Author: Marc Staimer, President Dragon Slayer Consulting

W H I T E P A P E R

Why Oracle Exadata Blows Away POWER9 for Oracle Databases

Separating Fact from Fiction

Why Oracle Exadata Blows Away POWER9 for Oracle Databases

Executive Summary

It's common for a technology vendor to present their products in the best light possible. IBM is no different in that respect. They proclaim their latest POWER9-based Power series system running Oracle Database delivers superior performance at a lower cost than Exadata. That's an audacious claim considering Exadata is purpose-built specifically for Oracle Database. What makes it so bold is how IBM backs up its claims. It is a real head scratcher. IBM focuses on the Power Systems architecture, POWER9 chip architecture, Flash System 9xxx Storage architecture, and ambiguous total cost of ownership comparisons.

Each of these arguments are meant to imply superiority without evidence or substantiation. They are suggestive at best with no correlation to actual Oracle DB performance. The Power Systems and POWER9 architectures may or may not provide good general database performance. IBM argues that the P9 chip is optimized for AI, in-memory OLTP, and data warehousing or analytics, making it a better Oracle DB performer. IBM will compare their POWER9 CPU to the latest Intel x86 CPUs utilized in Exadata, pointing out every perceived technical advantage regardless of the impact on Oracle Database workloads. They point out that by delivering more cores per socket, more threads per core, intelligent threading (workloads can be executed utilizing 1-8 threads per core automatically or determined by the sysadmin), and intelligent caching (dynamically varying cache utilization to minimize overall cache latency), will automatically deliver better Oracle Database performance and especially OLTP performance.

But does it? IBM provides no actual benchmark results to back that up. They only have middle of the road benchmark results from past generations (P8 and P7) Power series for the Oracle e-Business Suite and Oracle Siebel CRM 8.1 Industry Applications and Scalability tests. More threads may sound impressive, but more threads are valuable only when there are a lot of idle processes, which is not the case with Oracle DB. Intelligent memory caching is required to mask compute to storage latency issues. And there is always never enough cache because DRAM and especially non-volatile DRAM (NVDIMM) is expensive[1]. Not enough cache means more frequent cache misses and inconsistent Oracle Database performance.

The IBM POWER9 architecture is not a bad architecture. The fact of the matter is that it's optimized for high-performance NUMA (non-uniform memory access) supercomputers such as the DoD's "Summit" supercomputer, and AI like Watson. But it is definitely not optimized for Oracle DB despite IBM's declarations.

Comparing POWER9 for Oracle Database to Oracle Exadata demands far more than comparing component specifications. It requires end-to-end real-world results from a complete system. In addition to performance, any comparisons must examine: ease-of-use specifically with implementation, operations, upgrades, patches and management; availability, reliability, and serviceability more commonly referred to as RAS; and of course, total cost of ownership or TCO. In other words, from the user and DBA points of view, any comparison between IBM POWER9 running Oracle DB and Exadata must be based on which one brings the best **HEAT**:

- **H**igh performance
- **E**ase-of-use in implementation, operations, patches, upgrades, and management
- **A**rchitecture
- **T**CO

This document examines how the IBM POWER9 running Oracle Database and Oracle Exadata compare when bringing the **HEAT**.

[1] Although non-volatile persistent memory is on the horizon, it will likely not be a drop-in replacement for DRAM. Operating systems, file systems, and application software will have to be reworked to effectively utilize this new class of memory. It's dubious to expect this to happen in the near future.

Table of Contents

Comparing the **HEAT**

IBM believes its general purpose POWER9 System architecture is better suited to run Oracle Database than Oracle's purpose-built Exadata. This is despite the fact Exadata is specifically co-engineered with Oracle Database. The best way to separate the wheat from the chaff is to see how each compare when bringing the **HEAT**.

High-Performance

There are several key criteria when measuring Oracle Database high performance. These include:

- 8K random SQL read IOPS – critical for OLTP
- 8K random SQL write IOPS – critical for OLTP
- Throughput in Bps – crucial for OLAP and Data Warehouses (DW)
- Oracle Database IO end-to-end latency – critical for OLTP
- Oracle Database Log File Sync and Log File Parallel Write performance consistency – affects both OLTP, OLAP, and DW
- Oracle RAC node failure detection and reconfiguration performance – application continuity

Note that CPU threads, cores, and sockets are not a DBA or user high-performance criterion. They are a means to achieving an end, not the end in itself. That part will be covered in the architecture aspect of **HEAT**. Oracle Database performance is a key criterion. IBM unfortunately offers little to no published Oracle Database performance numbers at this time. They assert they are faster without evidence. IBM can assert whatever it wants, but it requires empirical and repeatable evidence or proof points to back that up. Oracle Database is optimized for 8K IOPS. This makes 8K random SQL read and write IOPS performance directly correlated to Oracle OLTP performance. Throughput is crucial for OLAP and DW performance.

