EMC IT’S VIRTUAL ORACLE
DEPLOYMENT FRAMEWORK
VCE Vblock, VMware vSphere, High Availability, Distributed Resource Scheduler, vMotion, Templates

Enabling EMC’s Journey to a Virtual Data Center
Creating a Foundation for Oracle as a Service

ABSTRACT

Migrating from a physical Data Center to a virtual Data Center creates challenges, in terms of what are the best practices for deploying virtualized Oracle databases. This white paper will illustrate EMC IT’s framework for deploying virtualized Oracle databases. EMC IT’s Oracle virtual deployment models are the foundation for the “as-a-Service” cloud deployment model.

April 2014
# TABLE OF CONTENTS

**EXECUTIVE SUMMARY** ........................................................................................................... 4
  Audience ................................................................................................................................. 4

**EMC IT’S ORACLE DATABASE VIRTUALIZATION JOURNEY** .................................................. 5
  EMC IT Overview ..................................................................................................................... 5
  EMC Virtual Infrastructure Evolution ..................................................................................... 5
  EMC IT’s Database Virtualization – a Phased Approach ......................................................... 6
    Application Rationalization .................................................................................................... 6
  EMC IT’s Virtualization business drivers .............................................................................. 7
  Oracle Virtualization enablers .............................................................................................. 7

**EMC IT ORACLE VIRTUALIZATION INFRASTRUCTURE DEPLOYMENT** .............................. 9

**ORACLE VIRTUAL DEPLOYMENT FRAMEWORK** ............................................................... 11
  Model A – VMware only Solution .......................................................................................... 12
  Model B – VMware Solution with Clustering Software ....................................................... 12
  Model C – VMware Solution with Oracle RAC for Availability ........................................ 12
  Model D – VMware Solution with Oracle RAC for Scalability .......................................... 12
  Oracle Virtual Deployment Model Creation ......................................................................... 12
  VMware Converter tool ........................................................................................................ 12
  VM Creation/Database Migration Method ............................................................................... 12
  VMware Templates ................................................................................................................ 13

**EMC IT’S ORACLE VIRTUAL DEPLOYMENT MODELS** .......................................................... 14
  Legacy - Oracle RAC Grids ..................................................................................................... 14
  Legacy Infrastructure ............................................................................................................ 14
  Current Deployment Model - Business Critical (BC) Oracle Grid ....................................... 15
  Tomorrow’s Deployment Example - Database (DB) VM Clusters ......................................... 15
  Database (DB) VM Clusters components ............................................................................. 15
  Model A - Oracle Single Instance Deployment Model – vSphere 5.0 ................................ 16
  vSphere Database Clusters ..................................................................................................... 16
  Model B - Oracle Mission Critical Single Instance Deployment Model ........................... 18
  Mission critical Single Instance SAP infrastructure ......................................................... 18
  High Consolidation Ratio Model .......................................................................................... 19
  Model infrastructure ............................................................................................................. 19
  Model VMware Infrastructure ............................................................................................... 20
  Model Oracle components .................................................................................................... 20
  Model C (Availability) and Model D (Scalability) - Oracle Mission Critical RAC Multi-Node Deployment Model – VMware 5.0 ............................................................... 21

**ORACLE VIRTUAL DEPLOYMENT MODELS BENEFITS** ....................................................... 23
  Oracle Single Instance ........................................................................................................... 23
  Oracle RAC ........................................................................................................................... 23
  Oracle RAC to Single Instance Use Case ............................................................................ 23
  Oracle Enterprise Edition to Standard Edition Use Case ....................................................... 24

**EMC IT ORACLE VIRTUAL DEPLOYMENT MODELS LESSONS LEARNED** ........................ 25
  NUMA and the Oracle Database ............................................................................................ 25
  Memory Access ...................................................................................................................... 25
  Huge Pages in Linux .............................................................................................................. 26
  Transparent Pages .................................................................................................................. 26
  Dynamic Coalescing .............................................................................................................. 26
  Latency ................................................................................................................................ 26
  BIOS settings ......................................................................................................................... 26
  Storage – Virtual Luns (VMDK) ............................................................................................. 26
EXECUTIVE SUMMARY

Migrating from a physical Data Center to a virtual Data Center requires best practices for deploying virtualized Oracle databases. This white paper will illustrate EMC IT’s framework for deploying virtualized Oracle databases, including both business supporting and mission critical applications, which will become a foundational for the “as-a-Service” cloud deployment model.

