle databases grow, every added core creates additional license and support fees. HPE 3PAR helps reduce Oracle sprawl and simplifies instance consolidation, driving higher TPS and providing a reduced storage footprint. What's more, HPE 3PAR Adaptive Data Reduction technologies, such as compression, can boost storage efficiency while helping enterprises bypass costly Oracle compression license fees.
- **High Availability and Data Protection** – Many Oracle environments face challenges with a growing primary database with an increasing number of applications writing to that database. As databases get larger, backup windows, recovery point objectives (RPO), and recovery time objectives (RTO) become harder to meet. HPE 3PAR offers a broad set of solutions that drive high availability for Oracle, including snapshots and Remote Copy. For direct snapshot copies to HPE StoreOnce Systems as the target backup appliance, HPE 3PAR includes Recovery Manager Central for Oracle (HPE 3PAR RMC-O) software at no additional cost, delivering fast backup and recovery for Oracle data. The HPE StoreOnce Catalyst Plug-in for Oracle is also offered free of charge and is tightly integrated with Oracle Recovery Manager (RMAN), giving the Oracle database administrator complete visibility and control for backup and recovery tasks. For more information about this solution, see HPE Reference Architecture for Comprehensive Oracle Backup, Restore and Disaster Recovery using HPE RMC and HPE StoreOnce. In addition, HPE 3PAR Peer Persistence can be deployed with Oracle RAC to provide customers with a highly available stretched cluster. Peer Persistence can also be utilized in conjunction with HPE Serviceguard for Linux in a single instance Oracle environment to provide automated fail-over protection with no data loss (RPO of zero).

Taken together, these features help Oracle database and storage administrators manage even the most demanding Oracle environments, delivering the performance, data protection, efficiency, and high availability needed to keep critical applications and business processes up and running.

For more details about the HPE 3PAR features that benefit Oracle environments, see Best Practices for Oracle Database on HPE 3PAR StoreServ Storage.

HPE Application Tuner Express (HPE-ATX)

HPE Application Tuner Express is a utility for Linux® users to achieve maximum performance when running on multi-socket servers. Using this tool, you can align application execution with the data in memory resulting in increased performance. The tool does not require any changes to your applications, but runs alongside them. Because many x86 applications today were designed for older and smaller systems, HPE-ATX was designed to take advantage of the resources of newer servers to run workloads more efficiently. HPE-ATX offers the following launch policies to control the distribution of an application's processes and threads in a NUMA environment:

- Round Robin: Each time a process (or thread) is created, it will be launched on the next NUMA node in the list of available nodes. This ensures even distribution across all of the nodes.
- Fill First: Each time a process (or thread) is created, it will be launched on the same NUMA node until the number of processes (or threads) matches the number of CPUs in that node. Once that node is filled, future processes will be launched on the next NUMA node.
- Pack: All processes (or threads) will be launched on the same NUMA node.
- None: No launch policy is defined. Any child process or sibling thread that is created will inherit any NUMA affinity constraints from its creator.

HPE-ATX is fully supported by Hewlett Packard Enterprise and can be downloaded from the HPE My License Portal.

Oracle Database 18c

Oracle Database 18c is the latest generation of the widely-used Oracle Database. It adds new functionality and enhancements to features previously introduced in Oracle Database 12c, including:[4]

- Multitenant Architecture
- In-Memory Column Store
- Native Database Sharing
- Additional capabilities for enhanced database performance, availability, security, analytics, and application development.

Of particular interest for this Reference Architecture is the In-Memory Option, which is available with Oracle Database Enterprise Edition. Oracle Database In-Memory is a suite of features that greatly improves performance for real-time analytics and mixed workloads. The In-Memory

[4] Introducing Oracle Database 18c, https://www.oracle.com/technetwork/database/oracledatabase18c-wp-4392576.pdf

Column Store is the key feature of this product. The In-Memory Column Store stores tables and partitions in memory using a columnar format optimized for rapid scans. Oracle Database manages data in columnar and row formats simultaneously. The In-Memory features can accelerate analytic queries by orders of magnitude without sacrificing OLTP performance or availability. The In-Memory column store resides in the In-Memory Area, which is an optional portion of the system global area (SGA). With Oracle Database 18c, the management of the In-Memory column store can be automated, so the database takes care of populating objects in memory and removing those that are no longer needed. Unlike other in-memory databases, Oracle Database In-Memory does not require the entire database or entire tables to fit in memory.

Solution components

This solution features the HPE Superdome Flex with HPE 3PAR StoreServ 9450 All Flash storage array. A diagram of the hardware components is shown in Figure 3.