Oracle publishes and keeps up-to-date Exadata performance metrics on its data sheets. Keep in mind that Exadata is purpose-built for Oracle Database performance. It's co-engineered with the Oracle Database to obtain paramount performance. The results are industry leading. The latest Exadata X7 produces up to 4.8 million random 8K SQL read IOPS per rack, up to 4.3 million random 8K SQL write IOPS per rack, as much as 350 GBps throughput per rack. These are_actual applicable Oracle DB SQL IOPS. They are not IOmeter based 4K IOPS that have no correlation to Oracle Database performance. And since Exadata X7 scales-out to 18 racks, that performance scales to 86.4 million random 8K SQL read IOPS, 77.4 million random 8K SQL write IOPS, and 6.3 TBps throughput. These performance numbers are currently unprecedented.

These metrics are measured from within the Oracle Database SQL statements being run. Most other storage and system vendors publish numbers measured at the component level, not from within the Oracle Database or any other database for that matter. SQL Flash IOPS means the IOPS are measured from within the SQL engine of the Oracle Database and exposed in AWR reports. Those other storage and system vendors publish IOPS numbers as measured from the storage array, or from a benchmark program running on the server. They do not correlate to Oracle Database SQL IOPS.

Latency is another underrated essential Oracle Database performance factor. Latency is cumulative and impacts the response time of every database transaction. It is less of a factor with OLAP and DW while still effecting the response time to first byte. Oracle engineered Exadata to provide a very low and consistent Oracle Database OLTP latency of 0.25 ms. Exadata reduces Oracle Database latencies via Smart Flash Logging; extensive internal use of RDMA (remote direct memory access) between compute nodes and between compute and storage cells; offload of SQL, XML, JSON, RMAN Backup, Data file vs. REDO IO segregation, encryption/decryption, and fast data file creation. Offloading moves the executable closer to the data. It takes far less time to move the executable and the processing results than it takes to move the input data. IBM does not publish Oracle Database latencies at all.

Log File Sync performance is where latency has an oversized impact on Oracle Database response time. More specifically the outliers. Log File Sync affects every connected session and user in the entire Database. Analogous to a traffic jam, latency outliers have a ripple effect lasting much longer than necessary, after they occur. Oracle Exadata Smart Flash Logging[2] eliminates Log File Sync and Log File Parallel Write latency outliers providing a consistent latency of no more than 1 ms.

The IBM Power Series based on the POWER9 architecture does not have anything equivalent to Oracle Exadata Smart Flash Logging or Smart OLTP Caching. Oracle Database users running on IBM Power Series anecdotally report waits of 64 ms for as high as 5% of the Log File Sync and Log File Parallel Write operations. Granted, 5% of operations means they are outliers. However, as previously stated, outliers typically have a major impact of the database's performance and user productivity.

Oracle added automatic Smart OLTP Caching to Exadata with Oracle Database 19c. Smart OLTP caching is unique to Exadata and it maintains OLTP service performance levels even when there is a storage failure. It eliminates cache misses on storage failures or repairs, thereby eliminating service interruptions. IBM POWER9 Systems have nothing equivalent.

In-Memory databases have become extremely popular because of the superior performance achieved through the use of DRAM. Oracle made Exadata the dominant in-Memory Oracle Database platform with Exadata exclusive functionality:

- In-Memory fault tolerance – duplicate all
- In-Memory on Active Data Guard
- Automatic in-Memory
- In-Memory External Tables
- In-Memory Row Store
- Faster Smart Scans using Column-level Checksum – new in 19c
- Faster ingest for Row Store – new in 19c
-

That unique functionality makes Exadata ideal for memory optimized workloads such as IoT.

Another key performance factor has to do with Oracle Real Application Clusters (RAC) nodal failures. The performance is the amount of time it takes for the Oracle Database RAC to detect a nodal failure and reconfigure affects the performance of each Oracle DB instance and the connected applications. Obviously faster is better. Slow RAC node reconfiguration performance can cause uncommitted transactions to be disconnected, Oracle Database instance sluggishness, and/or application timeouts forcing DBA intervention. Oracle Exadata detects and reconfigures in real-time, taking typically 3 seconds or less. It also transfers transactions in mid-flight from the failed node providing Application Continuity.

IBM Power Series POWER9 RAC implementations must wait for TCP/IP timeouts to detect a RAC node failure. This causes reconfigurations to take 2 minutes and frequently longer. Transferring transactions from the failed node in mid-flight is inconsistent. There have been very few IBM Power Series RAC implementations historically.

The performance numbers and consistency show that Exadata is engineered for mission-critical applications and their data. It delivers incomparable high performance and high volume OLTP, OLAP, DW, and extremely fast recovery from RAC node failures. The only conclusion that can be reached is that Exadata is the Oracle DB performance leader by a significant margin.

High Performance Advantage Oracle Exadata.