This paper illustrates EMC IT’s approach by showing its Oracle Virtual Deployment Framework powered by Intel® Xeon®. This framework answers the questions of when and what to deploy using Oracle virtual deployment models based upon two key business service levels of availability (99.9 to 99.999%) and scalability (from 1 CPU to N CPUs). These models make virtual deployment models of Oracle simplified, accelerated and give cost savings, consolidation and efficiency to EMC IT operations and infrastructure.

The Oracle Virtual Deployment Framework has the following models:

• **Model A** – VMware Solution with VMware technologies only. This solution is well suited for most deployments and provides a high level of availability.

• **Model B** – VMware Solution with Clustering Software. This solution provides additional availability and flexibility, but adds complexity. It is appropriate where very high levels of availability are required.

• **Model C** – VMware Solution with Oracle RAC for Availability. This solution should only be used systems that require zero downtime, however the complexity may introduce unplanned downtimes.

• **Model D** – VMware Solution with Oracle RAC for Scalability. This solution is well suited for databases that need to scale to 40+ CPUs, while using commodity hardware.

EMC is using these virtual deployment models to deploy its mission critical applications.

EMC is reducing its Operational Expense (OPEX, cost to run) and Capital Expense (CAPEX, cost to buy) in EMC’s Data Center by creating the following virtual database models:

• **Model A** – higher consolidation levels (physical to virtual servers), reduced CAPEX and OPEX, improved agility; deploy Oracle in minute versus months

• **Model B** – movement away from Oracle RAC model, significant CAPEX savings (No Oracle RAC licenses needed) and OPEX (Simplified Oracle Database Management, fewer OS images)

• **Model C** – higher (denser) consolidations, driving reduced CAPEX/OPEX

• **Model D** – improved resource utilization and hardware independence

**Audience**

This white paper is intended for CIOs, Oracle architects, virtualization architects, storage architects, Oracle Database Administrators (DBAs), and virtualization, server and network administrators.
EMC IT’S ORACLE DATABASE VIRTUALIZATION JOURNEY

EMC IT Overview
EMC IT supports over 60,000 users with approximately 500 applications on 22 PB of storage, 11,000 OS images of which 94% are fully virtualized, in 5 data centers across the globe.

EMC Virtual Infrastructure Evolution
To better understand EMC’s Oracle Database journey it is important to look at the virtual infrastructure journey EMC has taken from a dedicated “custom build” to the “self-service cloud” architecture and deployment models.

As the following diagram illustrates, Figure 1, EMC IT has traveled four phases on its Virtual Infrastructure journey:

Figure 1. EMC IT’S Virtual Infrastructure Journey
EMC IT’s Database Virtualization – a Phased Approach

The road to a virtualized data center is one of a careful phased approach to bring together the three key components in any organization but especially in a data center: people, process and technology into the planning, building and deploying the virtualized Oracle environments. This section highlights EMC’s phased approach.

Application Rationalization

Before any application is promoted to a virtual application, EMC IT performs a thorough process of application rationalization. This process analyzes EMC’s application portfolio, rationalizes the need for the applications, and standardizes and integrates it into EMC IT’s application environments to support its overall business strategy.

EMC IT database virtualization journey is a phased one. As you can see in the high-level illustration:

---

Figure 2. EMC IT’s Oracle Database Virtualization - Phased Journey
Phase 1 is plan, build, and deploy

Plan the migration from physical server(s) to virtual machine(s) (VM). Build the virtual machine (VM) infrastructure. Deploy all new builds virtualized. The following are examples of phase 1

- Commitment of all new builds as virtual
- Development and test application tiers
- Initial Conversions – Use VMware Converter to convert physical OS to a virtual OS

Phase 2 is virtualize

All lessons learned in Phase 1 are now virtually deployed on all Oracle deployments:

- All new Oracle builds
- All Oracle single instances
- RAC for availability
- Test and Non-Mission Critical production
- RAC to Single Instance

Phase 3 is finalize

The remaining Oracle deployments:

- Mission critical applications
- Ultra high volume applications

EMC IT’s Virtualization business drivers

The following are the business drivers for virtualization

1. **Consolidation** – reduction in the Data Center
2. **Cost savings** – servers, storage, retired legacy platforms, software licenses
3. **Operational agility** – faster deployments both OS and database
4. **Standardize** – compute platform x86 OS platform Linux, virtualization – VMware
5. **Simplify** – operations enabled by the standardize driver

The above business drivers map to Capital Expense (CAPEX, cost to buy) which is found in drivers 1 and 2 (Consolidation, Cost Savings). Operational Expense (OPEX, cost to run) is seen in drivers 3 thru 5 (Operational agility, Standardization and Simplify).