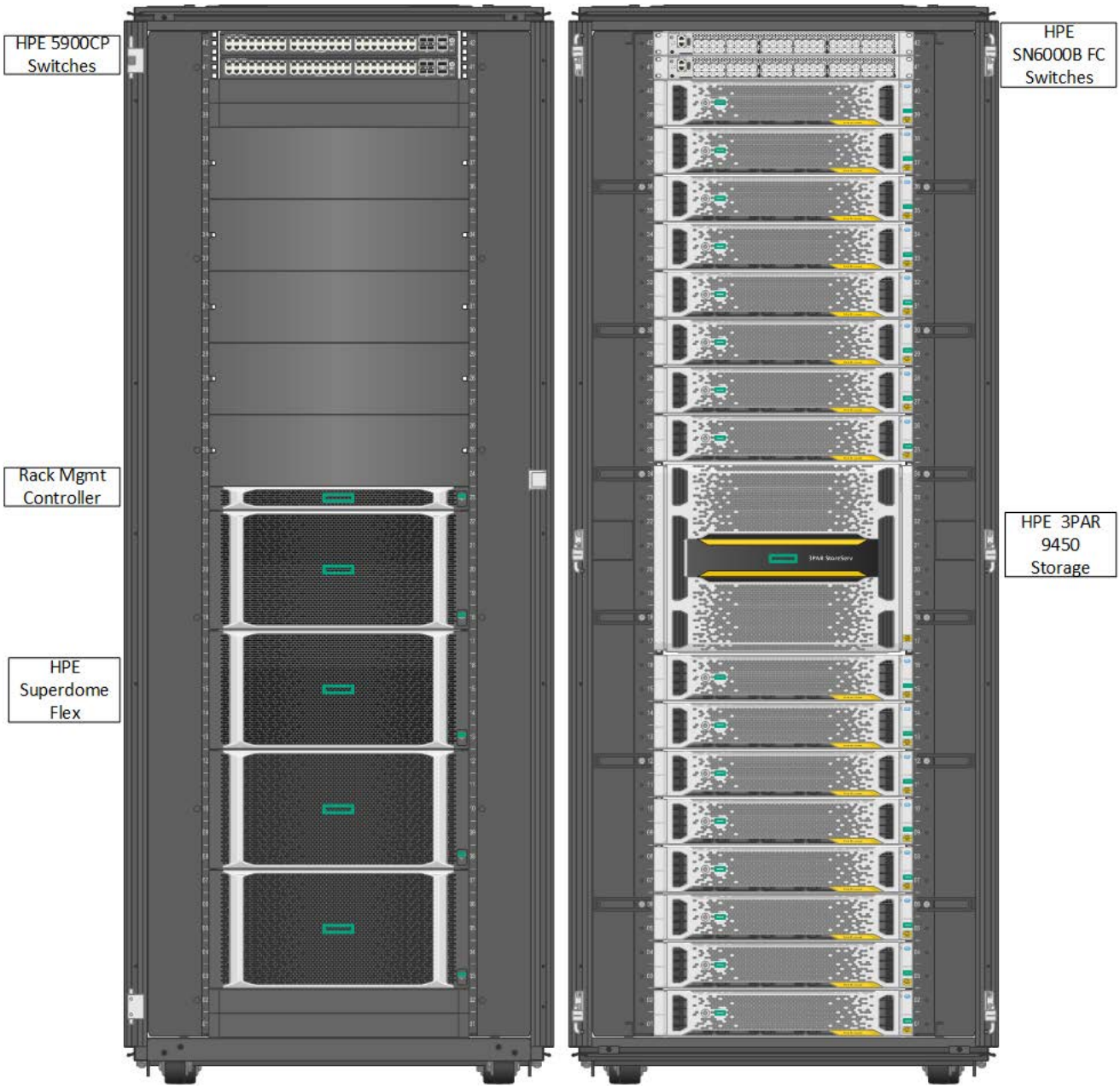

Figure 3. Solution hardware diagram

Hardware

A four-socket chassis is the core building block for the HPE Superdome Flex. Expanding beyond four sockets requires the use of expansion chassis. Each chassis supports four processors, 48 memory DIMM slots, up to 16 I/O slots, eight fans and four power supplies. Up to seven additional chassis can be added to a single HPE Superdome Flex. For this Reference Architecture, the server had one base chassis and three expansion chassis, for a total of 16 processors and 12TB of memory. Each chassis included the hardware listed in table 1.

Table 1. HPE Superdome Flex Configuration

HPE Superdome Flex (Chassis Configuration)	
Processor	Four 28-core Intel® Xeon® 8180 processors at 2.50GHz
Memory	3TB memory (48 x 64GB HPE DDR4 SmartMemory LRDIMMs)
FC HBAs	4 X 16Gb HPE SN1200E 16Gb DP FC HBA

The HPE 3PAR StoreServ 9450 all-flash storage array was configured as follows.

Table 2. HPE 3PAR StoreServ 9450 configuration

HPE 3PAR StoreServ 9450 all-flash storage array	
Nodes	4 nodes
Cache	896GiB
Drive enclosures	16
Drives	160 x 400GB SAS SFF SSDs
FC ports	8 x 4-port 16Gb Fibre Channel HBAs

Software

The following software was configured on the system:

- Red Hat Enterprise Linux version 7.5
- HPE Foundation Software 1.2.1-1
- HPE Application Tuner Express 1.0.1-103.15

Application software

Oracle Database 18c Enterprise Edition (18.3.0.0.0) was used for this Reference Architecture.

Best practices and configuration guidance for the solution

Install HPE Foundation Software

The HPE Foundation Software should be installed on the server after installing the OS. This software consists of packages designed to ensure the smooth operation of the server. It includes Data Collection Daemon (DCD) for Linux, an agentless service that proactively monitors the health of hardware components in the server. For instructions on installing this bundle, see the HPE Superdome Flex Server OS Installation Guide.

Configure kernel boot options

The HPE Foundation Software sets some kernel boot options that are key to optimal performance. Some additional parameters were also set manually. The boot options listed in table 3 were used for optimal performance for this reference architecture. Note that numa_balancing=0 and transparent_hugepage=never are both recommended by Oracle. When NUMA balancing is enabled, the kernel migrates a task's pages to the same NUMA node where the task is running. Due to the size of the Oracle SGA and frequency of task migrations, the migration of pages can be expensive, and optimal performance can be achieved by disabling NUMA balancing. Transparent huge pages are enabled by default in RHEL,

and Oracle recommends disabling this setting to avoid memory allocation delays at runtime. Note that while dynamically-allocated transparent huge pages were disabled, statically-allocated huge pages were configured via the vm.nr_hugepages kernel parameter as described in the RHEL OS settings section below. The option intel_idle.max_cstate=1 allows Cstate 1 transitions and encourages TurboBoost functionality. The multi-queue block I/O queuing options were set because high block soft IRQ times were reported by the linuxki tool. Note that multiqueue support must first be added in the Emulex lpfc driver by including lpfc_use_blk_mq=1 in the /etc/modprobe.d/lpfc.conf file. See Appendix G for further details.

Table 3. Kernel boot options for optimal performance

Kernel boot option	Description	How set
numa_balancing=0	Disable automatic numa balancing	HPE Foundation Software
intel_idle.max_cstate=1	Allow only C-state 1 transitions to encourage Turboboost clock speed increases	HPE Foundation Software
transparent_hugepage=never	Disable transparent huge pages	HPE Foundation Software
cgroup_disable=cpu	Disable cgroups for the CPU	Manually added to /etc/default/grub
scsi_mod.use_blk_mq=y	Enable multi-queue block I/O queuing	Manually added to /etc/default/grub
dm_mod.use_blk_mq=y	Enable device mapper to use blk-mq	Manually added to /etc/default/grub

RHEL OS settings

A complete list of the RHEL tuning parameters is shown in Appendix B. The shmmax parameter and the number of huge pages were set large enough to contain the Oracle SGA (buffer cache), which included the in-memory tables for the OLAP workload.

