



White Paper

Virtualization—The Catalyst for Change: Best Practices for Virtualizing IBM DB2 with Technologies from VMware, Intel, and NetApp

Peter Kokosielis and Sunil Kamath, IBM
Kshitij A. Doshi, Intel
Jawahar Lal, NetApp
Bob Goldsand, Jeffrey Buell, and Robert Campbell, VMware

August 2010 | WP-7109

EXECUTIVE SUMMARY

From an architecture and operational efficiency standpoint, virtualization offers enormous benefits for customers who virtualize compute resources in their data centers. The question to ask is not “Can we deploy virtualization technologies for the business application?” but “How fast can we deploy?” This white paper is the result of a collaboration among IBM, VMware, NetApp, and Intel. It describes test configurations, performance results, and best practices for deploying IBM DB2® on Linux® in a VMware vSphere™ environment running on servers with Intel® Xeon™ 5500 and 5600 Series processors and connecting to NetApp® storage systems.

TABLE OF CONTENTS

1	CATALYST FOR CHANGE	3
2	VIRTUAL DATA CENTER SOLUTION	4
2.1	IBM DB2 FOR VIRTUALIZED ENVIRONMENTS	4
2.2	VMWARE VSPHERE	4
2.3	INTEL XEON 5500 AND 5600 SERIES PROCESSORS	6
2.4	STORAGE VIRTUALIZATION WITH NETAPP	8
3	SOLUTION ARCHITECTURE	11
4	TESTING AND PERFORMANCE RESULTS	15
4.1	DECISION SUPPORT WORKLOAD PERFORMANCE	16
4.2	OLTP WORKLOAD PERFORMANCE	18
4.3	VCPU OVERALLOCATION TEST	19
5	BEST PRACTICES	20
5.1	INTEL BEST PRACTICE	20
5.2	NETAPP BEST PRACTICES	21
5.3	VMWARE BEST PRACTICES	22
5.4	DB2 BEST PRACTICES	23
6	CONCLUSION	24
7	REFERENCES	24
7.1	INTEL REFERENCE INFORMATION	24
7.2	NETAPP REFERENCE INFORMATION	25
7.3	VMWARE REFERENCE INFORMATION	25
7.4	IBM DB2 REFERENCE INFORMATION	26
8	APPENDIXES: CONFIGURATION PARAMETER SETTINGS	26
8.1	APPENDIX A: DSS WORKLOAD SETTINGS	26
8.2	APPENDIX B: OLTP WORKLOAD SETTINGS	30
8.3	APPENDIX C: VCPU OVERALLOCATION OLTP WORKLOAD SETTINGS	32

1 CATALYST FOR CHANGE

Virtualization is reshaping the data center world and application deployment environments. With the cost of doing business rising every day, organizations are looking for ways to reduce overhead and consolidate resources in a single, easy-to-manage environment. This trend is producing demand for more integrated technologies and fewer parts to manage. As businesses turn to consolidating servers, storage, and networking hardware, they see the benefits of:

- Increasing resource usage
- Reducing space needs
- Reducing power consumption
- Reducing cooling expenses
- Reducing the TCO
- Increasing the ROI of the data center
- Improving IT agility
- Reducing provisioning time

Business processing and data analysis applications are among the workloads being migrated on a large scale from isolated execution on physical servers to consolidated execution on virtual servers. These workloads are rich in their use of databases, either directly for On Line Transaction Processing (OLTP) and Decision Support System (DSS) or through other functions such as enterprise resource planning (ERP) and customer relationship management (CRM) processing. A recent International Data Corporation (IDC) white paper, "[Optimizing Hardware for x86 Server Virtualization](#)," states that 38% of all spending on virtual machines (VMs) in 2009 was for business processing workloads, and over two-thirds of current virtual servers support production workloads. Initially motivated by resource and energy-sharing efficiencies from server consolidation, the use of virtualization for mission-critical purposes such as database services is increasingly driven by the business-agility benefits of more fluid deployments, elastic resourcing, failover execution, and policy automation. Complementary innovation in processors, hypervisors, software, and I/O subsystems underlies the delivery of these benefits.

This white paper describes solution architecture and best practices for deploying IBM DB2 databases on Linux in a VMware virtual infrastructure environment running on a server equipped with Intel processors and connected to NetApp unified storage. Test results from performance and scalability tests performed under DSS and OLTP workloads are also described. Some key benefits of the overall solution are:

- **Reduced costs with VMware virtualization.** With virtualization, customers can reduce the number of physical servers, which not only lowers hardware operation and maintenance costs but also decreases the physical data center space use and lowers overall power consumption and cooling expenses. A VMware virtualization technology with Intel processors and NetApp unified storage can unlock the full power of the hardware for DB2 environments by running multiple workloads and operating systems on each server. This consolidation can provide a cost-effective solution and potentially higher ROI when compared to deployments without virtualization.
- **Intel processors optimized for virtualization.** Intel processors and platforms implement virtualization assists to remove or reduce the burden on hypervisors, to properly intercept sensitive operations, and to preserve the illusion of unshared hardware. From one generation to the next, they also implement optimizations to reduce the latencies of transferring control from guests to host by reducing the number of cases in which such transfers need to take place. These changes reduce overhead from virtualization and improve the ability to scale performance when consolidating increasing numbers of virtual servers on shared infrastructure.
- **NetApp unified and efficient storage solutions for virtualization.** NetApp's unified storage runs with Data ONTAP[®], a proprietary operating system that is highly optimized for virtualized environments. Customers can deploy DB2 databases on NetApp FAS or V-Series storage and leverage existing networking infrastructure such as Fibre Channel (FC), Fibre Channel over Ethernet (FCoE), iSCSI, and Network File System (NFS). The unified nature of NetApp products offers a cost-effective storage solution. NetApp FAS and V-Series storage arrays have been fully tested and certified for use in FC- and IP-based VMware environments. In addition, by leveraging NetApp storage efficiency, intelligent caching capabilities, and ease of manageability, customers can save on their storage investment without trade-offs.

- **High availability.** A platform enabled by VMware virtualization technology can provide high availability (HA) for DB2 environments without the need for clustering at the VM level. VMs are no longer tied to the underlying server hardware and can be moved across servers at any time with VMware VMotion™. VMware HA offers greater levels of availability over solutions designed to protect just the server.
- **DB2 autonomies leveraging virtualization.** The DB2 database's ability to adapt its resource requirements automatically to dynamic environments makes it a great citizen within a VM framework. Through features such as deep compression and scan sharing, DB2 databases make efficient use of I/O, reducing the burden on hypervisors in managing resources.

Various DB2 deployment scenarios and configurations were tested and are described in Section 4.

2 VIRTUAL DATA CENTER SOLUTION

The solution presented in this white paper demonstrates how customers can leverage virtualization technologies from VMware, Intel, and NetApp for their IBM DB2 environments. This section highlights the main features of IBM DB2 9.7, VMware vSphere, Intel processors, and NetApp storage that facilitate development of highly optimized virtual application solutions.

2.1 IBM DB2 FOR VIRTUALIZED ENVIRONMENTS

The IBM DB2 9.7 data server is optimized to deliver industry-leading performance, availability, and reliability across multiple database workloads while lowering administration, storage, and server costs. The DB2 data server is engineered to perform well in virtual environments through a variety of features:

- **Autonomic self-tuning memory and automatic configuration parameters.** With autonomic self-tuning memory and automatic configuration parameters, DB2 can configure its heaps and memory usage in dynamic environments. This capability can be leveraged in VM templating and cloning so that minimal database configuration is required when rapidly deploying multiple DB2 instance VMs, even if individual instance VMs are configured for different resource capacity.
- **DB2 deep compression.** DB2 9.7 deep compression provides row compression of data and index pages, as well as temporary tables and other objects, which allows DB2 9.7 to reduce substantially the amount of storage and buffer memory required for a given-size database. Thus, I/O throughput requirements on a system, which traditionally have been very expensive to virtualize, can be reduced.
- **Scan sharing.** Scan sharing is a feature new to DB2 9.7. It allows subsequent large queries to exploit the scans that may have been performed for previous queries, thus reducing the need for disk I/O in virtual environments.

A DB2 data server uses the operating system-based scheduler for its server threads so that it can react dynamically to changes in the CPU capacity of a VM. In addition, a DB2 data server can monitor its own resource consumption to minimize allocation when usage is low, so that a virtual machine monitor can distribute resources more efficiently to other VMs in need.

Most important, DB2 9.7 differentiates itself from other database management systems by offering subcapacity licensing to enable customers to consolidate their infrastructure effectively and flexibly and to reduce their TCO. Customers can track and manage their own software license usage and pay only for the physical capacity that was used.

IBM DB2 engineering works closely with third-party vendors such as VMware, NetApp, and Intel to bring to market best-of-breed solutions that deliver maximum value to customers.

2.2 VMWARE VSPHERE

VMware vSphere is the leading virtualization solution, providing multiple benefits to IT administrators and users. VMware vSphere provides a layer of abstraction between the resources required by an application and operating system, and the underlying hardware that provides those resources. VMware vSphere provides the following benefits:

- **Consolidation and infrastructure optimization.** VMware virtualization technology allows you to consolidate multiple physical servers. This reduces the total required number of servers and related IT hardware in the data center, with little or no decrease in overall performance. VMware technology also

lets organizations pool common infrastructure resources to optimize their use. Through reduction in physical servers, real estate, power, cooling requirements, and IT costs, customers can reduce their own TCO for running a wide range of business- and mission-critical applications.

- **Ease of provisioning and management.** VMware infrastructure encapsulates an application in a self-contained image. With the addition of VMware vCenter™ management software, organizations can use templates to create golden masters of VM environments. The golden masters then can be duplicated or moved using VMware VMotion technology. VMware vSphere, along with vCenter and VMotion, provide new ways to manage IT infrastructure. They can help you spend less time on repetitive tasks such as provisioning, configuration, monitoring, and maintenance.
- **Increased application availability and business continuity.** VMware HA can shorten unplanned downtime and provide higher service levels to an application. In the case of an unplanned hardware failure, VMware HA restarts affected VMs on another host in a VMware cluster. Thus, VMware HA, along with other VMware technologies, helps minimize planned downtime and lets organizations recover quickly from unplanned outages. Entire virtual environments can be backed up and migrated with no service interruption.

Figure 1 illustrates the VMware infrastructure solution.

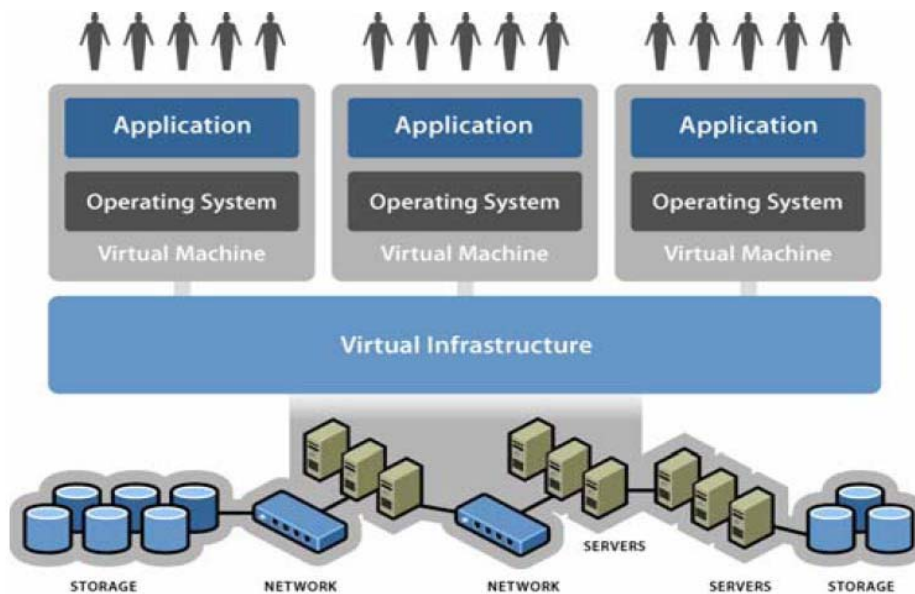


Figure 1) VMware infrastructure solution (graphic supplied by IBM).