Ease-Of-Use In Implementation, Operations, Patches, Upgrades, & Management

Oracle Exadata is engineered specifically to allow DBAs to comfortably implement, operate, patch, and manage without requiring dedicated teams of server, network, or storage experts. In addition, no expertise is required for compute servers, storage servers, storage media, networking, interconnect, or any Exadata component. Exadata automates and minimizes tuning, partitioning, indexing, aggregating, and adjusting. It still empowers the DBA with server, storage, and network expertise to go beyond the built-in automation if they want to improve performance further.

[2] Spinning disk in Exadata is generally FASTER than Flash because it uses a DRAM-based controller cache. However, that controller cache sometimes gets filled up and the IO waits on disk. That's when Smart Flash Logging comes into play and knocks down the outliers. Those IO outliers on these critical IO operations in POWER9 systems generate system-wide slowdowns.

IBM POWER9 offer a general-purpose (GP) computing platform optimized for high performance computing (HPC) and artificial intelligence according to IBM public declarations and publications. It requires separate storage systems and networking. The GP Power System architecture demands expertise in the multiple disciplines of Power servers, AIX operating system (OS), storage networking, SAN storage, Oracle Database, and server clustering for Oracle RAC. In direct contrast, the Oracle Exadata only requires the database management administration (DBMA) team. This is more than a philosophical difference. It translates into reduced expertise necessity, training, admins, management, human errors, downtime, and budget. This becomes quickly evident when implementing Oracle Database on Exadata for the first time.

Oracle made Exadata intuitively simple and easy to implement, commonly completed in 1.5 days, including cleanup, with Oracle doing the install utilizing automated tools. But, the same cannot be said for Oracle Database on POWER9 or any IBM Power System. Implementations will typically take many days to weeks to complete. First the POWER9 compute nodes, network, and storage must be brought up. The Oracle Database must then be installed on each node and each Oracle Database instance configured. Then the tuning process begins and is ongoing for each Oracle Database instance, POWER9 compute nodes, networking, and storage. These are non-trivial processes with no automation, entailing extensive knowledge, skills, experience, manual effort, and time. The differences in implementation illustrate the Exadata usability advantage. But the difference in patch management demonstrates the ongoing Exadata operations ease-of-use advantage.

Patches have always been a headache for IT infrastructure admins. Patch processes grow continually more complicated and error prone with system complexity. This is an essential problem area that Oracle tackled and extensively simplified for Exadata. A single bundled patch is issued per quarter complete for ALL of the Exadata components. That patch includes all of the required patches for the entire Exadata ecosystem including the OS on both storage and database servers, InfiniBand switches, Oracle Grid infrastructure (ASM and Clusterware), Exadata storage software, Oracle Database, and everything else that is part of an Exadata system. One patch[3], one implementation, highly automated. Exadata has only a few configurations that need to be supported versus conventional system architectures with nearly infinite component permutations. And every permutation demands dozens of labor-intensive patches. Patches have to be found, downloaded, and implemented in the correct sequential order, not in parallel, on the right hardware, at the right time, several times a quarter or more frequently.

It is the latter that POWER9 Oracle Database DBAs must endure. Implementing these patches requires time, expertise, patience, and downtime. Patching has to be done frequently. Because patches tend to be disruptive they are typically scheduled. Scheduling generally takes place on weekends and holidays when they will (hopefully) be the least disruptive. Keep in mind that each of the system components are from different vendors. Patches are not necessarily tested in the POWER9 Oracle Database configuration. Individual patches are unaware of how they will affect the Oracle Database and the rest of the system in general. This in turn raises the probability of significant problems to arise forcing the patches to be backed out. It's more likely to be a major problem when multiple patches are implemented to multiple components concurrently. Determining the root cause patch to the problem in this situation is a frustrating exercise. Many admins will simply back out all the patches and reschedule for another time, which of course wastes yet more time and adds more cost.

But what are IBM's and Oracle's directions to simplify their Oracle Database architectures? Of the two only Oracle has made it clear that they are blazing new trails in database innovation and ease-of-use. The latest Exadata 19.1 raises ease-of-use yet again to new levels. In combination with the latest Oracle 19c Database Exadata now adds an autonomous type of optimizer. The autonomous optimizer automates performance monitoring of Oracle Database processes, memory, and network usage with online and real-time analytics. It additionally automates indexing with an expert machine learning engine based on what a skilled performance engineer would do. Reinforcement learning enables that expert machine learning engine to

[3] Exadata allows Oracle Database to run at a lower patch/version level than the Exadata infrastructure components. Databases can reside at a LOWER patch level when required. Oracle Databases don't automatically have to be upgraded at the same time on Exadata.

learn from its own actions as all candidate indexes are validated before being implemented completely automatically and transparently.

Automated security protects against runaway SQL statements, while the advanced intrusion detection environment (AIDE) automatically checks critical files every night via sha256, modified time, items added, items removed, and more.