HPE 3PAR 9450 StoreServ All Flash array volumes

The HPE 3PAR StoreServ 9450 All Flash array was configured with the volumes listed in table 4. Sixteen volumes were configured in an Oracle ASM disk group named DATA, which contained the Oracle tablespaces, indexes, undo tablespace, and temp tablespace. Eight volumes were used for the ASM disk group REDO_A, and eight more were consumed for REDO_B, which contained the Oracle redo logs.

For the OLAP 300GB and 3TB tests, there was sufficient space on the DATA disk group to hold the schemas. For the OLAP 10TB test, a separate ASM disk group, utilizing another 16 volumes, was created to hold the schema.

Table 4. HPE 3PAR volumes

Quantity and size	RAID level, type	Description
1 x 100GB	RAID10, thin	Oracle binaries
16 x 768GB	RAID10, thin	Oracle ASM DATA disk group
16 x 128GB	RAID10, full	Oracle ASM REDO disk groups, 8 for REDO_A and 8 for REDO_B
16 x 896GB	RAID5, thin	Oracle ASM disk group for OLAP 10TB schema

Oracle configuration

A complete list of the Oracle parameters set for various tests is provided in Appendix D.

The Oracle SGA was set large enough to minimize physical reads. For the OLTP-only tests, the SGA was set to 929GB. For the various OLAP tests, it was increased sufficiently to allow putting all of the OLAP tables in memory, as shown in table 5. Note that the Oracle parameter inmemory_size sets the size of the In-Memory Column Store, and is allocated from the SGA.

Table 5. Oracle SGA settings

Test	sga_target	inmemory_size
OLTP + OLAP 300GB	1379GB	450GB
OLTP + OLAP 3TB	4TB	3TB
OLTP + OLAP 10TB	10TB	9TB

Due to compression, the inmemory_size parameter can be set to a smaller value than the size of the schema, which was the case for the 10TB schema. Table 6 shows the on-disk size of the tables for the 10TB schema as compared to their size in memory.

Table 6. Oracle SGA settings

Table	On-disk size (GB)	In-memory size (GB)	Compression ratio
LINEITEM	8165.70	4434.55	1.84
ORDERS	1792.02	1663.73	1.08
PARTSUPP	1302.72	1182.19	1.10
PART	296.46	140.1	2.12
CUSTOMER	240.42	257.78	.93
SUPPLIER	14.55	16.60	.88

Two redo log files of 700GB each were configured to minimize log file switching during the performance tests. Customer implementations should determine the log file size required to meet their business needs.

An undo tablespace of 800GB was created to minimize overhead due to filling up the tablespace during a benchmark run.

A temp tablespace of 1TB was created. For the OLAP 10TB test, this was expanded to 6TB.

Capacity and sizing

Workload description

Oracle performance tests were conducted using HammerDB, an open-source tool. For this reference architecture, HammerDB 3.0 was used to implement an OLAP-type workload, as well as an OLTP workload.

For the online analytical processing (OLAP) workload, with the exception of sorting and the update function, the entire test is based on reading a large amount of data. This test is meant to emulate a Decision Support System (DSS), which represents a typical workload of business users inquiring about the performance of their business. This test is represented by a set of business-focused ad-hoc queries and the tests were measured by the amount of time taken to complete each discrete query as well as the amount of time to complete all of the queries. In all, 22 separate queries are part of this test scenario. The timed results were normalized and used to compare test configurations. Other metrics measured during the workload came from the operating system.

The tests were performed on schema sizes of 300GB, 3TB and 10TB. In all cases, the default compression type, “memcompress for query low”, was utilized, since this offers the best query performance.

For the online transaction processing (OLTP) workload, HammerDB provides a real-world type scenario that consumes both CPU for the application logic and I/O. The HammerDB tool implements an OLTP-type workload with small I/O sizes of a random nature. The transaction results were normalized and used to compare test configurations. Other metrics collected during the testing came from the operating system and/or standard Oracle Automatic Workload Repository (AWR) statistics reports.

The OLTP test, performed on a schema with 5,000 warehouses and 500GB in size, was both highly CPU and moderately I/O intensive. The environment was tuned for maximum user transactions. After the database was tuned, the transaction rates were recorded at various Oracle connection counts. Because customer workloads vary in characteristics, the measurement was made with a focus on maximum transactions.

Analysis and recommendations

To demonstrate the capacity of the HPE Superdome Flex and the performance advantages of HPE-ATX, the following tests were run:

- OLTP tests with and without HPE-ATX
- OLTP scale-up tests (with HPE-ATX)
- OLTP stand-alone, OLAP stand-alone, and OLTP plus OLAP, to compare performance of mixed workloads versus stand-alone (all OLTP tests with HPE-ATX).

HPE-ATX results

HPE Application Tuner Express was utilized to start the Oracle listener processes with the round robin flat policy. This ensured that the listener processes were evenly distributed across all of the nodes (sockets) of the server. The following HPE-ATX command was used:

```
hpe-atx -p rr_flat -l listener1.log lsnrctl start
```

Oracle OLTP tests with 400 Oracle connections were run with and without HPE-ATX, to demonstrate the performance benefits of this tool. The tests were run on 4, 8, 12, and 16 socket configurations. Figure 4 shows the relative increase in Oracle throughput (transactions per minute) when using HPE-ATX as compared to the same test without ATX for each configuration. As more nodes were added, the benefits of HPE-ATX increased due to the increased NUMA latencies with additional nodes. For the base HPE Superdome Flex configuration of one chassis (four processors), there was a 30% improvement in throughput, and that increased to a 67% performance improvement for the four chassis (16 processors) configuration. HPE-ATX was used for all subsequent tests.