Oracle Virtualization enablers

These are the technologies that enable Oracle virtualization today for EMC IT:

- Intel’s latest CPU technologies (Xeon/Nehalem)
  - CPU advances, coupled with ever increasing memory density means that the amount of CPU and memory available in a single server is many times what is needed by most single instance databases. This means that the only way to drive efficient use of server resources in many cases is to virtualize.
- Migration to an open hardware and software platform
  - Please review the following:

  [EMC IT’s "On-Ramp" to the Journey to the Private Cloud](#)
• VMware vSphere
  - With the release of vSphere 4 EMC was able to create VMs with 8vCPUs and 255GB of memory, enough to take on large
database workloads. With vSphere 5, these limits increased even further to 32 vCPUs and 1TB of memory, enough to
handle extremely large database workloads.

  Update: vSphere 5.1 scales this limit to 64 vCPUs

• Availability and Resource Pooling
  - Please review the following:
    http://www.vmware.com/products/high-availability/overview.html
    http://www.vmware.com/products/drs/overview.html

• VMware Oracle Support Policy
  - VMware ownership of Oracle technical issues:
    - http://www.vmware.com/support/policies/oracle-support.html
The following diagram illustrates the deployment strategy used for Oracle instances types:

**Figure 3. Oracle Virtualization Deployment Model and Oracle instances**

As seen in the diagram below, within one or more ESX Clusters, Oracle databases are separated into similar groups. This is primarily done for Oracle license compliance, but it also allows EMC to group the resources by similar footprints. These groups can be physically separate ESX clusters or logically separated using VMware’s DRS host affinity groups.

DRS, or Distributed Resource Scheduler, is a unique technology that allows vSphere to balance available computing power amongst all ESX servers within a resource pool. It manages this by performing VMotions of VMs from more loaded ESX servers to lesser-loaded ESX servers. This allows for peak utilization of all compute resources and enables significantly higher consolidation ratios than other virtualization technologies. By assigning a VM to a specific resource pool, it prevents that VM from starting or being moved to a server outside that pool.
The Oracle Database groups that EMC have defined are the following:

- Oracle Enterprise Edition
- Oracle Standard Edition
- Oracle RAC
- Oracle Enterprise Edition - Externally facing

Within these groups, there are many different types of VM categories. For the smallest databases, EMC combines several application databases onto one VM. This can be either physically separate databases or multiple schemas in one database. For larger databases, or mission critical databases, EMC dedicates a single VM for them. This allows greater availability, since there are not competing requirements between applications. This approach allows EMC to achieve 99.99% uptime for these databases.
The decision when and what to virtualize is often debated and is not generally based on facts. For instance, many infrastructure administrators believe everything should be virtualized, just because they can be. On the other hand, many DBAs will resist virtualization due to the belief that databases cannot be virtualized efficiently and therefore it should not be done.

The reality is somewhere in between. There are many advantages to virtualization and some limitations, but generally, all database and applications will gain benefits simply by virtualizing them.

EMC IT has developed guidelines based on those advantages and limitations, as seen in Figure 5.

![Figure 5. EMC IT’s Virtual Deployment Model Framework](image)

Using two driving business service levels, availability and scalability, EMC IT created a virtual deployment framework. It uses a spectrum of availability to meet EMC business users.