Oracle has made multiple pronouncements making it clear their focus on putting more of the expertise into the system instead of the administrators. The goal is to make the Oracle Database much simpler to utilize. This has become evident with the Oracle Autonomous Database Cloud based on the Oracle 18c and now 19c. The Oracle Autonomous Database Cloud is essentially a 'self-driving' database with no human tuning, patching, repairing, or human intervention that eliminates manual labor. It is important to note that neither the Oracle Database 18c or 19c are by themselves are autonomous. The autonomy is additional capability Oracle engineered top of Oracle 18c and 19c with the Exadata platform. Oracle's Autonomous Database Cloud leverages artificial intelligence (AI) and machine learning algorithms to automate pedantic and time-consuming DBA functions. This frees them up to focus on higher-level work and much greater productivity.

Exadata is clearly a strategic platform for Oracle that powers its flagship cloud services; the same cannot be said for other vendors pushing hardware for Oracle environments.

Table 1: Oracle Autonomous Database Cloud

- *Auto provisions, secures, monitors, backs up, recovers, and troubleshoots with no human intervention.*
- *Instantly grows and shrinks compute or storage elastically without downtime or human intervention.*
- *Self-Securing, lowering risk.*
- *Adaptive intelligence-enabled cyber threat detection and remediation.*
- *Automatic data encryption, security patches application with no downtime or human intervention.*
- *Self-Repairing, delivering higher availability.*
- *Automated protection from downtime.*
- *Up to 99.995 percent availability; < 2.5 minutes downtime per month, including planned maintenance.*

Oracle has made it crystal clear that there are no plans to make the Oracle Autonomous Database or any exclusive features, such as Automatic Indexing, found on Database 19c, available on IBM Power Systems or any other platform other than Exadata. This provides customers with additional business value for their Exadata investments.

Oracle engineered Exadata to be easily and intuitively managed by a single DBA and continues raising the bar in ease-of-use. IBM did not do the same for POWER9 running the Oracle Database nor do they appear to be advancing in this category.

Ease-of-Use Advantage Oracle Exadata

Architecture

System architectures are optimized to a specific purpose. No architecture can be all things to all applications. There are always tradeoffs. Oracle and IBM take very different approaches to the Oracle Database system architecture.

IBM has generally optimized its Power Systems for general purpose (GP) compute – the traditional or legacy server/network/SAN storage architecture. The latest Power iteration - POWER9 has been optimized for Linux high performance computing (HPC a.k.a. supercomputing), and artificial intelligence (AI). Like all GP servers, the POWER9 GP versions are architected to perform 'good enough' for most applications running concurrently. This means GP Power Systems do not provide best-in-class performance for any given application. It's a different matter when the Power Systems are utilized for HPC or AI. IBM has included functionality into the POWER9 CPU such as tight coupling with the GPU, to accelerate HPC and AI processing. This is demonstrated by IBM's US DoD contracts for super computers. The latest Summit HPC computer is rated as the fastest PETAFLOPS super computer in the world at the time this paper was written. It's also evidenced by IBM's advancements in AI-driven analytics with Watson.

IBM would have IT organizations believe that an architecture that drives impressive supercomputer and AI analytics will naturally be ideal for database OLTP and analytics such as Oracle Database. Nice if it were true. There's simply no factual evidence to back it up. These vastly different applications have very different architectural demands. IBM released POWER9 servers with Linux first and AIX later. IBM's focus going forward is clearly Linux. Oracle Database runs on AIX GP Power Systems, not the Linux versions. Even on IBM AIX GP Power Systems, it runs with less than impressive results. A perfect example of the differences is clustering.

IBM POWER9 Systems for HPC do an outstanding job of clustering specifically for the MPI (Message Passing Interface) protocol. But there are very few Oracle RAC implementations on IBM Power Systems. Both HPC and Oracle RAC are clusters and the HPC implication is that Oracle RAC should run quite well there. And yet it does not.

IBM would need to partner with Oracle to co-engineer Power Systems to be optimized as effectively for Oracle Database as the Oracle Exadata. No such partnering agreement currently exists. As a result, some of the latest Oracle Database innovations such as the Oracle Automatic Indexing found on Database 19c, are not compatible with IBM's Power Systems and likely never will be.

Oracle Database has run on GP architectures since its inception and continues to do so; however, there are performance and functional optimizations that cannot take place unless the hardware is co-engineered with the software. IBM cannot do this with Power Systems whereas Oracle does so with Exadata. Co-engineering enables several Oracle Database CPU-intensive functions to be offloaded to the Exadata storage processors, thereby reducing latency. Executables and processing results are much smaller than the data being processed and thus takes much less time to move. Offloading also frees up the database server CPU cores for additional Oracle Database processes. Performance is noticeably better as storage servers accelerate data-intensive workloads. Exadata offloads multiple CPU-intensive processes including.