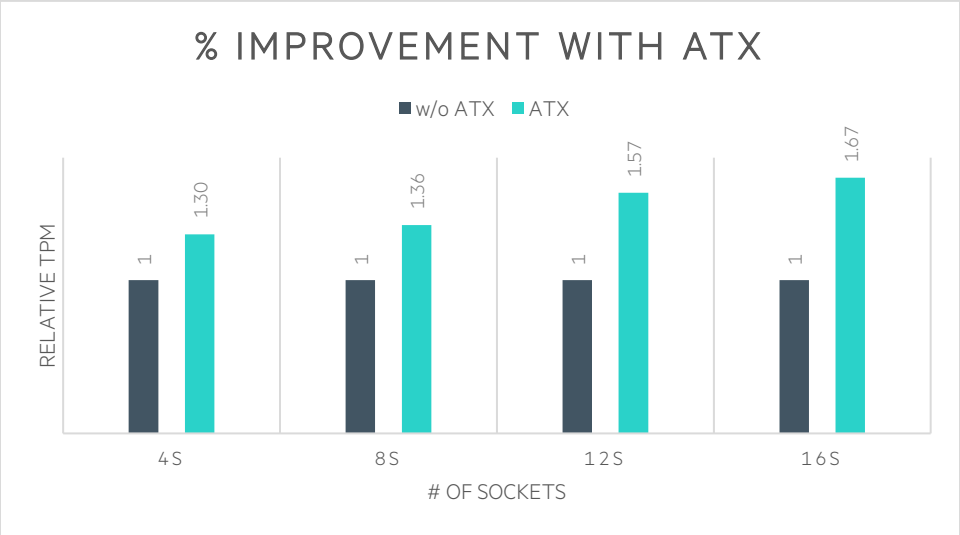

Figure 4. Performance improvement with HPE-ATX

OLTP scale-up results

Oracle OLTP throughput was compared for a single chassis, four-processor server and a two chassis, eight-processor server. For the four-processor configuration, peak throughput was achieved with 400 Oracle connections. For the eight-processor configuration, two Oracle instances were used to avoid Oracle lock contention. Peak throughput was obtained with 500 Oracle connections for each instance. The eight-processor configuration achieved 60% more TPM than the four-processor server, as shown in figure 5.

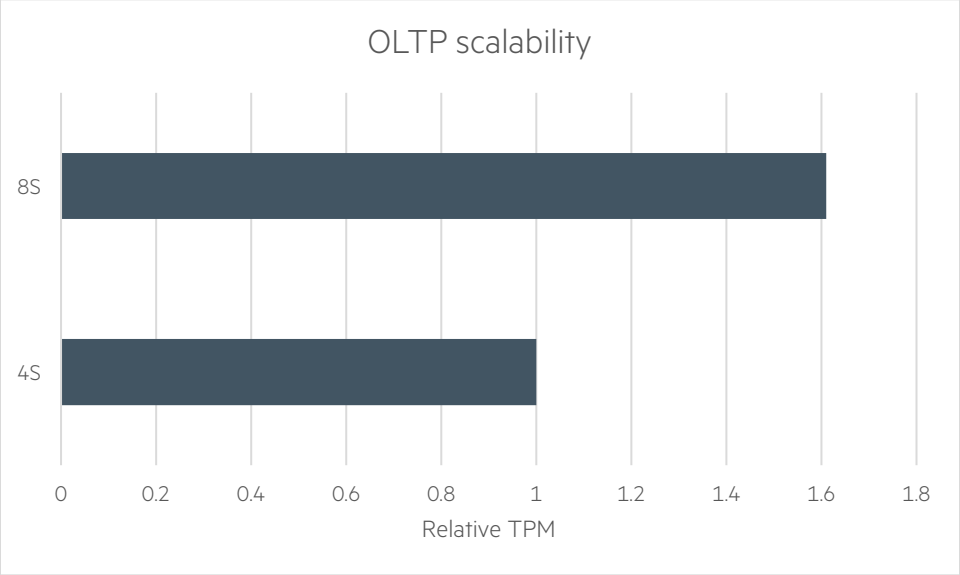

Figure 5. Oracle OLTP scalability for four versus eight-processor configurations

OLTP plus OLAP mixed-workload results

One of the key benefits of the large capacity of HPE Superdome Flex servers is the ability to run multiple types of workloads simultaneously. Tests were run with both OLTP and OLAP workloads to determine the impact of the simultaneous workloads as compared to running each workload on its own. For the OLAP workload, three different schema sizes were tested (300GB, 3TB and 10TB) to demonstrate the capacity of the Superdome Flex and the impact on an OLTP workload as the size of the OLAP workload increased. The OLAP workload utilized Oracle's in-memory feature, taking advantage of the large memory capacity available for the Superdome Flex.

Figure 6 shows the relative throughput achieved for the OLTP workload with varying Oracle connection counts when running stand-alone as compared to running the test simultaneously with a 300GB OLAP workload. For the mixed workload test, the OLTP transaction rate was within 89 to 93% of the rate for the stand-alone test. This demonstrates the ability to add an OLAP workload with little impact on the transaction processing throughput.