Running DB2 databases on VMware vSphere can improve overall availability of the database system while reducing capital and operational costs. In addition, vSphere provides the following benefits to DB2 virtualized deployments:

- **Isolation.** VMware ESX provides a safe, secure, and scalable environment for running multiple DB2 instances, DB2 databases, and multiple guest operating systems on the same physical server.
- **Rapid provisioning.** VMware ESX provides a nimble environment for rapidly provisioning DB2 servers. This rapid provisioning is the result of VMware's unique encapsulation feature of using templates and clones (and leveraging built-in DB2 features such as the db2relocatedb utility).
- **Change management.** Virtualizing DB2 allows you to migrate DB2 VMs across systems during data center maintenance operations and other changes. VMware virtualization also allows you to roll back instantly application VMs during problem resolution for any previous change applied.
- **Improved availability and recovery.** You can use VMware VMotion, VMware HA, and VMware Fault Tolerance to increase greatly the uptime of IBM DB2 servers. Virtualized deployments can minimize planned downtime and enable quick recovery from unplanned outages, providing the ability to back up and migrate entire virtual environments with little or no service interruption.

DB2 data servers can be efficiently consolidated, rapidly provisioned, and highly optimized in your virtualized data center by leveraging the power of infrastructure virtualization solutions provided with VMware vSphere. As illustrated in Figure 2, you can centrally manage your infrastructure from a single console.

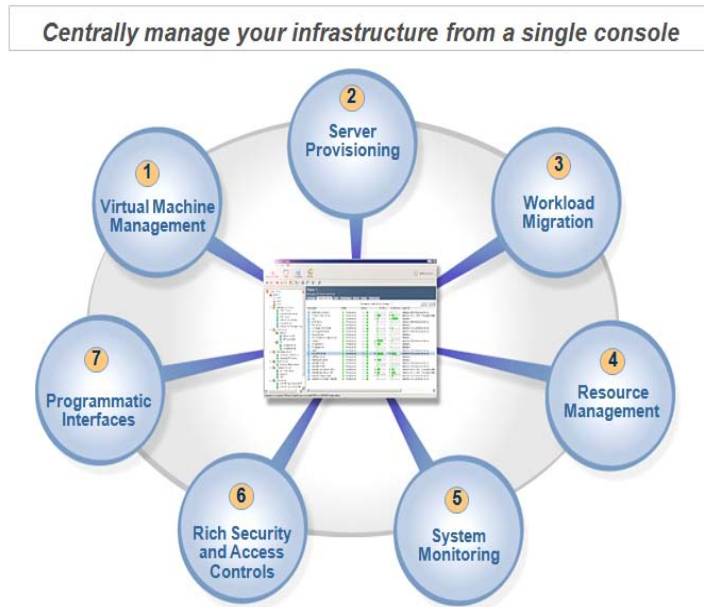


Figure 2) VMware vSphere and vCenter infrastructure management (graphic supplied by IBM).

Successfully deploying DB2 data servers on a VMware virtual infrastructure is not significantly different from deploying DB2 data servers on physical servers. DBAs can leverage their current understanding of database application deployments while obtaining all the benefits associated with virtualization.

A large majority of DB2 applications, including those using very large databases, can be deployed on a virtual infrastructure. When run on VMware vSphere, those applications match and can even exceed the performance achieved on physical servers. Each VM can scale to 8 virtual CPUs (vCPUs), 256GB of memory, and support I/O-intensive applications. Your applications achieve far greater scalability per physical host by scaling out on multiple VMs, which is often the only way to leverage the capacity of the latest multicore servers. To minimize server requirements, you can deploy DB2 data servers on most existing commodity x86 servers with VMware virtualization. VMware vSphere and VMware ESX 4 can also take advantage of newer virtualization support features, such as those built in to the latest Intel processors, increasing virtualization performance to rival the performance of any high-end physical server.

2.3 INTEL XEON 5500 AND 5600 SERIES PROCESSORS

New processor architectures from Intel have incorporated several technology features that facilitate seamless and low-overhead virtualization. Intel designs provide high core counts per chip, together with increased memory capacities, addressability, and throughput. This combination of features is delivered on industry-standard components and is well supported across the breadth of commonly available hypervisor and operating system runtimes. These developments have rapidly accelerated the adoption of system virtualization and broadened its use. Once limited to vertical, vendor-specific environments characteristic of mainframes, virtualization is rapidly moving to horizontal, multivendor, standards-based mainstream usage.

The following list describes the features of Intel platforms and processors that assist system virtualization. The descriptions explain the performance and reliability impacts, and describe how best to benefit from continually increasing processor capabilities and energy efficiencies.

- **Intel VT-x.** A set of CPU features that make it possible for guest operating systems to run unmodified at their intended privilege level and with virtualization overheads that are small and decline over time. Recent Intel processors implement caching and other optimizations that accelerate the transfer of control into and out of the hypervisor and reduce the frequency of hypervisor interventions. In particular,

the Intel Xeon 5500 series processors used in the solution designs described in this white paper implement the following optimizations:

- **Extended Page Tables (EPTs).** These tables, maintained by software but used by hardware, provide translations from guest physical memory addresses to host server physical memory addresses. This translation allows a server to use both the page tables set up by a guest operating system in a VM and the EPT values set up by the host hypervisor; the server does not need to use a prior method in which the hypervisor maintains the full end-to-end translation structures by shadowing any modifications the guest operating system performed on its own. EPTs remove one of the biggest sources of performance overhead for address space-intensive applications such as databases, Web servers, and file servers. VMware uses EPT by default if it is available, but it provides a virtual machine-by-machine override controlled by VMware options. NetApp recommends that you use EPTs both for the improved performance they provide and for reducing the amount of memory that is used for shadow page tables in the absence of EPTs.
- **Virtual processor ID (VPID).** With this capability, each translation look-aside buffer (TLB) entry is associated with a VPID tag, which makes it unnecessary to flush the TLBs from an executing VM upon a transfer of execution context. For short-lived transitions into and out of the hypervisor, the VPID capability improves the efficacy of TLBs, thereby reducing the number of page table walks that are otherwise necessary.
- **Flex priority, or APIC TPR virtualization.** This feature virtualizes the Advanced Programmable Interrupt Controller (APIC) Task Priority Register (TPR), which is a register consulted by the APIC hardware to determine whether a particular task can be interrupted by an arriving system interrupt. Virtualizing the APIC TPR is done through a shadow register that captures any priority modifications performed by the guest. This is followed by a filtering action that determines whether or not the hypervisor needs to obtain control for reflecting the action by the guest into the real TPR that is consulted by the APIC hardware. Provisioning this functionality greatly reduces the possible harm from device drivers or operating systems that modify the TPR excessively. Ordinarily, this feature should be enabled by default. NetApp recommends its use.
- **Flex migration.** Flex migration provides compatibility with previous versions of Xeon processors, beginning with the Intel Xeon 5100 and continuing with Intel Xeon 7300 and Intel Xeon 7400. This compatibility allows you to migrate VMs across a diverse pool of Intel servers and may be particularly useful for configurations in which mission-critical database execution is staged over multiple generations of physical servers.

The Intel Xeon 5500 processors used in the solution design evaluated in this white paper also support a set of optimizations that are described as Virtualization Technology for Directed I/O and abbreviated as Intel VT-d. We did not employ VT-d for the performance tests. VT-d allows a system administrator to bind I/O devices (such as network and storage interface ports) to VMs and then bypass the hypervisor-based processing or copying of data during I/O operations to or from the bound devices. Such bypass reduces the processing overhead because the hypervisor is bypassed. It also improves isolation between the VMs by enforcing page permissions in hardware. It is recommended under some situations, such as those in which network-intensive workloads are unable to use the available network throughput under virtualization because of the additional latency of I/O processing and context switching in the hypervisor.

ARCHITECTURE FEATURES OF INTEL XEON 5500 AND HIGHER SERIES PROCESSORS

Intel Xeon 5500 and higher series processors use the following features to consolidate database applications through virtualization:

- Hyperthreading (HT)
- Turbo mode processing

HT couples with the high number of cores per chip. With the HT feature, Intel provides two logical CPUs for each physical core through transparent multiplexing of computational threads in each core. This allows per-core resources such as TLBs, on-chip caches, and functional units to be multiplexed as needed. The much increased hardware concurrency makes it easier for a hypervisor to schedule a larger number of virtual CPUs on the same host. It also reduces the likelihood of scheduler thrashing when a spike in the processing load might otherwise cause the scheduling of much larger numbers of active vCPUs than the number of physical CPUs available. The energy cost from hyperthreading is minimal because the processor schedules resources that are otherwise idle. The HT feature is normally enabled by default in Intel Xeon processors.

Under turbo mode, if the total available thermal budget is larger than that being used by the active cores, the active cores are permitted to execute at higher than nominal frequencies. This frequency increase makes it possible for workloads that are particularly latency sensitive and that are not uniformly well threaded to take advantage of the ups and downs in processor use to obtain modest but nontrivial improvements in their rate of processing. For virtualization-based consolidation, enabling turbo mode may prove useful if the additional expense of energy required at the higher frequency is tolerable.

The Intel Xeon 5500 series processor used in this study was launched in early 2009. This processor has since been followed by Intel Xeon 5600 and Intel Xeon 7500 series processors. They have increased the per-chip core count by 50% and 100%, respectively, and implemented larger per-chip caches. These architectural advances have brought corresponding performance advantages to a number of virtualized and bare metal workloads. All these processors implement point-to-point, high-bandwidth, and low-latency serial interfaces between processors and between processors and memory modules. The latter is made possible through memory controllers that are integrated into the processor chips. These improvements from prior generations of Intel Xeon processors make it possible to handle gracefully the potentially larger cache misses resulting from consolidated multiple virtual server execution. With each new generation of processors, Intel also has been allowing larger ranges of virtual and physical memory spaces to be addressed, which makes it easier to support the memory demands that are typical of large-scale data processing applications.

2.4 STORAGE VIRTUALIZATION WITH NETAPP

With the exponential growth of data and digitized information, storage has become a critical component of doing business. NetApp understands the challenges that customers face and has designed business solutions that integrate well with solutions from technology partners such as IBM, VMware, and Intel. NetApp has addressed the need to do more with less in a virtual environment. By using a combination of [RAID-DP](#)[®], [deduplication](#), [FlexVol](#)[®], [FlexClone](#)[®], thin provisioning, [FlexShare](#)[®], [Flash Cache](#), and [Snapshot](#)[™] technologies, plus tools such as SnapManager[®] for Virtual Infrastructure (SMVI), VMware vCenter plug-ins Virtual Storage Console (VSC), and the Rapid Cloning Utility (RCU), NetApp enables customers to achieve storage savings and operational efficiency in a virtual environment.