- SQL offload
- XML offload
- JSON offload
- Encryption/decryption offload
- RMAN Backup (BCT) Filtering
- Fast Data File Creation

Oracle Database on IBM Power Systems runs on a GP Power System. This is an average fit at best. A deeper dive shows why. There are several individually managed parts to this implementation including the Power CPU (POWER8 or the more recent POWER9), for the AIX operating system (Oracle Database on Power Linux is not supported), PowerVM (AIX Hypervisor), storage network - generally the Brocade 6th generation Fibre Channel (FC), FC host-based adapters (HBA), and SAN connected storage system. The system generally utilizes currently installed or new GP flash or HDD storage. IBM will likely propose their all-Flash System 9000 series if there is an opportunity to sell new storage.

As previously discussed, the GP Power System architecture has challenges when it comes to ease-of-use. That architecture also negatively impacts Oracle Database performance. There is always going to be additional latency between the server and the storage based on speed of light latencies and the FC storage network. These latencies help explain the inconsistent Log File Sync and Log File Parallel Write performance. One reason why Oracle RAC on Power Systems tends to be rare is the use of high latency TCP/IP networking, resulting in relatively slow node failure detection and reconfiguration.

Oracle put extensive research and development (R&D) time, money, and effort into reducing latencies for the Oracle Database within the Exadata architecture. Including but not limited to:

- Moving/offloading many of the executables closer to the data as previously discussed
- Utilizing remote direct memory access (RDMA) built into Infiniband between compute nodes
- Utilizing RDMA Infiniband between compute nodes and storage nodes.
- Smart flash logging as previously discussed

In contrast, IBM put the vast majority of its R&D efforts into its Power CPUs for HPC and to a lesser extent, AI. It is why IBM primarily touts the POWER9 chipset and not Power Systems. IBM implies that its latest and greatest POWER9 System CPU makes it superior for Oracle Database because the CPU has traits[4] that are superior to the latest and greatest Intel X86 CPU utilized in Exadata. The fallacy of that logic is overwhelming. Oracle Database performance is not only about the CPU performance. The CPU is just one of several factors affecting Oracle Database performance. Others include storage, storage services, storage processing, network, memory, interconnect, and more. The POWER9 is the latest version of the Power System CPU. And that one component has NOT been optimized for the Oracle Database. POWER processors were more powerful than Intel when they first came onto the scene many years ago. Since then, that gap has narrowed significantly to the point where POWER9 simply doesn't offer much of an edge. And while the POWER9 runs both AIX and Linux operating systems, the Oracle Database is only supported on AIX Power Systems.

Another argument IBM touts as an advantage against the Intel CPUs used in Exadata is chip-based hypervisor (PowerVM) vs. the software hypervisors (KVM, ESXi, etc.) that consume resources. Their argument has some merit when the Oracle Database on Exadata is running in a VM under a hypervisor. It has no validity when the Oracle Database is running in a Docker container. Containers eliminate the hypervisor and utilize less resources. And just like VMs they can run the applications that utilize the Oracle Database.

Additional key aspects of system architecture are open versus proprietary and future sustainability. What is IBM's roadmap? Can it be sustained past the current generation?

Oracle Exadata is based on the market standards of off-the-shelf hardware and software such as the Intel x86 CPUs, Mellanox Infiniband, SSDs, HDDs, Linux OS, Docker containers, etc. Intel CPUs power the vast majority of servers, desktops, and laptops the world over. There is typically more than one source for the vast majority of system components. Exadata continues to be updated with the latest and greatest components in the market. The future sustainability is assured. Oracle has additionally made backwards compatibility a hallmark of Exadata. It has been and continues to be simple to add racks of new versions of Exadata to previous generations. It has only been recently that the latest X7 dropped compatibility with the decade old X1 and X2. The future is openly tied to the industry which continues to advance.

The situation with Power Systems is considerably more tenuous. Only IBM uses Power Systems and their market is shrinking. Power CPUs are proprietary with no second source. Lower and shrinking volumes make them significantly more expensive. IBM's investment in new hardware has been on a decline while they have been divesting themselves of complete lines of hardware. IBM does have a roadmap beyond the POWER9 CPU but there are no commitment dates at the time this document was written. There is considerable doubt as to whether or not IBM can generate enough market to sustain the Power Systems architecture.

Oracle Exadata is architected from the ground up to get the most performance, reliability, and ease-of-use out of the Oracle Database based on open industry standard hardware and software. It is co-engineered with the Oracle Database to provide unique and significant database advantages as seen in Appendix A.

IBM's GP Power Systems are based on a proprietary CPU and OS that based on IBM reporting, has a shrinking demand. There is no co-engineering with Oracle Databases. Exadata is a much better architectural fit for Oracle Database than Power Systems.

Architecture Advantage Oracle Exadata

TCO

Total Cost of Ownership (TCO) comparisons are frequently perceived as complicated, difficult to explain, and can cause situations to devolve down a rat hole. It does not have to be that way. The key is normalizing the data so that the comparisons are apples-to-apples and not apples-to-grapes. It also requires gaining agreement on assumptions. This section goes through a detailed process explaining how to normalize and calculate TCO. The table below summarizes the results. The details that explain how these numbers were

[4] More cores per socket, more threads per core, intelligent threading enabling some workloads to be executed utilizing 1-8 threads per core automatically or determined by the sysadmin, and intelligent caching, which dynamically varies cache utilization to minimize overall cache latency.

achieved are important to understand and to defend the TCO comparisons and are detailed in the table below for reference.