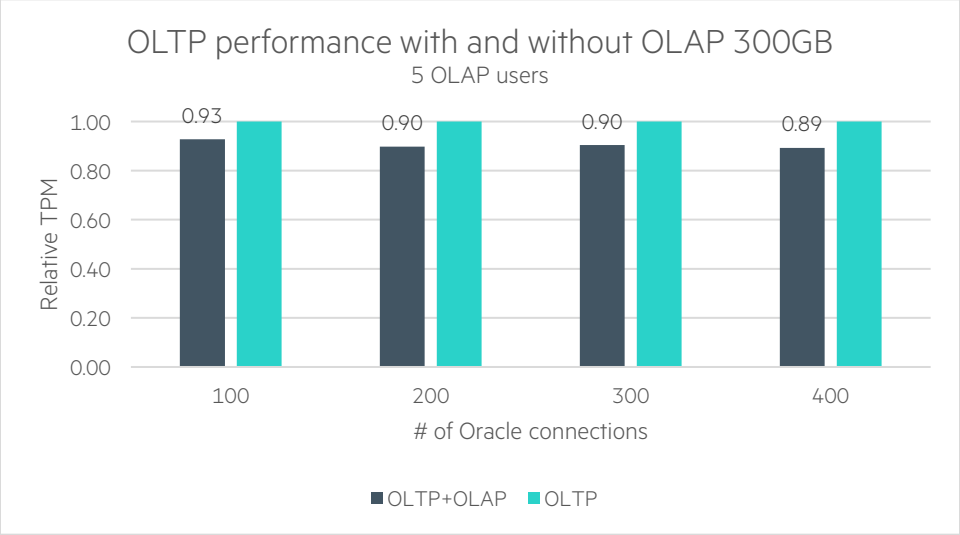

Figure 6. Oracle OLTP throughput with and without OLAP 300GB workload

The same tests were run with an OLAP workload with a 3TB schema. In this case, with the mixed workload, the OLTP transaction rate was 87 to 91% of the rate when running the stand-alone OLTP tests, as shown in figure 7. Even with this larger OLAP schema, OLTP performance remained fairly close to stand-alone transaction rates.

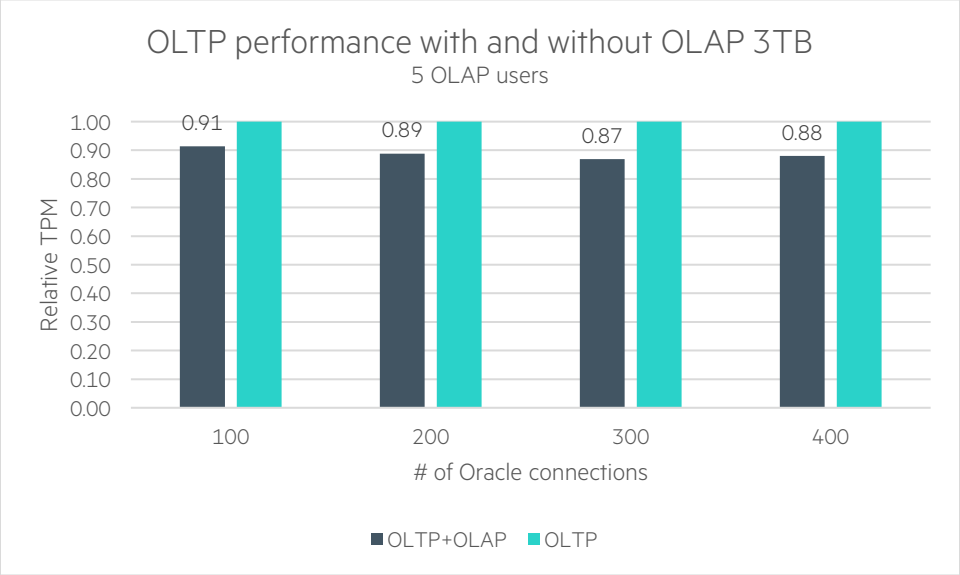

Figure 7. Oracle OLTP throughput with and without OLAP 3TB workload

Figure 8 shows the results when running an OLTP workload in conjunction with an OLAP workload with a 10TB schema. In this case, with the mixed workload, the OLTP transaction rate was 84 to 87% of the rate for the stand-alone OLTP tests when using 100 to 300 Oracle connections. At the 400 connection load, there was a bigger impact on the OLTP workload, which achieved 74% of the throughput as the stand-alone OLTP test.

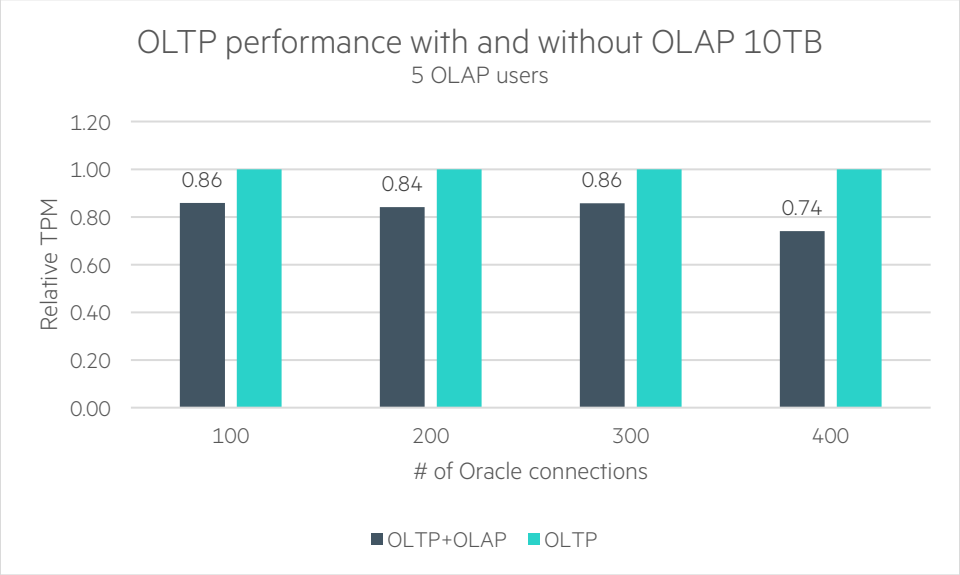

Figure 8. Oracle OLTP throughput with and without OLAP 10TB workload

The next set of graphs show the impact on OLAP performance when simultaneously running an OLTP workload. Figure 9 shows the relative time to complete the OLAP test for a 300GB schema for the mixed workload case as compared to running stand-alone. As the number of Oracle connection counts increased for the OLTP test, the time to complete the OLAP test increased from 6% up to 33%, as compared to the stand-alone OLAP test.