The following list provides brief explanations of these business solutions and tools:

- **RAID-DP.** RAID-DP provides performance that is comparable to that of RAID 10, with pricing comparable to RAID 4 and much higher resiliency than that of either RAID 10 or RAID 4. It provides protection against double disk failure, where RAID 5 can protect only against single disk failure. Because of increased reliability and decreased cost compared to similar environments, RAID-DP offers businesses a compelling total cost of ownership storage option without putting their data at increased risk. For more information about RAID-DP, see [“TR-3298: RAID-DP: NetApp Implementation of RAID Double Parity for Data Protection.”](#)
- **Thin provisioning.** Thin provisioning is a method of storage virtualization that allows storage administrators to address and oversubscribe storage in the same way that server resources such as RAM and CPU can be overprovisioned in a virtual environment, providing a level of additional storage on demand. Thin-provisioned storage is treated as a shared resource pool and is consumed only as each individual VM needs it. This sharing increases the overall use rate of storage by eliminating the unused but provisioned areas of storage that are associated with traditional storage methods. By allowing as-needed provisioning and space reclamation, thin provisioning can result in better storage use and smaller total capital expenditures on storage infrastructure. For details about thin provisioning, see [“TR-3563: NetApp Thin Provisioning.”](#)

NetApp thin provisioning extends VMware thin provisioning for virtual machine disk files (VMDKs) and allows logical unit numbers (LUNs) that are serving Virtual Machine File System (VMFS) datastores to be provisioned to their total capacity, yet consume only as much storage as is required to store the VMDK files, which can be of either thick or thin format. In addition, LUNs connected as Raw Device Mapping (RDM) can be thin provisioned. NetApp recommends that when you enable NetApp thin-provisioned LUNs, you deploy these LUNs in FlexVol volumes that are also thin provisioned with a capacity that is two times the size of the LUN. By deploying the LUN in this manner, the FlexVol volume acts merely as a quota. The storage consumed by the LUN will be reported in FlexVol and its containing aggregate. For further details, refer to [“TR-3749: NetApp VMware vSphere Best Practices.”](#)
- **Deduplication.** NetApp deduplication technology eliminates duplicate data at the storage level, thus making better use of storage. It provides space saving by removing redundant copies of blocks within a

volume. This process is transparent to applications and can be enabled or disabled on the fly. Deduplication technology enables multiple VMs to share the same physical blocks in a NetApp V-Series/FAS system in the same manner that VMs share system memory. It can also be introduced seamlessly into a virtual infrastructure without having to make any changes to the virtual environment's administration, processes, or tasks. Deduplication runs on the NetApp storage system at scheduled intervals and does not consume any CPU cycles on the hypervisor. Figure 3 illustrates use of NetApp deduplication technology for a VMware environment. For more information about NetApp deduplication technology, see "[TR-3505: NetApp Deduplication for FAS Deployment and Implementation Guide.](#)"

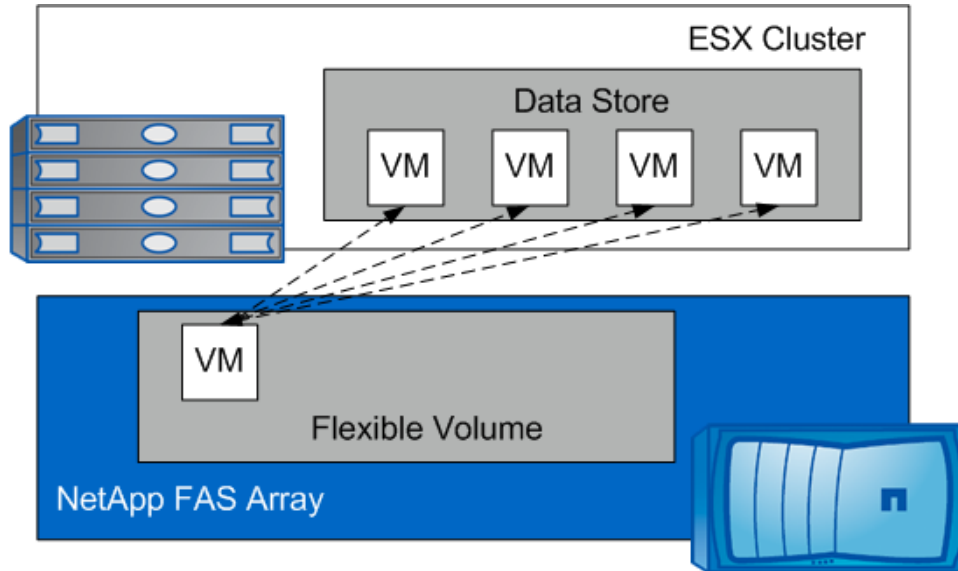


Figure 3) NetApp deduplication technology for VMware environments.

- **FlexVol volume, FlexClone technology, and Snapshot copy.** Virtualization abstracts physical resources and separates the manipulation and use of logical resources from their underlying devices and physical implementation.

FlexVol volume. NetApp has used this same technique to virtualize file volumes by adding a level of indirection, called a FlexVol volume, between client-visible volumes and the underlying physical storage. The FlexVol volumes are managed independent of lower storage layers. Multiple volumes can be dynamically created, deleted, resized, and reconfigured within the same physical storage container.

FlexClone technology. A FlexClone volume is a writable, point-in-time image of a FlexVol volume or another FlexClone volume and is based on NetApp Snapshot technology. FlexClone volumes add a new level of agility and efficiency to storage operations. They take only a few seconds to create and are created without interrupting access to the parent FlexVol volume. FlexClone volumes use space very efficiently, leveraging the Data ONTAP architecture to store only data that changes between a parent and its clone. The use of FlexClone technology in a virtual environment offers significant savings in cost, space, and energy. In addition to all these benefits, FlexClone volumes have the same high performance as FlexVol volumes.

Figure 4 illustrates how NetApp FlexVol volumes and FlexClone technology can virtualize storage and offer substantial return on investments for customers.

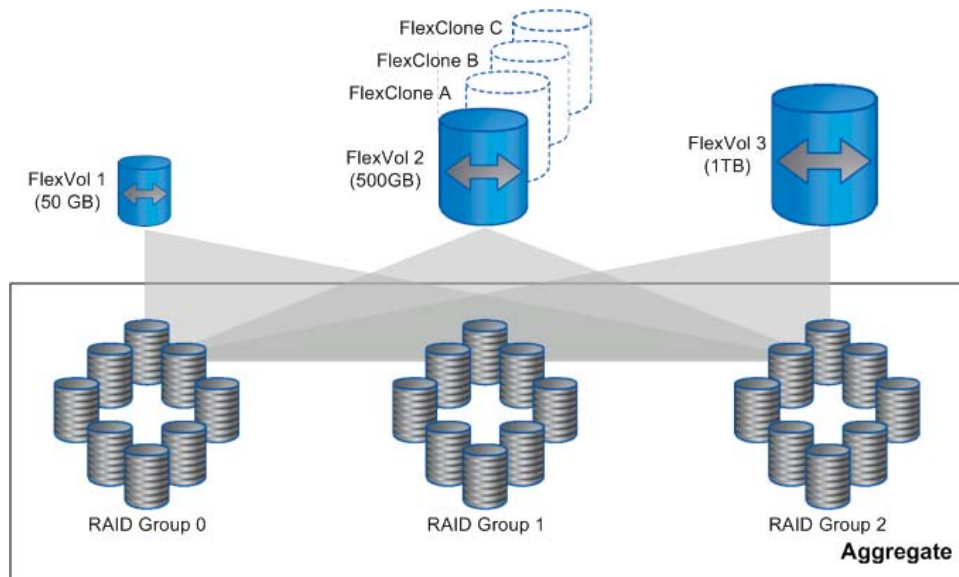


Figure 4) NetApp storage virtualization with FlexVol and FlexClone.

Snapshot copy. A Snapshot copy is a frozen, read-only image of a traditional volume, a FlexVol volume, or an aggregate that captures the state of the file system at a specific time. Snapshot copies provide a first line of defense to back up and restore data. NetApp Snapshot technology can be integrated easily into virtual environments, where it provides crash-consistent versions of VMs for the purpose of full VM recovery, full VM cloning, or site replication and disaster recovery. This is the only snapshot technology currently available that does not have a negative impact on system performance.

- **Flash Cache.** NetApp Flash Cache is intelligent read cache that combines software and hardware within NetApp storage controllers to increase system performance without increasing the disk drive count. It optimizes the performance of random read-intensive workloads such as OLTP databases, virtual infrastructure, file services, and messaging without using more high-performance disk drives. These intelligent read cache cards speed access to data by reducing latency by a factor of 10 or more compared to disk drives. Faster response times can translate into higher throughput for random I/O workloads. Flash Cache cards are implemented with software features in Data ONTAP.

NetApp Flash Cache gives customers performance that is comparable to that of solid-state disks (SSDs) without the complexity of another storage tier. Customers do not have to move data from tier to tier to optimize performance and cost. Active data automatically flows into Flash Cache because every volume and LUN behind the storage controller is subject to caching. The Flash Cache takes full advantage of Data ONTAP features such as deduplication, FlexClone, and FlexShare to deliver storage efficiency and performance. For deduplicated data, only one copy of data is kept in the read cache, block for block, as it is kept on disk.

The Flash Cache can reduce costs for storage, power, and rack space. These cards can be used in combination with Serial Advanced Technology Attachment (SATA), serial-attached SCSI (SAS), and FC drives for many workloads to increase storage capacity and reduce costs while maintaining a high level of performance. For further detail, refer to "[TR-3832: Flash Cache and PAM Best Practices.](#)"

- **vCenter plug-ins—VSC and RCU.** NetApp has introduced a new model of storage provisioning and management that allows the virtual infrastructure administrative team to deploy and manipulate datastores and VMs directly from raw storage pools provided by the storage administrators. This functionality is provided by two plug-ins for vCenter Server—the VSC and RCU. These plug-ins work in conjunction with each other and should be installed together.

VSC. VCS enables optimal availability and performance with ESX/ESXi hosts. Following are the core abilities of the VSC:

- Identify and configure optimal I/O settings for FC, FCoE, iSCSI, and NFS in ESX/ESXi hosts
- Identify and configure optimal path selection policies for existing FC, FCoE, and iSCSI datastores

- Monitor and report storage use at levels from the datastores to the physical disks
- Provide a central interface to collect data from storage controller, network switches, and ESX/ESXi hosts in order to aid in the resolution of I/O-related case issues

The VSC allows the automated configuration of storage-related settings for all ESX/ESXi 4.x hosts to connect to NetApp storage controllers. For more information, refer to "[TR-3749: NetApp VMware vSphere Storage Best Practices.](#)"

RCU. RCU provides a means to provision and manage optimally NetApp storage network (SAN) and network-attached storage (NAS) datastores along with providing a means to provision-zero cost VMs directly from within vCenter. Following are some of the core features of the RCU:

- Provision FC, FCoE, iSCSI, and NFS datastores
 - Automated assignment of multipathing policies to datastores with the VMware pluggable storage architecture for LUNs and Asymmetric Logical Unit Access (ALUA)-enabled LUNs, and distribution of NFS path connections based on a path round robin policy
 - Automated storage access control security implemented when datastores are provisioned; access control is in the form of LUN masking and NFS exports
 - Dynamically resize FC, FCoE, iSCSI, and NFS datastores on the fly
 - Provide a central interface to collect data from storage controller, network switches, and ESX/ESXi hosts in order to aid in the resolution of I/O-related case issues
- **SMVI.** SMVI works with VMware vCenter to automate and simplify management of backup and restore operations. The easy-to-manage tool allows VMware administrators to create application-consistent backups for their VMs. In addition, they can recover a datastore, VM, VMDK, or an individual file within a VM guest instantly. For more information, refer to "[TR-3737: SnapManager for Virtual Infrastructure Best Practices.](#)"

3 SOLUTION ARCHITECTURE

This solution showcases a deployment architecture virtualizing an IBM DB2 environment with VMware vSphere 4, using Intel virtualization-optimized processors and NetApp unified storage. The results from various workload tests demonstrate that the performance and scalability of the IBM DB2 virtual solution is suitable for production environments and the solution conforms to all best practice recommendations from IBM.

Each server used in the configuration was equipped with a QLogic QLE2462 dual port 4GB host bus adapter (HBA) with multipathing enabled for high availability and load balancing. We used FC connectivity between the database server and the NetApp storage. NetApp storage controllers were configured in active-active cluster mode to further enhance database availability.