Cost Comparison Summary			
	POWER9	**Exadata**	**Advantage**
Oracle Database Licensing Costs	1X	**62%**	Exadata
Oracle Database Implementation Costs	1X	**3.6%**	Exadata
Compute & Storage Networking Costs	**50%**	1X	POWER9
Storage Costs	1X	**62%**	Exadata
Patching & Upgrade Costs	1X	**2.8% to 4.2%**	Exadata

TCO is the combination of capital expenditures (CapEx) and operating expenditures (OpEx) over a specific time period. CapEx includes the purchase price of the system – compute, storage, storage network – and any additional supporting infrastructure. OpEx includes all software licensing costs, personnel or professional services for implementation/patches/upgrades/operations, power, cooling, fixed overhead allocation, and premium maintenance.

When using TCO as a comparison metric, it is crucial to ensure the configurations are equivalent. In other words, the costs must be normalized for accuracy, fairness, and true apples-to-apples. There are always assumptions to be made and it is crucial to eliminate those that tilt the playing field. For example: IBM has been known to exclude or minimize storage costs (CapEx and OpEx) when Power Systems TCO is compared to Exadata. IBM assumes that Power System running Oracle Database is connecting to already installed SAN storage. Even when that is true, there is still a storage requirement. The storage has to be provisioned for the Oracle Database. That provisioned storage is no longer available to other applications or usages. It is for Oracle Database consumption only[5]. The IT organization may have already purchased that storage, but all of the costs associated with the provisioned Oracle Database capacity will be allocated to the Power Systems TCO. Oracle Exadata on the other hand, always includes its own storage as part of the co-engineered system. Those costs are part of the Exadata hardware costs.

Comparing TCO in an objective manner is difficult at best and easy to manipulate. To make TCO comparisons objective requires normalizing costs so they are actually comparable. Note that there are several discretionary cost items such as premium maintenance that's above and beyond what comes with the hardware warranties. Oracle Database licensing today is subscription based so the maintenance is built into the annual license. In addition, there are OpEx cost items that vary by location such as power and cooling costs. Although these are legitimate costs, they are not included in this comparative TCO. This comparative TCO focuses on the biggest knowable costs. The largest cost as a percentage of the TCO is the Oracle Database software license. The other big-ticket items include: Oracle Database implementation costs; compute and storage networking costs; storage costs – the second biggest cost aspect of TCO; patching and upgrade costs.

Normalizing Oracle Database Licensing

In more multi-user environments, Oracle Database is licensed per CPU core. Oracle utilizes a processor core factor table. That factor is multiplied against the total number of cores and the Oracle Database license per core. The Oracle Database processor core factor for Intel cores (utilized in Exadata) is .5. For all Power cores the core factor is 1. The number of processing cores for each system must be determined for an equivalent amount of CPU processing in order to provide licensing normalization.

Normalizing Cores

Published benchmarks from IBM demonstrate that each POWER8 core gets best case approximately 20-24% more performance than the Intel X86 CPUs utilized in Exadata. In other words, a POWER8 core equals ~ 1.24 Intel cores. The POWER9 core benchmarks show approximately a 40% improvement over the POWER8 core. Therefore, it is ~ 1.736 X that of an Exadata Intel core. Theoretically, the Oracle Exadata would need 1.736 the number of cores as a POWER9 running Oracle Database. IBM TCO comparisons typically round that up to 2X. Both are incorrect.

Remember the Exadata architecture takes advantage of underutilized Exadata storage cell Intel cores for co-processing by offloading CPU-intensive functions. Power Systems cannot do this. The number of total Exadata storage cell Intel cores generally equals approximately 75% of the number of database compute cores. The number of idle Exadata storage cell intel cores varies but conservative have at least 50% available at any given time. That makes at least 37.5% more cores available for co-processing. Normalizing these numbers means that each IBM POWER9 System core processing actually equates into approximately 1.25X that of each Exadata Compute node Intel core because of the additional storage cell intel cores available for co-processing. For each IBM POWER8 System core processing equates into approximately .90X that of each Exadata Compute node Intel core because of the additional storage cell intel cores available for co-processing. Note: Oracle does not charge any Oracle Database licensing fees for the storage cores.

[5] Storage systems can technically be thin provisioned so available capacity can be shared; however, doing so with any database is not recommended. If the actual capacity required becomes unavailable the database can become corrupted and mission critical data lost.