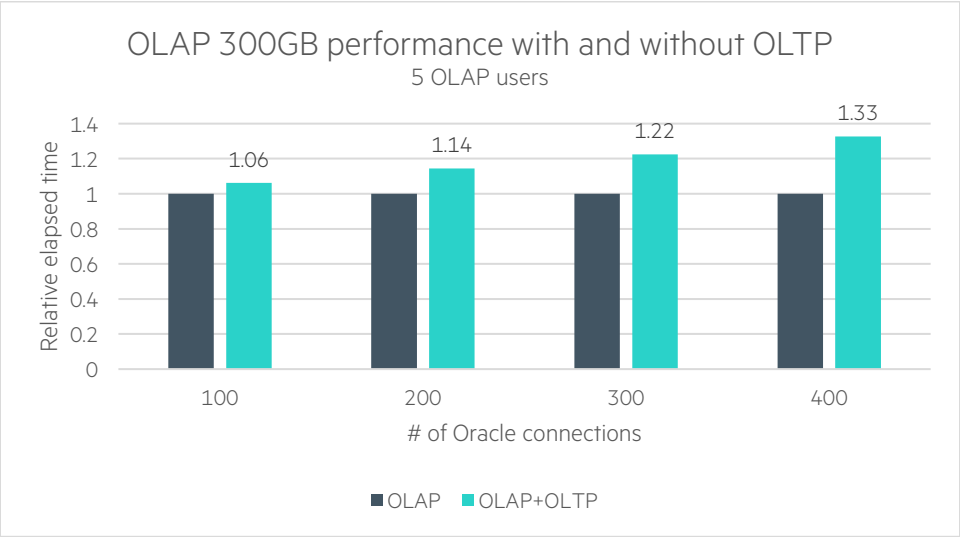

Figure 9. Oracle OLAP 300GB performance with and without OLTP workload

When the OLAP 3TB schema is used, the impact on the mixed workload was similar to the 300GB schema up through 300 Oracle connections for the OLTP workload, as shown in figure 10. There was a heavier impact on the OLAP workload when 400 connections were used for the OLTP workload.

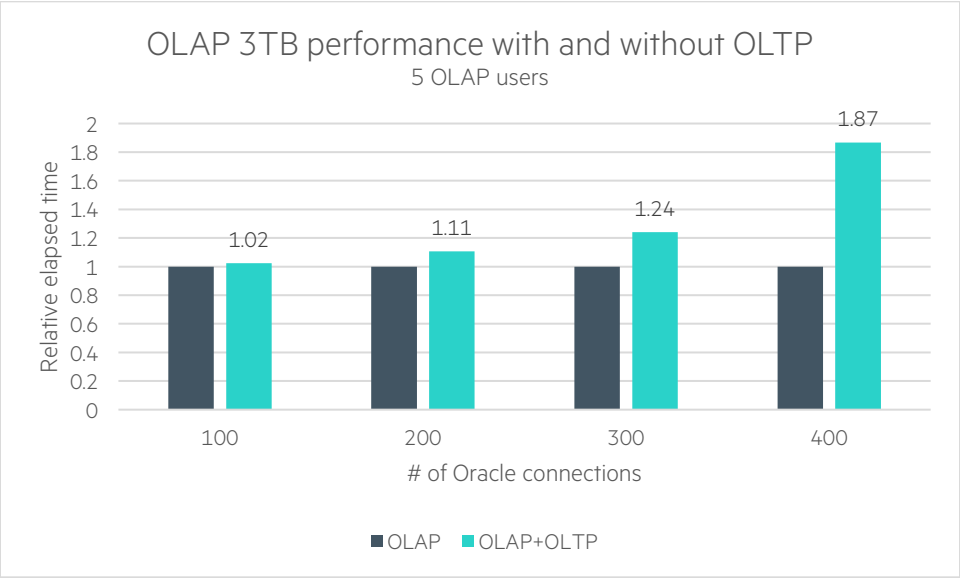

Figure 10. Oracle OLAP 3TB performance with and without OLTP workload

With a 10TB schema, the OLAP test consumed too much time to complete multiple runs with 5 OLAP users, so the tests were run with a single OLAP user (versus 5 OLAP users for the 300GB and 3TB OLAP tests). With just a single OLAP user, the impact of OLTP users on the OLAP test was smaller. The time to complete the OLAP test was only 6 to 14% longer when simultaneously running the OLTP workload than when running the standalone OLAP test, as shown in figure 11.

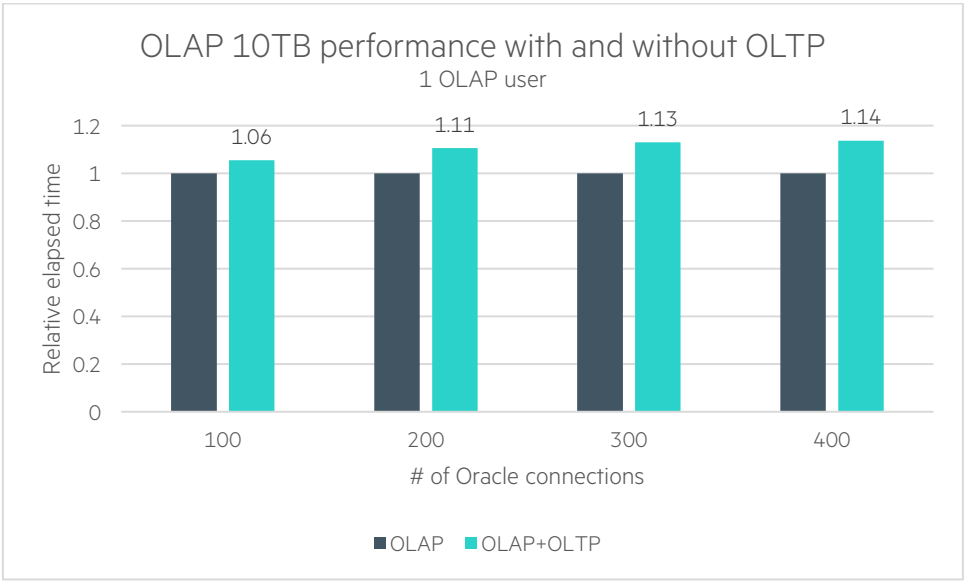

Figure 11. Oracle OLAP 10TB performance with and without OLTP workload

Summary

Combining online transaction processing (OLTP) with online analytical processing (OLAP) allows analytics to be run in real time on in-flight transactional data. The combination of HPE Superdome Flex with HPE 3PAR Storage and Oracle In-Memory allows customers to run both OLTP and OLAP workloads simultaneously without slowing transactions, delivering analytics on demand, and providing competitive advantage. In addition, HPE Application Tuner Express allows customers to obtain optimal performance as workloads scale up, taking full advantage of HPE Superdome Flex processing and memory capacities.