The following tables illustrate the availability criteria:

<table>
<thead>
<tr>
<th>Availability %</th>
<th>Downtime per year</th>
<th>Downtime per month*</th>
<th>Downtime per week</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.9% (&quot;three nines&quot;)</td>
<td>8.76 hours</td>
<td>43.2 minutes</td>
<td>10.1 minutes</td>
</tr>
<tr>
<td>99.99% (&quot;four nines&quot;)</td>
<td>52.56 minutes</td>
<td>4.32 minutes</td>
<td>1.01 minutes</td>
</tr>
<tr>
<td>99.999% (&quot;five nines&quot;)</td>
<td>5.26 minutes</td>
<td>25.9 seconds</td>
<td>6.05 seconds</td>
</tr>
</tbody>
</table>

* For monthly calculations, a 30-day month is used
Scalability criteria were based upon the number of physical CPUs needed by the application. The scale criteria grow from 8 CPUs to N (i.e. 32 physical CPUs) and map to corresponding virtual CPUs (vCPUs).

The following are the high level descriptions of EMC’s four Oracle Virtual Deployment Models:

**Model A – VMware only Solution**
This Oracle Virtual Deployment model is deployed on a “pure” VMware solution to deploy an Oracle virtualized instance. This solution relies entirely on VMware HA, DRS and SRM and the new VMware Guest App Monitor technologies.

**Model B – VMware Solution with Clustering Software**
This Oracle Virtual Deployment model is deployed on VMware software with the addition of clustering software, such as Oracle clusterware (CRS).

**Model C – VMware Solution with Oracle RAC for Availability**
This Oracle Virtual Deployment model is deployed on VMware software with the addition of Oracle RAC. This model is required to meet the business service levels of availability of 5x9s or 99.999% (less than 5.26 minutes of unplanned downtime per year).

**Model D – VMware Solution with Oracle RAC for Scalability**
This Oracle Virtual Deployment model is deployed on VMware software with the addition of Oracle RAC. This model is required to meet the scalability and processing power needed by the largest Oracle applications.

**Oracle Virtual Deployment Model Creation**
There are two methods to deploy an Oracle virtual environment.

**VMware Converter tool**
VMware’s vCenter Converter tool provides an easy-to-use solution to automate the process of creating VMware virtual machines from physical machines (running Windows and Linux), other virtual machine formats, and third-party image formats.

**Benefits**

- Convert physical machines running Windows or Linux operating systems to VMware virtual machines quickly and with little disruption or downtime.
- Convert third-party image or virtual machine formats such as Parallels Desktop, Symantec Backup Exec System Recovery, Norton Ghost, Acronis, StorageCraft, Microsoft Virtual Server or Virtual PC, and Microsoft Hyper-V Server virtual machines to VMware virtual machines.
- Enable centralized management of remote conversions of multiple physical servers or virtual machines simultaneously.
- Ensure conversion reliability through quiesced snapshots of the guest operating system on the source machine before data migration.
- Enable non-disruptive conversions through hot cloning, with no source server downtime or reboot.

**VM Creation/Database Migration Method**
The second method is to build a new VM and migrate the Oracle database from the physical server to the VM. This is a useful method if you’re trying to further standardize the version of the Operating System. Rather than keep the same version that is on the physical server, you’re free to start with a new version. While this requires slightly more effort to ensure Oracle database (DB) and Operating System (OS) compatibility, it can be worth it if your current physical environment lacks standardization. This method of migration also allows you to more easily change the physical layout of the drives that the database is installed on.

The database migration may be a physical file copy of the datafiles or a full database export/import. The migration could also be an in place migration by creating virtual raw device mapping, vRDM. With a migration to vRDMs, the actual database files do not require movement and allows for a very rapid cutover and backout, if you ever needed to. The only disadvantage to the RDM method is the loss of storage vMotion capabilities and complications introduced with SRM.

Both migration methods have their advantages. It is often beneficial to use a combination of both methods depending on the specific database being migrated.
VMware Templates

VMware Templates are Virtual Machine (VM) images that are copied to make new VMs. Using templates removes a lot of the repetitive, manual steps from the process of creating a new VM. Standard templates include the base Operating System (OS) as well as all the standard software that is generally installed on all machines such as anti-virus, backup agents, monitoring agents, etc. Using templates allows for faster deployments as well as providing a consistent configuration across your environment.