Calculating Oracle Database Licensing[6] Costs
POWER9 System Oracle Database Licensing = Total # of POWER9 cores required * 1 * Oracle Database license/core.
Exadata Oracle Database Licensing = Total POWER9 System Oracle Database licensing * 1.25 * .5 = **62.5%**
POWER8 System Oracle Database Licensing = Total # of POWER8 cores required * 1 * Oracle Database license/core.
Exadata Oracle Database Licensing = Total POWER8 System Oracle Database licensing * .9 * .5 = **45%**
Oracle Database licensing costs are frequently 70% of the total system cost.
Oracle Database Implementation Costs
Implementation costs can come from professional services, internal IT personnel, or a combination of both, and time to market. These costs are frequently left out of many TCO comparisons. However, the time required to implement these two systems is vastly different. Oracle Exadata comes fully assembled. Implementations predictably take no more than 3 days and usually less. IBM Power Systems implementations take considerably longer because none of the components are integrated. IT organizations spend weeks to months designing and planning the system even before the hardware shows up. The Power System, storage, and FC fabric have to be installed and cabled up upon delivery, the AIX OS installed, the Oracle Database installed, and then the entire system configured, trouble shot (there are always issues to be resolved), and tuned. IBM customers report deployment times to production ranging from a low of 3-4 weeks to 4-6 months.
These cost differences are huge. Assuming the cost per IT professional per hour is equal, the Power Systems require at a minimum 4X the personnel and substantially more costly hours. Exadata needs the DBMA team to deploy. Power Systems need the DBMA team, plus the FC storage networking team, storage admin team, a facilities admin team for cabling-conduit-power, rack, etc. Assuming a best-case scenario for Power Systems and a worst-case scenario for Exadata, the personnel implementation costs = (21 days/3 days) * at least 4X personnel = **28X** more commercial costs for Power Systems. Or put another way, Exadata personnel implementation costs run **1/28th** that of Power Systems running Oracle Database.
That's a major difference and it's not the biggest implementation difference. That 3-4-week time difference to production can be huge. It can mean the difference in **$ millions** in revenue gained wit Exadata or lost in the case of Oracle Database on Power Systems. These costs are non-trivial.
Compute and Storage Network Costs
Compute costs are the smallest of the big cost numbers typically accounting for no more than 15% of the total Exadata cost and less than 3% of the Power System costs. IBM has priced Power Systems very competitively with Intel based servers. Compute costs are an area where IBM Power System, POWER8 or POWER9, will come in at approximately **50%** or less than Oracle Exadata Compute nodes on a MSRP basis.
Storage Costs
Storage cost comparisons are a tad tricky because of architectural performance differences.
Higher latency issues with the Power Systems between the compute nodes and the storage may require faster, lower latency, expensive storage with limited storage functionality such as the IBM Flash System 900. Or the IT organization may require more storage functionality such as snapshots, dedupe, compression, replication, thin provisioning, etc. that's more expensive, and has higher latency such as the IBM Flash System 9000.
But many Oracle Database functions don't necessarily need all-flash and are well suited for lower cost HDDs. Exclusive Oracle Database performance functionality found only in Exadata means use of HDDs will not reduce database OLTP, OLAP, or DW performance. Oracle Exadata storage cells can be all flash, all HDD, or combinations. Doing an apples-to-apples comparison will come down to the storage performance required. Generally, IBM storage tends to be expensive compared to other systems in the market. IBM storage compared to equivalent Exadata storage cells runs approximately **62%** more expensive. Because storage is a much bigger cost than compute, IBM essentially gives back all of the savings it has from the compute nodes and much more because of its storage cost.
Patching and Upgrade Costs
As previously discussed, Oracle delivers a single Exadata patch/upgrade per quarter. That patch/upgrade covers all aspects of the Exadata system at once including the Oracle Database. The process is highly automated taking little time with minimal to no disruption.
IBM does not have a single patch/upgrade. There are patches/upgrades for the Oracle Database, Power System, AIX OS, FC HBAs, brocade FC switches, and the storage system or systems. These patches/upgrades are not coordinated among the different vendors, IBM does not aggregate nor automate them, and they come at different times. There can be quite a few in a quarter. Even if there is only one per major subsystem, that's 6 unique patches/upgrades that haven't been tested together. Many of those patches/upgrades are disruptive or potentially disruptive, requiring scheduling in off hours on weekends and holidays. In many parts of the US and the world, the personnel must be paid overtime for those hours which is typically time and a half.

[6] All of these calculations are relatively close approximations and actual configuration numbers will vary somewhat.

Calculating the cost reveals Power Systems demand at least 6X the effort and time of Exadata systems. And Power Systems require at least 4 different admins (DBA, Power System admin, FC storage network admin, and storage admin) instead of 1 for each patch. This is to ensure any given patch doesn't cause issues for the other system subsystems. That means Power Systems require a minimum of 6 patches * 4 admins vs Exadata with 1 patch * 1 admin time each quarter. In other words, Power Systems consume 24X the personnel time for patches and upgrades and that assumes none of the patches have to be backed out and redone at a later date. And Power Systems cost **24X** to 24 * 1.5 (time and a half for overtime) or **36X** the cost of Exadata for patches/upgrades.