For the tests described in this white paper, we used two NetApp FAS 3070 controllers, each attached to four disk shelves loaded with fourteen 15K RPM, 144GB FC disk drives. The controllers were configured in cluster mode. For the database server, we used an IBM x3850 M2 with 24GB RAM and Intel Xeon X5570 processor. Figure 5 illustrates the server, storage, and software components used in the solution test environment.

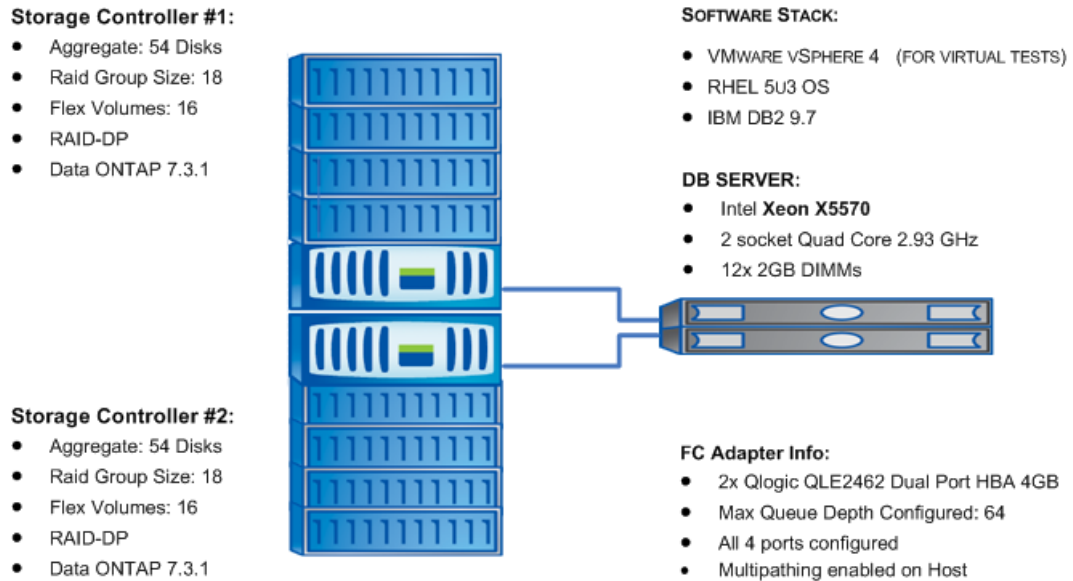


Figure 5) Storage, server, and software components used for the test environment.

Our main objective for this white paper is to study the benefits and impact of virtualizing the database using the DB2 virtualization feature and virtualization technologies from NetApp, VMware, and Intel. We performed the following tests to complete this study:

- Storage and server layout for bare metal DSS workload testing
- Storage and server layout for DSS workload testing in a virtualized environment
- Server configuration and storage layout for OLTP workload testing in a bare metal environment
- Server configuration and storage layout for OLTP workload testing in a virtual environment
- Server configuration and storage layout for vCPU overallocation testing

The following sections describe the test environment configuration for each test.

STORAGE AND SERVER LAYOUT FOR BARE METAL DSS WORKLOAD TESTING

The following list details the storage and server layout for the bare metal DSS workload testing:

- IBM DB2 9.7 server configured with 8 separate partitions, which is the maximum
- Enabled automatic storage and index compression
- 32 LUNs on the storage controllers
- 16 LUNs for each storage controller
- Sixteen 100GB LUNs for data
- Sixteen 40GB LUNs for logs

Figure 6 illustrates the server configuration and storage layout for DSS workload testing in a bare metal environment.

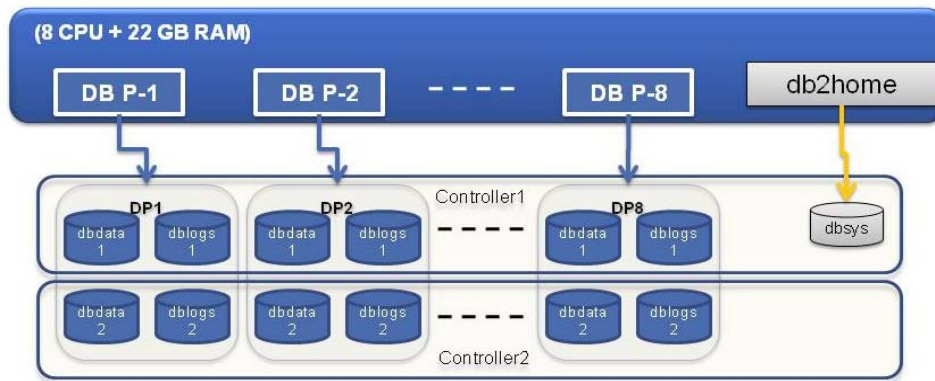


Figure 6) Server configuration and storage layout for DSS workload testing in a bare metal environment (graphic supplied by IBM).

STORAGE AND SERVER LAYOUT FOR DSS WORKLOAD TESTING IN A VIRTUALIZED ENVIRONMENT

The following list details the storage and server layout for the DSS workload testing in a virtualized environment:

- IBM DB2 9.7 server configured with 8 vCPUs and 22GB of RAM
- 32 RDM LUNs on the storage controllers
- 16 RDM LUNs for each storage controller connected to database partitions created in the VMware environment
- Sixteen 100GB LUNs for data
- Sixteen 40GB LUNs for logs
- Database home directory also on a NetApp storage volume

Figure 7 illustrates the server storage configuration for DSS workload testing in a virtualized environment.

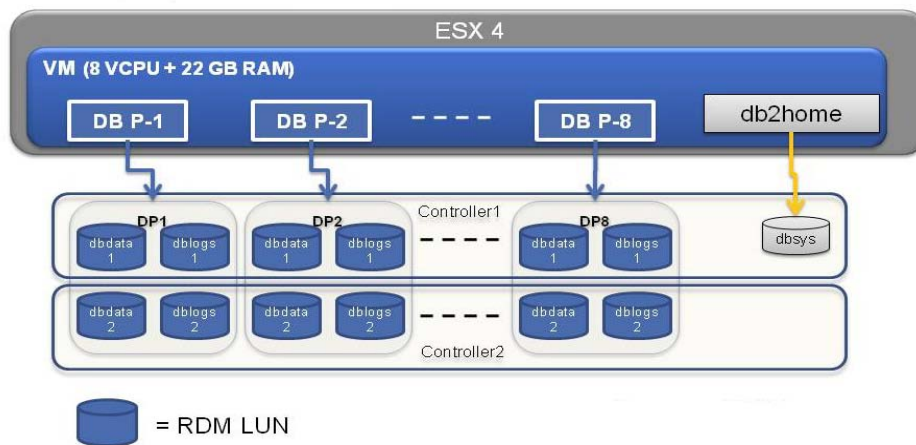


Figure 7) Server configuration and storage layout for DSS workload testing in a VMware virtual infrastructure environment (graphic supplied by IBM).

SERVER CONFIGURATION AND STORAGE LAYOUT FOR OLTP WORKLOAD TESTING IN A BARE METAL ENVIRONMENT

The test environment for OLTP workload testing in a virtual environment was similar to a bare metal environment.

The following list details the server configuration and storage layout for the OLTP workload testing in a bare metal environment:

- 1 DB2 instance and a database in a VM
- Data on NetApp storage
- SSD device for logs
- 8 vCPUs
- Server equipped with 22GB of RAM

Figure 8 illustrates the server configuration and storage layout for OLTP workload testing in a bare metal environment.

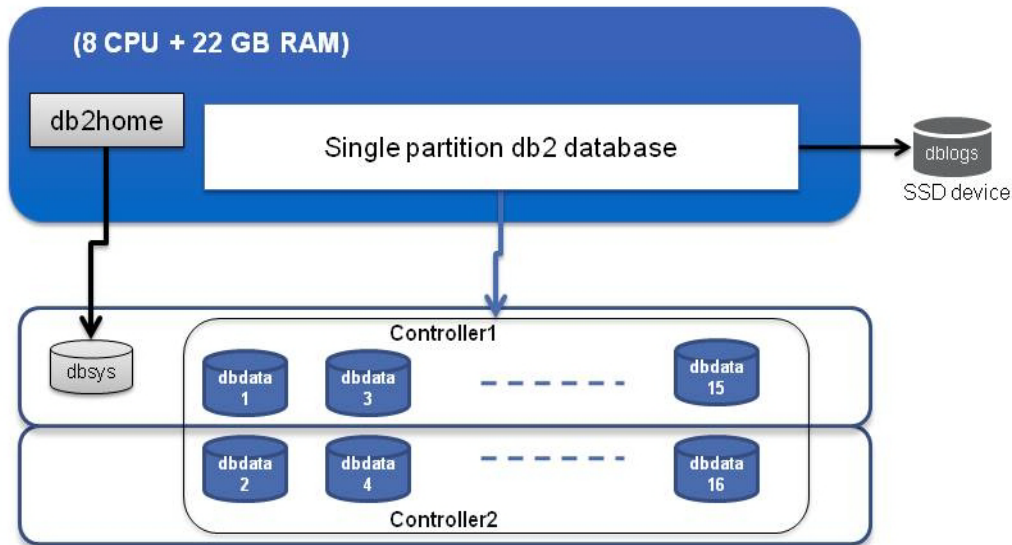


Figure 8) Server configuration and storage layout for OLTP workload testing in a bare metal environment (graphic supplied by IBM).

SERVER CONFIGURATION AND STORAGE LAYOUT FOR OLTP WORKLOAD TESTING IN A VIRTUAL ENVIRONMENT

The test environment for OLTP workload testing in a virtual environment was similar to a bare metal environment.

The following list details the server configuration and storage layout for the OLTP workload testing in a virtual environment:

- 1 DB2 instance and a database in a VM
- Data on NetApp storage
- SSD device for storing logs
- 8 vCPUs
- Server equipped with 22GB of RAM

Figure 9 illustrates the server configuration and storage layout for OLTP workload testing.

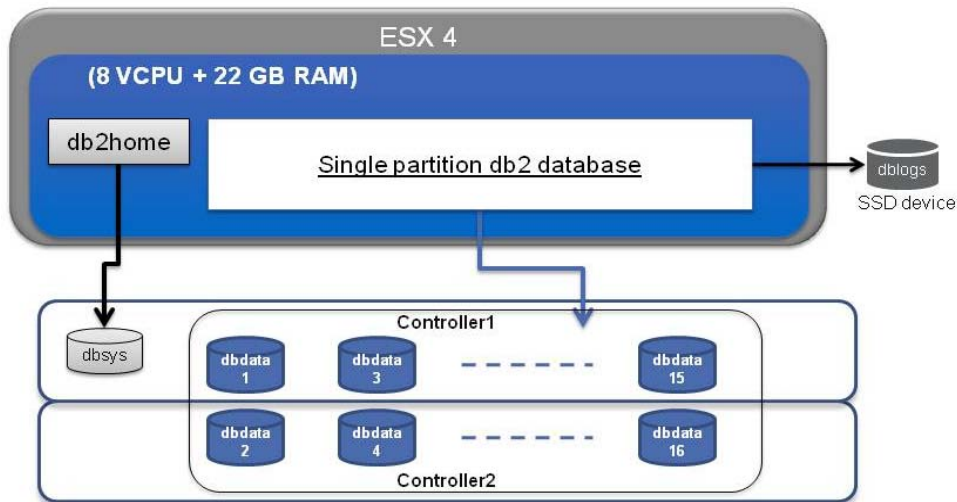


Figure 9) Server configuration and storage layout for OLTP workload testing in a virtual environment (graphic supplied by IBM).