For running databases on VMware, the templates can be enhanced further. By adding the base install of Oracle software, deployment time can be reduced by hours and can also include best practices such as hugepage configuration, Oracle required kernel parameters, and standardized directory locations. With the addition of a little scripting, in the templates, the process can include renaming a pre-created database and incorporating all best practices into the init.ora, as well as security best practices both inside the database and in the listener configuration. These additions extend the benefits of templates up into the database, removing all of the manual steps in the creation of a new database VM.
EMC IT’S ORACLE VIRTUAL DEPLOYMENT MODELS

This sections focuses on EMC IT’s Oracle virtual deployment models A, B, C, and D and will illustrate them with specific examples after first describing the previous and existing models:

**Legacy - Oracle RAC Grids**
EMC has consolidated 48 standalone Business Supporting/Business Critical Oracle database servers into two sets of Oracle 2-node RAC on physical systems:

- Production
- Development and Test

See figure 6 below.

**Legacy Infrastructure**
Before the consolidation, there were 51 Oracle databases running under 9 different Oracle versions and three different OS (MS Windows, Linux, and Solaris) and on 51 different physical servers. Three of these applications were actually decommissioned.

The consolidated Business Critical Oracle Grid runs on one Oracle version and one version of Red Hat Linux OS and with only 3 (production, test, and development) databases that host 38 separate business critical or business supporting applications, some of the applications do not have a test environment. It uses only 4 physical servers, a 2 node RAC for production and a 2 node RAC that is shared for test and development.

This greatly simplified system maintenance tasks, reduced database management overhead and freed up DBA time for more proactive initiatives.

The achieved consolidation ratio for database is about 13:1 and application to database ratio is an astonishing 38:1. The realized cost saving was about 2.5 million US dollars over a three year of period.

*Figure 6. EMC’s Legacy Business Critical and Today’s Oracle RAC Grid Deployments*
Current Deployment Model - Business Critical (BC) Oracle Grid

The current Business Critical (BC) Oracle Grid is hosted on x86-64 platform with Red Hat Linux Enterprise Edition 5.x and uses Oracle Automatic Storage Management (ASM) for database storage management.

Oracle Real Applications Cluster (RAC) provides the high-availability and load balancing for the databases running in BC Oracle Grid. The Grid database is a multi-tenant database where individual applications have their schema(s) within the single grid database.

Each of the applications is assigned separate service names for database connections.

Tomorrow's Deployment Example - Database (DB) VM Clusters

EMC is in the process of building and deploying Database (DB) VM Clusters with the purpose of hosting all Business Critical and Business Supporting databases in the cluster. The DB VM Cluster is being deployed on VMware vSphere 5.0 and is only supporting VMware HA which provides 99.9% uptime. At EMC, only Mission Critical databases require more than 99.99% uptime. At EMC, only Mission Critical databases require a disaster recovery site. Since this ESX cluster will not host mission critical databases, vSphere Site Recovery Manager, SRM, will not be required.

Database (DB) VM Clusters components

The database VMs have a dedicated ESXi cluster and are following the standard building blocks (Figure 7) with the following default specifications:

- 4 vCPUs x 32 GB RAM
- Red Hat Enterprise Linux 5.6
- Oracle version: 10.2.0.4, 11.1.0.7, or 11.2.0.3
- VMFS
- Shared datastores dedicated for data, redo, temp and archive
- FAST VP with tiered storage

The multi-tenant VMs as shown in Figure 7 are replacing the current physical BC Grid production, test, and development databases and will continue hosting multiple applications as schema(s) in the database. Other VMs, in the cluster are hosting individual databases for specific non-Mission Critical applications. There are also VMs that host small business supporting applications that have very low usage patterns. These VMs have multiple databases running on them. This method is used to drive even higher efficiencies out of the database tiers by reducing the OS images and databases that need to be supported.

All new database VMs will be built using predefined templates and will be available immediately for deployment. Each of the database VMs could support multiple databases and one of three Oracle versions (10.2.0.4, 11.1.0.7, and 11.2.0.3). The database storage will be pre-allocated to database VMs and will be shared by databases on the same ESXi cluster.