Implementing a proof-of-concept

As a matter of best practice for all deployments, Hewlett Packard Enterprise recommends implementing a proof-of-concept using a test environment that matches as closely as possible the planned production environment. In this way, appropriate performance and scalability characterizations can be obtained. For help with a proof-of-concept, contact an HPE Services representative (hpe.com/us/en/services/consulting.html) or your HPE partner.

Appendix A: Bill of materials

Note

Part numbers are at time of publication/testing and subject to change. The bill of materials does not include complete support options or other rack and power requirements. If you have questions regarding ordering, please consult with your HPE Reseller or HPE Sales Representative for more details, hpe.com/us/en/services/consulting.html.

Table 7. Bill of materials

Qty	Part number	Description
		Rack Infrastructure
1	P9K16A	HPE 42U 800x1200mm Advanced Shock Rack
1	P9K16A 001	HPE Factory Express Base Racking Service
2	JG838A	HPE FlexFabric 5900CP 48XG 4QSFP+ Switch
		HPE Superdome Flex
1	Q2N05A	HPE Superdome Flex Base Chassis
2	Q2N43A[5]	HPE Superdome Flex 480GB SATA SSD
1	Q2N42A	HPE Superdome Flex DVD-R Drive
3	Q2N06A	HPE Superdome Flex 4s Expansion Chassis
16	Q6L90A	HPE Superdome Flex Intel Xeon-Platinum 8180 (2.5GHz/28-core/205W) Processor Kit
48	Q2N39A	HPE Superdome Flex DDR4 256GB (4x64GB) Mem Kit
4	Q2N08A	HPE Superdome Flex PCIe FH 12-slot 3 Riser Kit
16	Q0L14A	HPE SN1200E 16Gb 2p FC HBA
1	Q2N16A	HPE Superdome Flex 16-socket Interconnect and Scale Activation Kit
1	Q2N07A	HPE Superdome Flex Mgmt Controller
4	P9Q61A	HPE G2 Basic 3Ph 17.3kVA/C13 NA/JP PDU

[5] This disk drive is superceded by part number R2A72A in February, 2019.

Qty	Part number	Description
		HPE 3PAR 9450 Storage
1	P9K10A	HPE 42U 600x1200mm Adv G2 Kit Shock Rack
1	Q0E92A	HPE 3PAR 9450 2N+SW Storage Base
2	Q7F41A	HPE 3PAR 9450+SW Storage Node
4	Q0E96A	HPE 3PAR 9000 4pt 12Gb SAS HBA
8	Q0E97A	HPE 3PAR 9000 4pt 16Gb FC HBA
1	P9M30A	HPE 3PAR Direct Connect Cabling Option
16	Q0E95A	HPE 3PAR 9000 2U SFF Drive Enclosure
160	Q0F40A	HPE 3PAR 9000 400GB+SW SFF SSD
32	716197-B21	HPE Ext 2.0m MiniSAS HD to MiniSAS HD Cbl
1	Q0F86A	HPE 3PAR StoreServ RPS Service Processor
1	Q1H95A	HPE 3PAR 1U Rack Accessories Kit
4	P9Q41A	HPE G2 Basic 4.9kVA/(20) C13 NA/JP PDU
1	BW932A	HPE 600mm Rack Stabilizer Kit
1	L7F19A	HPE 3PAR All-in Sngl-sys SW Latest Media
2	QK753C	HPE SN6600B 16Gb 48/24 FC Switch

Appendix B: RHEL kernel settings

The following RHEL kernel parameters were set in the /etc/sysctl.conf file. The shmmax parameter and nr_hugepages were set large enough to accommodate the memory requirements of the 10TB OLAP plus the OLTP workloads. The aio-max-nr had to be increased from its typical setting of 1048576 to 3145728 for the 3TB OLAP tests.

```
kernel.sem = 250 32000 100 128
kernel.shmall = 4294967295
# set large enough for 10TB OLAP plus OLTP workloads
kernel.shmmax = 12094627905536
fs.file-max = 6815744
kernel.shmmni = 16384
# increase for 3TB OLAP test
#fs.aio-max-nr = 1048576
fs.aio-max-nr = 3145728
net.ipv4.ip_local_port_range = 9000 65500
net.core.rmem_default = 1048576
net.core.wmem_default = 1048576
net.core.rmem_max=26214400
net.core.wmem_max=26214400
net.ipv4.tcp_rmem = 1048576 1048576 4194304
net.ipv4.tcp_wmem = 1048576 1048576 1048576
#configure huge pages for 10TB OLAP plus OLTP workloads
vm.nr_hugepages = 5505020
vm.hugetlb_shm_group = 507
kernel.numa_balancing = 0
```

Appendix C: Oracle user account limits

The following settings were included in the file /etc/security/limits.d/oracle-limits.conf:

```
oracle   soft   nofile    1024
oracle   hard   nofile    65536
oracle   soft   nproc    16384
oracle   hard   nproc    16384
oracle   soft   stack    10240
oracle   hard   stack    32768
oracle   hard   memlock 12240656794
oracle   soft   memlock 12240656794
```

Appendix D: Oracle initialization parameters

The following Oracle parameters were set in the init.ora initialization file. The sga_target and inmemory_size settings were for the 10TB OLAP tests. See Table 5 for the settings for the other tests.

```
*.audit_file_dest='/u01/app/oracle/admin/oraflex/adump'
*.audit_trail='db'
*.compatible='18.0.0'
*.control_files='+DATA/ORAFLEX/CONTROLFILE/current.261.994612837'
*.db_block_size=8192
*.db_create_file_dest='+DATA'
*.db_name='oraflex'
*.diagnostic_dest='/u01/app/oracle'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=oraflexXDB)'
*.local_listener='LISTENER_ORAFLEX'
*.nls_language='AMERICAN'
*.nls_territory='AMERICA'
*.open_cursors=3000
*.pga_aggregate_target=316760M
*.processes=3000
*.remote_login_passwordfile='EXCLUSIVE'
*.sga_target=10T
*.undo_tablespace='UNDOTBS1'
_high_priority_processes='VKTM*|LG*'
lock_sga=TRUE
use_large_pages='ONLY'
_fast_cursor_reexecute=true
_enable_NUMA_support=true
result_cache_max_size=0
shared_pool_size=236223201280
commit_logging=BATCH
commit_wait=NOWAIT
_undo_autotune=false
_in_memory_undo=TRUE
_trace_pool_size=0
audit_trail=NONE
pre_page_sga=FALSE
trace_enabled=FALSE
#for OLAP tests only
inmemory_size=9T
optimizer_dynamic_sampling=4
```