SERVER CONFIGURATION AND STORAGE LAYOUT FOR VCPU OVER-ALLOCATION TESTING

The following list details the server configuration and storage layout for vCPU overallocation testing:

- 4 VMs
- 6 vCPU and 5GB RAM for each VM
- Dedicated data and log containers for each VM
- SSD devices for storing logs

Figure 10 illustrates the server configuration and storage layout for vCPU overallocation testing.

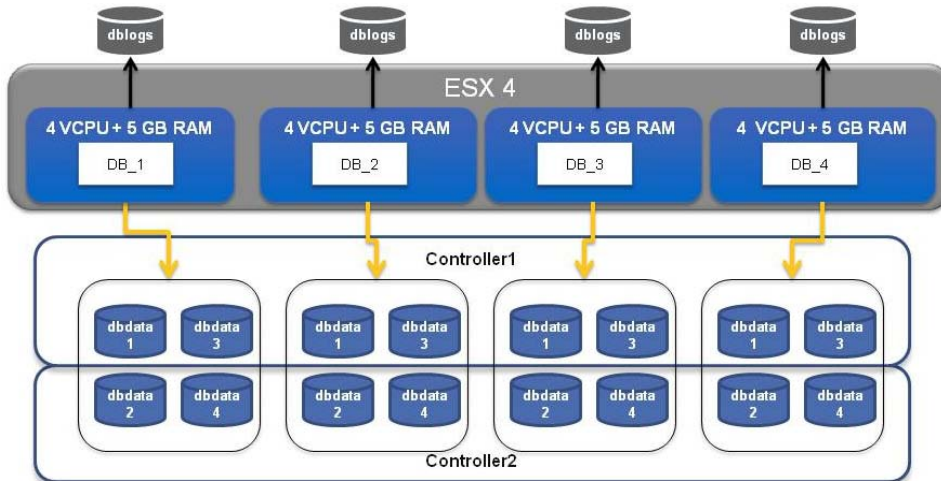


Figure 10) Server configuration and storage layout for vCPU overallocation testing (graphic supplied by IBM).

4 TESTING AND PERFORMANCE RESULTS

Technology improvements in all layers, from the microchip architecture of processors to the virtualization, database, and storage software, have contributed to closing the gap in performance between applications running in a traditional bare metal environment and applications running in a virtual environment. In this section, we describe the test plan used to observe whether these improvements are sufficient for database

workloads and to judge their suitability in virtual environments. To compare against bare metal performance, both DSS and OLTP workloads are evaluated at nontrivial scale factors of 100GB-sized raw databases with significant I/O requirements. We conducted a third test to demonstrate the DB2-stated best practice of not allocating more vCPUs than physical processors. Refer to Section 8, “Appendixes: Configuration Parameter Settings,” for database configuration parameters.

4.1 DECISION SUPPORT WORKLOAD PERFORMANCE

TEST OBJECTIVE

Our objective for this test was to demonstrate that virtualizing the DB2 DSS environment has relatively low to no overhead and it can meet the high I/O requirements. To measure the impact of virtualization, we compared virtual system performance against a bare metal baseline of equivalent configuration.

In this test, we used an eight-processor DB2 configuration with a 100GB (raw) database within a single OS image. We compared performance for a bare metal physical server versus vSphere 4 (single VM) to measure the virtualization overhead. Row and index compression were used on all tables.

TEST CONFIGURATION

The DB2 data server was configured to use eight logical partitions within a single OS image. The database size was 100GB (raw). We used row and index compression on all tables.

The physical system had the HT feature disabled for this specific test so that we could properly compare eight online processors in a bare metal Linux configuration with a maximum of eight vCPUs allowed in a single vSphere 4 VM. Database configuration was identical for both bare metal and virtual environments. Section 8, “Appendixes: Configuration Parameter Settings,” contains the configuration parameters used for this test.

WORKLOAD DESCRIPTION

This test comprises two components:

- A serial component that executed 22 queries back to back and measured the time of each query
- A throughput component that executed five threads running a set of seven ad hoc queries in random order and measured the time required for all threads to complete

SERIAL TEST RESULTS AND CONCLUSIONS

For the serial component of the test, VM performance was 97% of bare metal performance. This result was attributed to less I/O throughput being required because of DB2 compression and the serial nature of this test, and that the spare CPU capacity available for decompression purposes did not take away from other DB2 server computational needs. The graph in Figure 11 compares VM and bare metal serial performance.

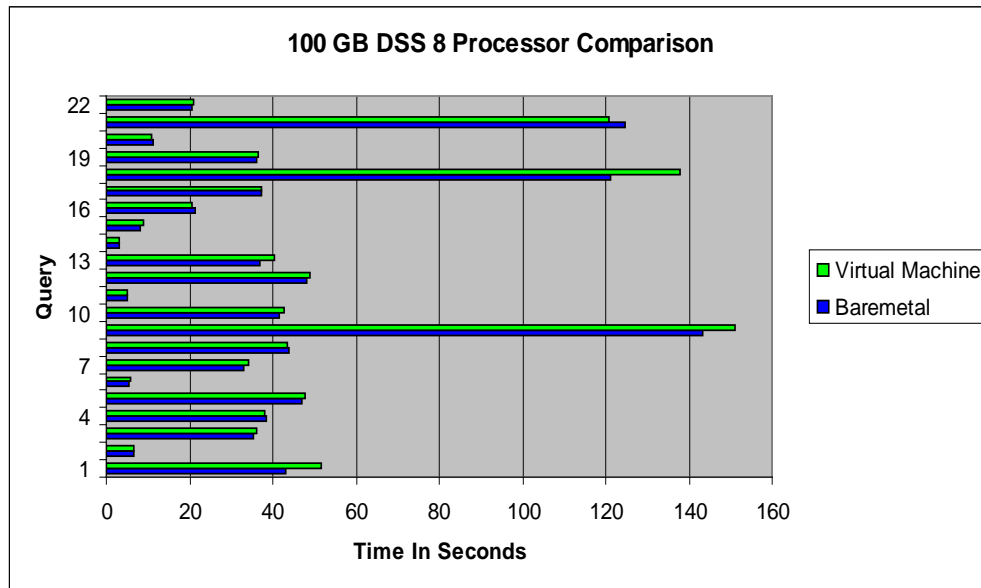


Figure 11) Serial test performance comparison (graphic supplied by IBM).

THROUGHPUT TEST RESULTS AND CONCLUSIONS

For the throughput component of the test, VM performance was significantly close to bare metal performance. This test exerted greater demands for both CPU and I/O because there were five concurrent threads running complex queries at the same time. The overall disk I/O rate was observed up to 1.25GB/s and the CPU was fully used. The substantial I/O demands that are traditionally expensive to virtualize likely brought down the result compared to that of the serial test. Overall, the VM fared well against the bare metal environment for this workload profile. The graph in Figure 12 compares the throughput performance of the VM and bare metal environment.

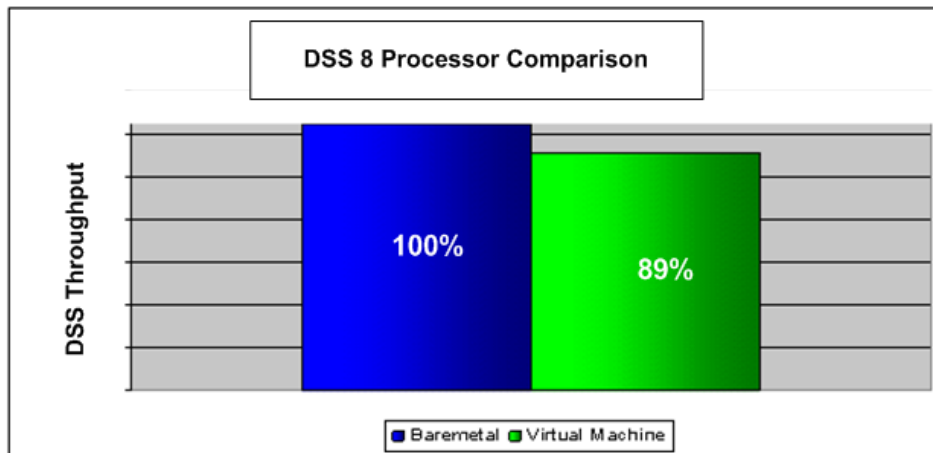


Figure 12) Throughput test performance comparison (graphic supplied by IBM).

In the graph, the throughput metric is calculated as the number of concurrent threads multiplied by the number of queries executed each thread, multiplied by the scale factor of 100GB, divided by the average time it took for each thread to complete. A higher-value result is better. Figure 6 illustrates the server and storage layout used in the bare metal environment and Figure 7 illustrates the server and storage layout used in the virtual environment.

4.2 OLTP WORKLOAD PERFORMANCE

TEST OBJECTIVE

Our objective for this test was to measure the impact of virtualization for an OLTP workload environment with high random I/O demands. To do this, we compared the virtual performance against an equivalently configured bare metal baseline.

TEST CONFIGURATION

The OLTP workload consisted of a 100GB (raw) database that also used DB2 row compression on its tables and indexes within a single OS image.

The physical system had HT disabled to compare properly eight online processors in a bare metal Linux environment against the maximum eight vCPUs allowed in a single vSphere 8 VM. Log I/O was off-loaded on an SSD device to keep data I/O balanced on the storage systems. The database configuration was identical for both bare metal and virtual environments. See Section 8, "Appendixes: Configuration Parameter Settings," for the configuration parameters used in this test. Figure 8 illustrates the server and storage layout used in the bare metal environment. Figure 9 illustrates the server and storage layout used in the virtual environment.

WORKLOAD DESCRIPTION

The workload used 120 concurrent clients issuing transaction requests to the database at a ratio of 30% reads to 70% writes. The database was configured to use only 3GB of the available 22GB for buffer pool use, enabling the storage system to get sufficient exercise. Linux OS large pages were enabled for the buffer pool to reduce the effects of TLB cache misses.

TEST RESULTS AND CONCLUSIONS

The graph in Figure 13 compares the OLTP performance of VM and bare metal.

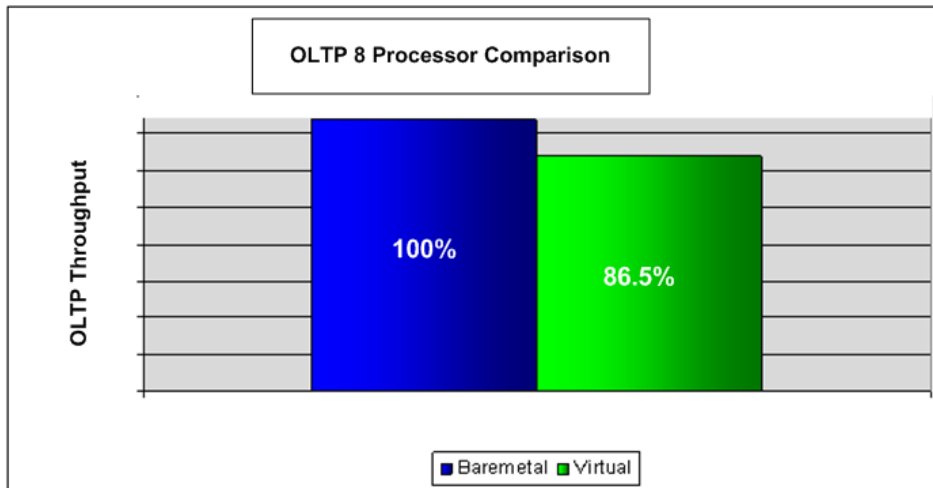


Figure 13) OLTP workload test performance comparison (graphic supplied by IBM).

The graph shows that, for this OLTP workload, the VM delivered 86.5% of the performance of the bare metal environment. On bare metal, the disk reached 31000 I/Os per second (IOPS), compared to 25500 IOPS on the VM.