By any objective analysis, the Oracle Exadata TCO comes in considerably less than the Power Systems TCO. Each situation will vary. However, expect Exadata TCO to generally be significantly less than Power Systems. How much less will depend on the variables described in the chart.

TCO Advantage Oracle Exadata

Summary and Conclusion

Both IBM Power POWER9 Systems and Oracle Exadata run Oracle Database. IBM's POWER9 is a general-purpose system without any specific optimization for Oracle Database. Oracle Exadata systems are specifically co-engineered with the Oracle Database to get the most possible out of the Oracle Database. As a result, of the two only Oracle Exadata brings the **HEAT**: high-performance, ease-of-use, architecture, and TCO. The best choice for running the Oracle Database is clearly the Oracle Exadata Database Machine.

One side note. When an IBM Power Systems user switches their Oracle Database to Exadata, they also free up substantial Power System resources for other applications getting a significant bonus.

For More Information on Oracle Exadata

Go to: http://www.oracle.com/us/products/database/exadata/overview/index.html

Paper sponsored by Oracle. **About Dragon Slayer Consulting:** Marc Staimer, as President and CDS of the 21-year-old Dragon Slayer Consulting in Beaverton, OR, is well known for his in-depth and keen understanding of user problems, especially with storage, networking, applications, cloud services, data protection, and virtualization. Marc has published thousands of technology articles and tips from the user perspective for internationally renowned online trades including many of TechTarget's Searchxxx.com websites and Network Computing and GigaOM. Marc has additionally delivered hundreds of white papers, webinars, and seminars to many well-known industry giants such as: Brocade, Cisco, DELL, EMC, Emulex (Avago), HDS, HPE, LSI (Avago), Mellanox, NEC, NetApp, Oracle, QLogic, SanDisk, and Western Digital. He has additionally provided similar services to smaller, less well-known vendors/startups including: Asigra, Cloudtenna, Clustrix, Condusiv, DH2i, Diablo, FalconStor, Gridstore, Nexenta, Neuxpower, NetEx, NoviFlow, Pavilion Data, Permabit, Qumulo, SBDS, StorONE, Tegile, and many more. His speaking engagements are always well attended, often standing room only, because of the pragmatic, immediately useful information provided. Marc can be reached at marcstaimer@me.com, (503)-312-2167, in Beaverton OR, 97007.

Appendix A: Exadata Unique Oracle Database Functionality

Smart Scan
Automatic Indexing

Bloom Filtering

Smart Flash Logging

Database Aware FlashCache

Smart Write Back

Reverse Offload

Fast Node Death Detection

Fast Network Failure Detection

Reduced Brownout for Instance Failure

I/O Resource Manager

Pure Columnar Flash Cache

Hybrid Columnar Compression (HCC)

Backup Change Tracking Filtering in Storage

Cross Node Parallel Query

ExaFusion Direct-to-Write Interconnect

In-Memory Fault Tolerance

In-Memory on Active Data Guard

Automatic In-Memory

In-Memory External Tables

In-Memory Row Store

Faster Smart Scans using Column-level Checksum

Faster Row Store Ingest

Smart Fusion Block Transfer

Appendix B: Exadata Many HA Features Supporting Most Stringent SLAs

Fast node & cell death detection

Fast network failure detection

Redundancy protection on cellsrv shutdown

Reduced brownout for instance recovery

ILOM hang detection & repair

Redundancy protection on cell shutdown

Automatic ASM mirror read on IO error corruption

IO error prevention w/Exadata disk scrubbing / ASM corruption repair

Exadata HARD

Corruption prevention w/HARD support

Elimination of false positive drive failures

Redundancy Check during power down

Blue OK-to-remove LED light notification

Active Active IB Network

Exadata Smart Write Back, Smart Flash Logging, Smart Scan, & Reverse Offload

Fastest Redo Apply & Instance Recovery

Efficient resilver rebalance after flash failure

I/O latency capping for reads & writes

Cell IO timeout threshold

Smart Write Back Flash Cache persistence

I/O & Network Resource Management

Health factor on predicatively failed disks

Disk confinement

IO hang detection & repair

Cell-to-Cell offload for Disk Repair

Cell-to-Cell Rebalance Preserves Flash Cache

Exadata Elastic Configuration

Drop hard disk for replacement

Drop BBU for Replacement

Appliance mode support

Cell Alert Summary

Flash & Disk Life Cycle Management Alerts

Automatic LED support for disk removal

Auto online

Auto disk management

Priority rebalance support

EM failure reporting

Failure Monitoring on database servers

Updating database nodes with patchmgr

Optimized & Faster Exadata Patching

Custom Diagnostic Package for Cell Alerts

VLAN support & Automation

Exachk - full stack health check with critical issues alerts

Smart Rebalance of High Redundancy Disk Groups