Appendix E: multipath.conf

The following entries were included in the /etc/multipath.conf file. Note that the round-robin path selector was utilized, rather than the default service-time path selector, due to the large number of paths to the HPE 3PAR storage. In addition, aliases were created for each HPE 3PAR volume, for usage in the udev rules file, dm-permission.rules, described in Appendix F. Only one of the alias entries is shown here, due to the large number of volumes that were used for the testing.

```
defaults {
        user_friendly_names yes
        find_multipaths yes
}
devices {
        device {
                vendor "3PARdata"
                product "VV"
                path_grouping_policy "group_by_prio"
                path_selector "round-robin 0"
                path_checker "tur"
                features "0"
                hardware_handler "1 alua"
                prio "alua"
                failback immediate
                rr_weight "uniform"
                no_path_retry 18
                fast_io_fail_tmo 10
                dev_loss_tmo "infinity"
        }
}
        multipath {
                wwid 360002ac0000000000000009100021014
                alias data01
        }
```

Appendix F: udev rules

One udev rules file was created to set parameters for the HPE 3PAR SSDs, and a second file was created to set ownership and permissions for the Oracle volumes. The file /etc/udev/rules.d/10-3par.rules set the rotational latency, I/O scheduler, rq_affinity and nomerges parameters. For SSDs, the rotational latency is set to zero. The I/O scheduler was set to noop. Setting rq_affinity to 2 forces block I/O completion requests to complete on the requesting CPU. The file contained the following settings:

```
ACTION=="add|change", KERNEL=="dm-*", PROGRAM="/bin/bash -c 'cat
/sys/block/$name/slaves/*/device/vendor | grep 3PARdata'", ATTR{queue/rotational}="0",
ATTR{queue/scheduler}="noop", ATTR{queue/rq_affinity}="2", ATTR{queue/nomerges}="1",
ATTR{queue/nr_requests}="128"
```

The /etc/udev/rules.d/dm-permission.rules file included the following settings for the HPE 3PAR volumes used for the Oracle ASM disk groups:

```
ENV{DM_NAME}=="data01", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data02", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data03", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data04", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data05", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data06", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data07", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data08", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data09", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data10", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data11", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data12", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data13", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data14", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data15", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="data16", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo01", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo02", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo03", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo04", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo05", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo06", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo07", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo08", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo09", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo10", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo11", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo12", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo13", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo14", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo15", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="redo16", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch01", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch02", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch03", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch04", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch05", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch06", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch07", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch08", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch09", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
```

```
ENV{DM_NAME}=="tpch10", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch11", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch12", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch13", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch14", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch15", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
ENV{DM_NAME}=="tpch16", OWNER:="oracle", GROUP:="oinstall", MODE:="660"
```

Appendix G: Emulex fibre channel adapter settings

The following settings were included in the file /etc/modprobe.d/lpfc.conf for the Emulex fibre channel adapter. The lpfc_use_blk=1 option was required to enable multiqueue block devices. The lpfc_devloss_tmo and lpfc_lun_queue_depth settings are recommended in the HPE 3PAR Red Hat Enterprise Linux Implementation Guide. Note that initramfs must be rebuilt and the server rebooted for these settings to take effect. The following command was used to rebuild initramfs:

```
/sbin/dracut -v --force --add multipath --include /etc/multipath.conf /etc/multipath.conf
```

Here are the contents of /etc/modprobe.d/lpfc.conf:

```
options lpfc lpfc_use_blk_mq=1
options lpfc lpfc_devloss_tmo=14
options lpfc lpfc_lun_queue_depth=16
```

Resources and additional links

HPE Mission Critical Servers Reference Architectures, hpe.com/info/missioncritical-ra

HPE Reference Architectures, hpe.com/info/ra

Oracle Solutions, hpe.com/info/oracle

HPE Superdome High-End Servers, hpe.com/superdome

HPE Superdome Flex server architecture and RAS, https://h20195.www2.hpe.com/v2/Getdocument.aspx?docname=a00036491enw

HPE Superdome Flex Server OS Installation Guide, https://support.hpe.com/hpsc/doc/public/display?docId=a00038168en_us

Running Linux on HPE Superdome Flex Server, https://support.hpe.com/hpsc/doc/public/display?docId=a00058577en_us

HPE Application Tuner Express, https://myenterpriselicense.hpe.com/cwp-ui/evaluation/HPE-ATX

HPE 3PAR StoreServ Storage, hpe.com/3par

HPE 3PAR Red Hat Enterprise Linux Implementation Guide, https://support.hpe.com/hpsc/doc/public/display?docId=c04448818

Best Practices for Oracle Database on HPE 3PAR StoreServ Storage, https://support.hpe.com/hpsc/doc/public/display?docId=emr_na-a00038978en_us&docLocale=en_US

Oracle Database In-Memory Guide, https://docs.oracle.com/en/database/oracle/oracle-database/18/inmem/database-memory-guide.pdf

To help us improve our documents, please provide feedback at hpe.com/contact/feedback.

Make the right purchase decision. Click here to chat with our presales specialists.

Sign up for updates

© Copyright 2019 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Red Hat and Red Hat Enterprise Linux are trademarks of Red Hat, Inc. in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries. Oracle is a registered trademark of Oracle and/or its affiliates. Intel and Xeon are trademarks of Intel Corporation in the U.S. and other countries.

a00065205enw, February 2019