4.3 VCPU OVERALLOCATION TEST

TEST OBJECTIVE

The overallocation test was conducted to measure the impact of having more vCPUs allocated across all VMs than the number of physical CPUs available. This is not the same as overcommitting CPU resources because the workload was sized to use only the capacity of the physical machine. The effect of losing memory locality was also monitored.

TEST CONFIGURATION

HT was enabled to produce 16 execution threads to which the vCPUs could be mapped.

To conduct this test, we created four separate, identically configured VMs. Each VM had its data storage on the SAN with its log on a separate SSD device. The baseline test used four vCPUs in each VM, while the comparison test used six vCPUs in each VM, but without increasing the overall load. The database configuration was identical for all VMs. See Section 8, "Appendixes: Configuration Parameter Settings," for configuration parameters used for this test.

WORKLOAD DESCRIPTION

The OLTP workload was a scaled-down version of the workload described in Section 4.2. The databases were each 1.6GB in size and I/O per VM was measured at a modest 3500 IOPS.

TEST RESULTS AND CONCLUSIONS

Overall, the overallocated case showed 6.5% degradation at peak load as illustrated by the graph in Figure 14.

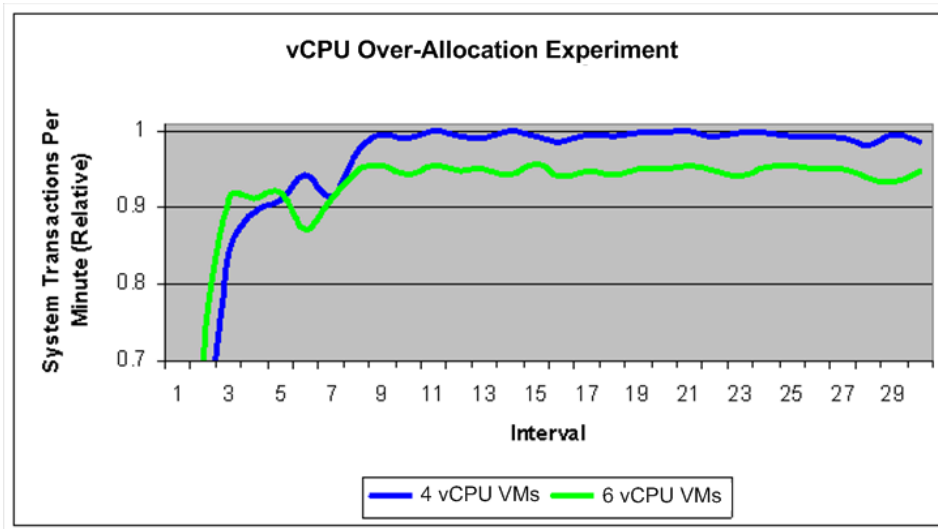


Figure 14) vCPU overallocation performance (graphic supplied by IBM).

Overall, a 50% increase in vCPUs over physical execution threads that yielded 6.5% degradation in peak performance is approximately equal to a 1.3% performance penalty for every 10% overallocation of vCPUs. Part of this penalty might be attributed to the additional overhead of scheduling vCPU resources and the associated cache efficiency loss. Most of this penalty, however, is the result of not always using local memory. On four-core processors, vSphere schedules 4-vCPU VMs on a single processor and uses memory local to the processor. For larger VMs, both CPU and memory are spread (interleaved) across both processors. On a two-socket system, this means that approximately half the memory accesses are remote. Remote access is slower than local access.

5 BEST PRACTICES

This section provides a summary of best practices that can be applied to virtual environments. In particular, this summary concerns best practices for deploying DB2 databases—from the configuration of processor, physical server, network and storage components, and systems to the configuration of vSphere 4 and IBM DB2 9.7.

5.1 INTEL BEST PRACTICES

To obtain the best performance and energy efficiency under OLTP, DSS, and a broad mix of other activities, you can selectively enable several features available in Intel Xeon 5500 or Intel Xeon 5600 series processor-based systems. By default, many of the features are set to a default value that is good for most cases. It is a good practice to examine the system BIOS settings to verify key settings as described in the following list:

- **Hardware virtualization.** This setting should be enabled.
- **Extended page tables (EPTs).** EPTs provide hardware support for virtualizing each CPU's memory management unit (MMU). VMware's vSphere 4 supports this capability. For processors that do not implement hardware support for MMU virtualization, VMware also provides software-based MMU virtualization. Thus, both hardware- and software-based MMU virtualization are available as options when using VMware with current-generation Intel processors.

When software MMU virtualization is employed, the hypervisor maintains shadow page tables that let the processor map guest virtual addresses directly to host physical memory addresses and keeps the shadow page tables consistent with the guest page tables. The maintenance and synchronization of these additional sets of page tables requires additional work and memory use by the hypervisor. Conversely, the use of EPTs eliminates the need for such overhead by having the hardware complement the translation provided by guest page tables with additional translation from each guest physical address to the final machine physical address. This additional translation increases the length of each page table lookup operation, but the processor also implements several hardware caching structures for page tables and previously translated addresses that make such table lookups infrequent.

As a result, EPTs can hurt performance by 1%–2% for very few workloads, but for the broad spectrum of typical workloads, EPTs improve performance significantly and further improve multi-VM scaling. Therefore, EPTs should be enabled and use of the hardware-assisted MMU virtualization, which is VMware's default policy, should remain operative unless the software alternative yields decidedly higher performance.
- **Hyperthreading technology (HT).** The HT capability in the processor should be enabled. This capability improves concurrent execution, reduces the impact of cache misses, and makes processor scheduling easier for the hypervisor. In rare server workload situations, HT can reduce performance by a few percent, but database-intensive work generally does not fall into that category. All hypervisors, including vSphere, reveal through administrative interfaces how many physical CPUs are available, so determining whether HT is enabled or disabled on a platform is easy.
- **Nonuniform memory access (NUMA).** This capability effectively determines the granularity at which the physical memory in a platform is interleaved across the processor sockets. When NUMA is enabled, a hypervisor can improve the likelihood of socket-proximal memory accesses by biasing memory allocation according to the physical socket from which a memory request is issued. The vSphere 4 hypervisor implements NUMA-aware scheduling of virtual CPUs, so it is generally recommended that NUMA be enabled. One situation in which it is not possible to recommend NUMA with certitude is when the physical machine is used to host just one guest. This is because the hypervisor may find it more beneficial to override the proximity consideration and spread out processing in order to balance the use of resources such as socket caches and memory channels across sockets.
- **Intel Turbo Boost Technology.** The Intel Turbo Boost capability dynamically allows a subset of active cores in each processor socket to operate at higher frequency when the remaining cores are underused. When enabled on the Intel 5500 or 5600 series processors, this capability has the potential to improve performance by up to 10%, particularly if a workload is unlikely to saturate the available number of physical CPUs. For the DB2 solution described in this white paper, the majority of database activities were expected to be highly concurrent, so the Intel Turbo Boost capability was of limited interest in our study.
- **Prefetch settings.** Typically, the system BIOS allows two prefetcher settings to be turned on or off. One is called the Hardware prefetch or Streaming prefetch, and the other is called the Adjacent sector

prefetch. Intel recommends using the default settings for these options because they are selected on the basis of performance measurements taken across many workloads.

5.2 NETAPP BEST PRACTICES

The following best practices are recommended by NetApp for solution configurations described in this white paper:

- **Large aggregate.** A large aggregate was created for each storage controller (54 disks). The RAID group size was set to 18, an optimal setting for the OLTP workload.
- **Separate data and logs.** Separate flexible volumes were created for data and transaction logs. The flexible volumes were automatically striped across all the disks in the aggregate.
- **RAID-DP.** RAID-DP offers high reliability compared to other RAID architectures. We used RAID-DP for the tests.
- **Partition and file system alignment.** NetApp LUNs should always be partitioned with a single primary partition. The partition serves two purposes. It functions as a label for the LUN, which helps the operating system identify the contents of the LUN. The partition is also used to align the host file system with the LUN. Aligning the host file system is necessary to achieve maximum performance of read and write I/O operations. The NetApp technical report "[TR-3747: Best Practices for File System Alignment in Virtual Environments](#)" contains additional detail on file system alignment.
- **One-to-one relationship between VMware datastores and FlexVol volumes.** NetApp recommends a one-to-one relationship between VMware datastores and flexible volumes when deploying virtual servers.
- For database environments, NetApp recommends the following options be modified for each flexible volume:
 - **Disable automatic NetApp Snapshot copies.** By default, when a volume is created, a default Snapshot schedule is set for it. A backup schedule for this particular environment is not required. However, if a backup schedule were required, you would have to back up the database based on a user-defined schedule. Therefore, automatic Snapshot scheduling was disabled on each volume for these tests.
 - **Update the access time of all files.** For volumes that are used by databases, the database management system manages the correct file-access time for the inodes. Therefore, the access time to update by storage is not required and this option was disabled.
 - **Set nvfail option on.** If this option is set to On, the NetApp storage system performs additional status checking at boot time to verify that the storage system's nonvolatile RAM (NVRAM) is in a valid state. This option is useful for databases. If any problems with NVRAM are found, the database instances are shut down and an error message is sent to the console to alert the database administrators.
 - **Set the Snapshot reserve.** When a new volume is created, by default Data ONTAP reserves 20% of the space for the Snapshot copies, which then cannot be used for the data. To better use the storage space, the Snapshot reserve was set to 0 on each volume.
 - **Max I/O per session.** This parameter controls the maximum outstanding commands per session. It was changed from the default value of 128 to 256 for database workloads.
- **Native Multipathing Plugin (NMP) and Path Selection Policy (PSP) settings.** Make sure that ESX multipathing policies are set to NMP and that PSP is set to round robin with ALUA enabled.
- **Igroup for LUN masking.** When provisioning LUNs for access using FC or iSCSI, the LUNs must be masked so that the appropriate hosts can connect only to them. With a NetApp FAS system, LUN masking is handled by the creation of initiator groups. NetApp recommends creating an igroup for each VMware cluster. NetApp also recommends including in the name of the igroup the name of the cluster and the protocol type (for example, DC1_FC and DC1_iSCSI).
- **VMkernel swap file on separate FlexVol volume.** NetApp recommends that the VMkernel swap file for every VM be relocated from the VM home directory to a datastore on a separate NetApp volume that is dedicated to storing VMkernel swap files. NetApp suggests creating either a large thin-provisioned LUN or a FlexVol volume with the Auto Grow feature enabled. Thin-provisioned LUNs and Auto Grow FlexVol volumes provide a large management benefit when storing swap files.
- **Maximum queue depth value.** The default maximum queue depth value for QLogic HBA is 32. NetApp recommends changing the value to 64 for better performance.

5.3 VMWARE BEST PRACTICES

At a high level, you can approach design and deployment of virtual infrastructure the same as deployment on physical hardware. A key point to remember, however, is that VMware virtual infrastructure is designed to allow sharing of physical resources (CPU, disk, memory, and networking) among various workloads. DB2 databases, on the other hand, can be extremely computationally intensive, especially regarding use of network, disk, and memory resources. Therefore, when deploying VMware virtual infrastructure for DB2, be sure that service levels can be maintained when DB2 servers share compute resources with other VMs on the same server. Design DB2 virtualized systems to avoid hardware resource bottlenecks.

The following list contains additional best practices for running DB2 on VMware:

- Follow best practices in place for virtual infrastructure systems designed using VMware vSphere. Refer to [“Performance Best Practices for VMware vSphere 4.0”](#) for more information.
- Use physical sizing guidelines and best practices in considering overheads for virtualized target platforms. When configuring DB2 database VMs, the total CPU resources needed by the VMs running on the system should not exceed the CPU capacity of the host. It is good practice to undercommit CPU resources on the host because the performance of your virtual database may degrade if the host CPU capacity is overloaded.

However, when using VMware vSphere advanced workload management features such as VMotion and VMware Distributed Resource Scheduler (DRS), the database is freed from the resource limitations of a single host. VMware VMotion enables DBAs to move running DB2 VMs from one physical ESX host to another and to balance available resources with little impact on end users. VMware DRS dynamically allocates and balances computing resources by continually monitoring the use of resource pools associated with VMs in a VMware cluster.

In general, 80% usage is a reasonable ceiling in production environments. Use 90% as an alert to the VMware administrator that the CPUs are approaching an overloaded condition and should be addressed. Ultimately, however, decisions regarding the desired load percentage should be made based on the criticality of the DB2 database being virtualized.

Even if some vCPUs are not currently being used, configuring a virtual DB2 database with the excess vCPUs can impose additional resource load on vSphere due to the unused vCPUs still consuming timer interrupts. VMware vSphere attempts to coschedule the multiple vCPUs of a VM, trying to run vCPUs in parallel as much as possible. Having unused vCPUs imposes scheduling constraints on the vCPU being used and can degrade overall performance.

- If possible, keep the total number of vCPUs of all VMs less than the number of physical execution threads in the system. If you need to overallocate, use priority methods such as CPU shares in vSphere to give certain VMs higher CPU priority than others. This eases scheduling impact.
- Set memory reservations equal to the size of the DB2 shared memory. When consolidating DB2 database instances, vSphere presents the opportunity for sharing memory across VMs that may be running the same operating systems, applications, or components. In this case, vSphere uses a proprietary transparent page sharing technique to reclaim memory, which allows databases to run with less memory than physical memory available. Transparent page sharing also allows DBAs to overcommit memory, without any performance degradation.
- In production environments, careful consideration should be taken when overcommitting memory and should only be introduced after collecting data to determine the amount of overcommitment possible. To determine the effectiveness of memory sharing and the degree of acceptable overcommitment for a given database, run the workload and use the `resxtop` or `esxstop` tools to observe the actual savings. While VMware recommends setting memory reservations in production environments equal to the size of the DB2 shared memory, you can introduce more aggressive overcommitment in nonproduction environments such as development, test, or QA. In these environments, a DBA can introduce memory overcommitment to take advantage of VMware memory reclamation features and techniques. Even in these environments, the type and number of databases that can be deployed using overcommitment largely depend on their usage characteristics and their criticality to the business.
- Install VMware tools on your guest OS.

VMware tools within the guest provide many benefits to your DB2 execution environment such as:

- Better time synch between the VM and the vSphere 4 host

- Better memory management capability by allowing the VMware balloon driver to increase guest memory pressure in times of physical memory pressure on the system and have the OS and DB2 react to it in a natural way
- Ability to use paravirt I/O drivers (vmxnet 2 and 3 for network) as well as PVSCSI for disk
- Other intangibles, such as enhanced drivers for mouse, video, networking, and the ability to write system scripts that can trigger on VM events to ease system maintenance, including suspend/resume/power on-off

Allow vSphere to choose the best VM monitor based on the CPU and guest operating system combination. Make sure that the VM setting has Automatic selected for the CPU/MMU virtualization option.

- VM sizing considerations in NUMA environments

VMware vSphere 4 can take advantage of the faster local memory access of NUMA systems by scheduling VM resources on a single NUMA node. However, this is only possible if those resources (virtual CPU and RAM) can fit within a single NUMA node.

NUMA information is not passed to vSphere 4 VMs, so there is no advantage in trying to make use of DB2's NUMA-aware capabilities.

- Virtual CPU (vCPU) allocation

When planning your vSphere 4 environment for DB2 use, NetApp recommends that the total number of live vCPUs in the total system be less than or equal to the number of logical CPUs, as seen by vSphere 4 in its host view. (Also refer to the recommendations for vCPU allocation described earlier in this section.) As shown by test results described in Section 4, performance may be degraded if you add more vCPUs even if you do not increase the load on the physical system. This is true especially if you do not take into account NUMA considerations.

If you need to allocate more vCPUs than logical CPUs, consider attaching a priority to certain VMs so that higher-priority VMs will not suffer in quality of service. You can do this through the VMware client from the Edit Virtual Machine Settings tab.

You can use various VMware performance measurement and analysis tools available with vSphere to validate system design and performance under different workloads and also to isolate and address specific performance issues. For additional performance troubleshooting information, see [“Performance Troubleshooting for VMware vSphere 4 and ESX 4.0.”](#)

For example, you can use vSphere monitoring and management tools to determine if CPU resources are overcommitted. You can also determine how to set CPU and memory reservations and check VM disk and network setup and performance.

5.4 DB2 BEST PRACTICES

NetApp recommends the following best practices when deploying your DB2 environment with VMware vSphere 4 virtualization:

- Put VM swap on dedicated disk

VMs require swap space at the vSphere 4 host level in case they need to be paged out in a stressed physical memory environment; however, performance of the affected machine will suffer. It is possible that you can avoid performance issues on other executing VMs by not sharing these disks with the ones used for your DB2 data and logs.

- Use DB2 automatic parameters whenever possible

By default, most initial DB2 9.7 parameters for both the instance configuration and the database configuration are set to AUTOMATIC and the self-tuning memory manager is enabled. These settings should be maintained in vSphere 4 environments because VMs can be reconfigured easily with different CPU and memory resources. This also eases DB2 VM cloning because the parameters do not require retuning once DB2 wakes up in a new VM environment.

If you plan to deploy a DB2 VM on another physical host (that is, using a CPU model different from the one it was created on or templated on), it is a good practice to reset the DB2 instance CPUSPEED parameter to get a more accurate picture of the CPU model on the new host. This can be done with the following command:

```
db2 update dbm cfg using cpuspeed -1
```

- Consider using DB2 deep compression for your tables
In DB2 9.7, tables that are compressed also have their indexes compressed. This leads to more efficient storage and memory use at the cost of additional CPU. In virtual environments, I/O tends to be more expensive to virtualize than CPU, so compression should be considered to take advantage of this trade-off.
- Use of OS large pages
Using OS large pages is particularly beneficial for OLTP workloads, for bare metal and virtual environments. Section 5.3, “VMware Best Practices” provides additional information.
By default, Linux uses 4kB pages, but by setting the `vm.nr_hugepages` using the `sysctl` utility (man `sysctl` for more info), you can set the number of 2MB pages available to your application. To enable large page use in DB2 data servers for the database shared memory, set the DB2 registry variable `DB2_LARGE_PAGE_MEM=DB`.
NOTE: Exercise caution to be sure that the correct amount of large pages has been configured in the kernel. This number can vary depending on how DB2 is tuned and, if not correct, it can lead to issues with database activation, out of memory errors, or OS paging.
- Allocate memory to VMs
For memory-intensive workloads like DB2, NetApp recommends that not all host memory be allocated to VMs. Memory management algorithms such as transparent page sharing are not effective when all, or nearly all, of the VM's memory is allocated to a database cache. Similarly, any ballooning in the VM leads directly to guest swapping and loss of performance. Some memory is needed for VM memory overhead and for the hypervisor itself. The amount needed depends on the size of the VM and other factors, but, generally, 10% of host memory should be conservative.
- Weigh VMFS considerations against RDM considerations
In general, NetApp recommends that you use VMFS for database data and logs because of vSphere 4's enhanced journaling and recovery mechanisms. The performance overhead compared to RDM is negligible.
However, in an enterprise SAN environment, you may want to make use of the SAN's LUN snapshot capabilities. In this case, RDM is necessary. In our testing, RDM was used because snapshots were used. This provided a flexible option for moving from a bare metal environment to a virtual environment without having to reformat constantly the files `system` on the LUNs and register them with vSphere 4.

6 CONCLUSION

Technology improvements in all layers, from the microchip architecture of processors to virtualization, database, and storage software, have contributed to closing the performance gap between an application running in a traditional bare metal environment and an application running in a virtual environment. The test results described in this white paper further prove that virtualization technologies have matured and are suitable for read- and write-intensive database workloads. The integration of virtualization technologies from VMware, NetApp, Intel, and IBM DB2 can deliver a positive return on investment from consolidation of the servers and storage, or new deployments in their data centers.

7 REFERENCES

7.1 INTEL REFERENCE INFORMATION

The following documents provide more information on Intel processors described in this white paper:

- VMware vSphere and Intel Xeon Processor 5500 Series. Delivering the IT Infrastructure of Tomorrow – Today:
http://ipip.intel.com/go/wp-content/themes/ipip/includes/campaigns/cloud/vmware_vsphere.pdf
- Intel Xeon Processor 5500 Series. An Intelligent Approach to IT Challenges:
<http://www.intel.com/assets/PDF/prodbrief/322355.pdf>
- Intel 64 and IA-32 Architectures Optimization Reference Manual:
<http://www.intel.com/assets/pdf/manual/248966.pdf>

7.2 NETAPP REFERENCE INFORMATION

The following documents provide more information on using NetApp storage:

- TR-3749: NetApp and VMware vSphere Storage Best Practices:
<http://media.netapp.com/documents/tr-3749.pdf>
- TR-3747: Best Practices for File System Alignment in Virtual Environments:
<http://media.netapp.com/documents/tr-3747.pdf>
- TR-3737: SnapManager 2.0 for Virtual Infrastructure Best Practices:
<http://media.netapp.com/documents/tr-3737.pdf>
- TR-3505: NetApp Deduplication for FAS and V-Series Deployment and Implementation Guide:
<http://media.netapp.com/documents/tr-3505.pdf>
- TR-3832: Flash Cache and PAM Best Practices Guide:
<http://media.netapp.com/documents/tr-3832.pdf>
- FlexVol: Flexible, Efficient File Volume Virtualization in WAFL:
<http://media.netapp.com/documents/FlexVols.pdf>

7.3 VMWARE REFERENCE INFORMATION

Refer to the following documents for more information on VMware virtual infrastructure configuration and performance optimization.

Performance papers:

- Performance Best Practices for VMware vSphere 4.0:
http://www.vmware.com/pdf/Perf_Best_Practices_vSphere4.0.pdf
- Virtualizing Performance-Critical Database Applications in VMware vSphere:
http://www.vmware.com/pdf/Perf_ESX40_Oracle-eval.pdf
- Comparison of Storage Protocol Performance in VMware vSphere 4:
<http://www.vmware.com/resources/techresources/10034>
- Performance Characterization of VMFS and RDM Using a SAN:
http://www.vmware.com/files/pdf/performance_char_vmfs_rdm.pdf
- Large Page Performance:
http://www.vmware.com/files/pdf/large_pg_performance.pdf
- Recommendations for Aligning VMFS Partitions:
http://www.vmware.com/pdf/esx3_partition_align.pdf
- Dynamic Storage Provisioning: Considerations and Best Practices for Using Virtual Disk Thin Provisioning:
<http://www.vmware.com/files/pdf/VMware-DynamicStorageProv-WP-EN.pdf>

Storage configuration and protocols:

- iSCSI SAN Configuration Guide:
http://www.vmware.com/pdf/vsphere4/r40/vsp_40_iscsi_san_cfg.pdf
- Fibre Channel SAN Configuration Guide:
http://www.vmware.com/pdf/vsphere4/r40/vsp_40_san_cfg.pdf
- NetApp and VMware vSphere Storage Best Practices:
<http://media.netapp.com/documents/tr-3749.pdf>

VMware knowledge base articles:

- Configuring Disks to Use VMware Paravirtual SCSI (PVSCSI) Adapters:
http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1010398

VMware Compatibility Guide:

- Search the VMware Compatibility Guide:
<http://www.vmware.com/resources/compatibility/search.php>

Multimedia (YouTube):

- EMC & VMware: IP Storage-iSCSI, NFS, or Fibre Channel?:
<http://www.youtube.com/watch?v=VO46FyxGf3M>
- VMware on NetApp Overview:
<http://www.youtube.com/watch?v=zURPzi2XGxQ>

7.4 IBM DB2 REFERENCE INFORMATION

The following documents provide more information on the IBM DB2 database management system:

- DB2 Virtualization (IBM Redbooks publication):
<http://www.redbooks.ibm.com/abstracts/sg247805.html>
- Best Practices for DB2 for Linux, UNIX, and Windows:
<http://www.ibm.com/developerworks/data/bestpractices/>
- IBM DB2 Database for Linux, UNIX, and Windows Information Center:
<http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/index.jsp>
- DB2 Virtualization Support:
<http://www.ibm.com/developerworks/wikis/display/im/DB2+Virtualization+Support>

8 APPENDIXES: CONFIGURATION PARAMETER SETTINGS

These appendixes list the configuration parameter settings used for the three main tests described in Section 4:

- DSS
- OLTP
- OLTP with vCPU overallocation

8.1 APPENDIX A: DSS WORKLOAD SETTINGS

Table 1) DSS workload settings.

Configuration Level	Parameter	Value
DBM CFG	CPUSPEED	-1
	NUMDB	1
	FEDERATED	NO
	DIAGLEVEL	3
	NOTIFYLEVEL	3
	DFT_MON_BUFPOOL	OFF
	DFT_MON_LOCK	OFF
	DFT_MON_SORT	OFF
	DFT_MON_STMT	OFF
	DFT_MON_TABLE	OFF

Configuration Level	Parameter	Value
	DFT_MON_TIMESTAMP	OFF
	DFT_MON_UOW	OFF
	HEALTH_MON	OFF
	AUTHENTICATION	SERVER
	CATALOG_NOAUTH	NO
	TRUST_ALLCLNTS	YES
	TRUST_CLNTAUTH	CLIENT
	FED_NOAUTH	NO
	DFTDBPATH	/home/db2inst
	MON_HEAP_SZ	AUTOMATIC
	JAVA_HEAP_SZ	2048
	AUDIT_BUF_SZ	0
	INSTANCE_MEMORY	560000
	BACKBUFSZ	1024
	RESTBUFSZ	1024
	AGENT_STACK_SZ	1024
	SHEAPTHRES	786432
	DIR_CACHE	YES
	ASLHEAPSZ	15
	RQRIOBLK	32767
	QUERY_HEAP_SZ	1000
	UTIL_IMPACT_LIM	10
	AGENTPRI	SYSTEM
	NUM_POOLAGENTS	5
	NUM_INITAGENTS	0
	MAX_COORDAGENTS	AUTOMATIC
	MAX_CONNECTIONS	AUTOMATIC
	KEEPFENCED	YES
	FENCED_POOL	AUTOMATIC
	NUM_INITFENCED	0
	INDEXREC	RESTART
	TM_DATABASE	1ST_CONN
	RESYNC_INTERVAL	180

Configuration Level	Parameter	Value
	MAX_QUERYDEGREE	ANY
	INTRA_PARALLEL	NO
	FEDERATED_ASYNC	0
	FCM_NUM_BUFFERS	AUTOMATIC
	FCM_NUM_CHANNELS	AUTOMATIC
	CONN_ELAPSE	10
	MAX_CONNRETRIES	5
	MAX_TIME_DIFF	60
	START_STOP_TIME	10
DB CFG-per partition	SORTHEAP	8192
	DBHEAP	10000
	CATALOGCACHE_SZ	(MAXAPPLS*5)
	LOGBUFSZ	2048
	UTIL_HEAP_SZ	16
	BUFFPAGE	1000
	STMTHEAP	10000
	APPLHEAPSZ	AUTOMATIC
	APPL_MEMORY	AUTOMATIC
	STAT_HEAP_SZ	AUTOMATIC
	DLCHKTIME	10000
	LOCKTIMEOUT	-1
	CHNGPGS_THRESH	60
	NUM_IOCLEANERS	AUTOMATIC
	NUM_IOSERVERS	AUTOMATIC
	INDEXSORT	YES
	SEQDETECT	YES
	DFT_PREFETCH_SZ	AUTOMATIC
	TRACKMOD	OFF
	DFT_EXTENT_SZ	32
	MAXAPPLS	AUTOMATIC
	AVG_APPLS	AUTOMATIC
	MAXFILOP	16384
	LOGFILSIZ	12800
LOGPRIMARY	50	

Configuration Level	Parameter	Value
	LOGSECOND	0
	BLK_LOG_DSK_FUL	NO
	BLOCKNONLOGGED	NO
	MAX_LOG	0
	NUM_LOG_SPAN	0
	MINCOMMIT	1
	SOFTMAX	100
	LOGRETAIN	OFF
	USEREXIT	OFF
	AUTO_MAINT	OFF
	AUTO_DB_BACKUP	OFF
	AUTO_TBL_MAINT	ON
	AUTO_RUNSTATS	ON
	AUTO_STMT_STATS	ON
	AUTO_STATS_PROF	OFF
	AUTO_PROF_UPD	OFF
	AUTO_REORG	OFF
	AUTO_REVAL	DEFERRED
	CUR_COMMIT	ON
	DEC_TO_CHAR_FMT	NEW
	ENABLE_XMLCHAR	YES
	WLM_COLLECT_INT	0
	MON_REQ_METRICS	BASE
	MON_ACT_METRICS	BASE
	MON_OBJ_METRICS	BASE
	MON_UOW_DATA	NONE
	MON_LOCKTIMEOUT	NONE
	MON_DEADLOCK	WITHOUT_HIST
	MON_LOCKWAIT	NONE
	MON_LW_THRESH	5000000
DB2SET	DB2_LARGE_PAGE_MEM	DB
	DB2_EXTENDED_OPTIMIZATION	Y
	DB2_ANTIJOIN	Y
	DB2MEMDISCLAIM	N
	DB2RQTIME	30

Configuration Level	Parameter	Value
	DB2OPTIONS	-T -V +C
	DB2BQTRY	120
	DB2_PARALLEL_IO	*:5

8.2 APPENDIX B: OLTP WORKLOAD SETTINGS

Table 2) OLTP workload settings.

Configuration Level	Parameter	Value
DBM CFG	DFT_MON_BUFPOOL	OFF
	DFT_MON_LOCK	OFF
	DFT_MON_SORT	OFF
	DFT_MON_STMT	OFF
	DFT_MON_TABLE	OFF
	DFT_MON_UOW	OFF
	DFT_MON_TIMESTAMP	OFF
	HEALTH_MON s	OFF
	AGENTPRI	59
	MAXAGENTS	175
	NUMDB	1
	MAXDARI	-1
	NUM_INITDARIS	0
	NUM_POOLAGENTS	150
	NUM_INITAGENTS	0
	CPUSPEED	-1
	AUTHENTICATION	client
	RQRIOLBK	4096
	MON_HEAP_SZ	4096
	DIAGLEVEL	3
NOTIFYLEVEL	3	
SVCENAME	60000	
DB CFG-per partition	AUTO_MAINT	OFF
	AUTO_DB_BACKUP	OFF
	AUTO_TBL_MAINT	OFF

Configuration Level	Parameter	Value
	AUTO_RUNSTATS	OFF
	AUTO_STATS_PROF	OFF
	AUTO_PROF_UPD	OFF
	AUTO_REORG	OFF
	MINCOMMIT	1
	LOGBUFSZ	8192
	LOGFILSIZ	16384
	LOGPRIMARY	50
	LOGSECOND	0
	NEWLOGPATH	/ssd1/logs
	MAXAPPLS	175
	DBHEAP	8192
	SORTHEAP	16
	APPLHEAPSZ	328
	SEQDETECT	NO
	NUM_IOCLEANERS	20
	CHNGPGS_THRESH	40
	NUM_IOSERVERS	8
	MAXFILOP	800
	PCKCACHESZ	1000
	SOFTMAX	1000
	LOCKLIST	5000
	DLCHKTIME	3000
	MAXLOCKS	100
	SHEAPTHRES_SHR	0
	SELF_TUNING_MEM	OFF
	MON_REQ_METRICS	NONE
	MON_ACT_METRICS	NONE
	MON_OBJ_METRICS	NONE
DB2SET	DB2_HASH_JOIN	OFF
	DB2ASSUMEUPDATE	ON
	DB2_APM_PERFORMANCE	ON
	DB2COMM	tcPIP
	DB2_MAX_NON_TABLE_LOCKS	500
	DB2_KEEPTABLELOCK	ON

Configuration Level	Parameter	Value
	DB2_SELECTIVITY	ON
	DB2_USE_ALTERNATE_PAGE_CLEANING	YES

8.3 APPENDIX C: VCPU OVERALLOCATION OLTP WORKLOAD SETTINGS

Table 3) vCPU overallocation OLTP workload settings (per VM).

Configuration Level	Parameter	Value
DBM CFG	DFT_MON_BUFPOOL	OFF
	DFT_MON_LOCK	OFF
	DFT_MON_SORT	OFF
	DFT_MON_STMT	OFF
	DFT_MON_TABLE	OFF
	DFT_MON_UOW	OFF
	DFT_MON_TIMESTAMP	OFF
	HEALTH_MON	OFF
	AGENTPRI	59
	MAXAGENTS	175
	NUMDB	1
	MAXDARI	-1
	NUM_INITDARIS	0
	NUM_POOLAGENTS	150
	NUM_INITAGENTS	0
	CPUSPEED	-1
	AUTHENTICATION	client
	RQRIOBLK	4096
	MON_HEAP_SZ	4096
	DIAGLEVEL	3
NOTIFYLEVEL	3	
SVCENAME	60000	
DB CFG-per partition	AUTO_MAINT	OFF
	AUTO_DB_BACKUP	OFF
	AUTO_TBL_MAINT	OFF
	AUTO_RUNSTATS	OFF

Configuration Level	Parameter	Value
	AUTO_STATS_PROF	OFF
	AUTO_PROF_UPD	OFF
	AUTO_REORG	OFF
	MINCOMMIT	1
	LOGBUFSZ	8192
	LOGFILSIZ	16384
	LOGPRIMARY	50
	LOGSECOND	0
	NEWLOGPATH	/ssd1/logs
	MAXAPPLS	175
	DBHEAP	8192
	SORTHEAP	16
	APPLHEAPSZ	328
	SEQDETECT	NO
	NUM_IOCLEANERS	20
	CHNGPGS_THRESH	40
	NUM_IOSERVERS	8
	MAXFILOP	800
	PCKCACHESZ	1000
	SOFTMAX	1000
	LOCKLIST	5000
	DLCHKTIME	3000
	MAXLOCKS	100
	SHEAPTHRES_SHR	0
	SELF_TUNING_MEM	OFF
	MON_REQ_METRICS	NONE
	MON_ACT_METRICS	NONE
	MON_OBJ_METRICS	NONE
DB2SET	DB2_HASH_JOIN	OFF
	DB2ASSUMEUPDATE	ON
	DB2_APM_PERFORMANCE	ON
	DB2COMM	tcPIP
	DB2_MAX_NON_TABLE_LOCKS	500
	DB2_KEEPTABLELOCK	ON
	DB2_SELECTIVITY	ON

Configuration Level	Parameter	Value
	DB2_USE_ALTERNATE_PAGE_CLEANING	YES

NetApp provides no representations or warranties regarding the accuracy, reliability, or serviceability of any information or recommendations provided in this publication, or with respect to any results that may be obtained by the use of the information or observance of any recommendations provided herein. The information in this document is distributed AS IS, and the use of this information or the implementation of any recommendations or techniques herein is a customer's responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. This document and the information contained herein may be used solely in connection with the NetApp products discussed in this document.