With these Oracle database VM building blocks at hand, EMC Database administrators (DBAs) will be able to fulfill the needs of application teams and / or projects for agile deployment of Business Critical or Business Supporting Oracle databases. The turnaround time for Oracle database requests will be reduced to less than 2 days from weeks of database build lead time.
Model A – Oracle Single Instance Deployment Model – vSphere 5.0
Database-as-a-Service (DBaaS) has been much discussed in general cloud discussions, but often the use case has not been made clear. EMC has been building this use case for several years and are now deploying based on it in EMC’s Data Center to solve real problems.

vSphere Database Clusters
EMC IT has developed VMware vSphere clusters specifically for database VMs. There is a cluster for production workloads as well as a separate cluster for non-production workloads. Both clusters reside on the same VCE Vblock Reference Architecture. The Vblock architecture makes it very easy to grow either cluster in response to increased demand. Having two separate clusters allows for the testing of upgrades in non-production without any worry of effecting production.

These clusters use resource pools to provide different levels of service depending on the needs of the particular workload. Less critical workloads are put in resource pools that will yield resources to the more critical workloads in the event of resource contention. Through this use of resource pools, EMC IT is able to safely over-provision resources without any detrimental effects to critical production workloads.
For Oracle databases, production is logically fenced to a smaller set of DRS groups, for licensing reasons, but the extra unused capacity can also be used to augment the rest of the cluster. This database cluster allows EMC IT to rapidly provision environments.

The results are EMC IT is no longer tied to procurement cycles and business justifications to get projects off the ground. It also allows EMC the ability to provide resources for a virtual lab, a place where DBAs, Developers and System Administrators can use the extra capacity to perform POCs, upgrade tests or tool trials. This method of combining lower tiered database VMs alongside higher tiered VMs and guaranteeing service levels allows a much higher consolidation ratio than ever before.

EMC IT is now able to deploy databases and application tiers in minutes to hours, versus days to. This ability gives EMC the agility to complete projects in a fraction of the time they were able to before. It also gives IT the ability to spend time on design and best practices, instead of procurement and build activities.
Model B – Oracle Mission Critical Single Instance Deployment Model

In the case of EMC’s ERP instance, EMC has a need for very high availability. Previously this would have required running many physical Real Application Cluster (RAC) databases to support a 4x9’s (99.99%) SLA. With the availability of VMware’s server virtualization technology EMC is almost able to achieve this. The added requirement of long patching times for the OS and RDBMS kernel prevented EMC IT from fully achieving the stated SLA with VMware HA alone. For this particular instance, EMC solved the problem by adding Oracle Clusterware.

Clusterware adds the ability for a database or application to fail-over to another VM. This failover works for planned and unplanned events. EMC IT deploys this technology in their SAP ERP system both for the databases as well as the SAP Central Instance components.

Clusterware provides a solution for several use cases in this deployment model:

- Hardware failure – Oracle CRS fails database to standby mode. Users momentarily impacted
- Database Listener failure – CR restart listener. No impact to users.
- Database failure – CRS restarts the database. Users momentarily impacted.
- OS or Database patching – Rolling upgrade. No impact to users.

Mission critical Single Instance SAP infrastructure

This deployment is a mission critical SAP landscape, utilizing VMware vSphere and Oracle CRS to support multiple single instance databases and SAP application components.

The SAP modules utilizing this infrastructure include:

- Enterprise Central Component (ECC)
- Business Warehouse (BW)
- Product Lifecycle Management (PLM)
- Supplier Relationship Management (SRM)
- Supply Chain Management (SCM)
- Business Intelligence (BI)

The underlying infrastructure is on a Vblock running VMware’s vSphere 5.0 virtualization platform, upgraded from vSphere 4.1. It provides many mission critical VMs within a single ESXi cluster along with a logically isolated Oracle CRS grid supporting many elements within the SAP stack.

In this ESXi cluster, 8 larger VMs house the Oracle single instance databases. See diagram below. They are isolated from the rest of the cluster due to Oracle license requirements and are additionally protected with Oracle CRS. The SAP CI components are also running in the CRS cluster and benefit from CRS HA functionality.

In the event that one of the database or database components, in this cluster crashes, CRS will automatically bring that resource back online. In the event that the crash was a result of an OS crash or server failure, the database is brought up on another VM immediately. The failed VM is then started back up and is then available as a failover target.
The remaining SAP application components run in the rest of the ESXi cluster and benefit from built-in VMware vSphere capabilities such as HA, vMotion, and DRS.

There are many advantages to this deployment beyond the typical server consolidation. This environment benefits from all the features of vSphere, including reduced hardware costs, easier/faster upgrade paths and improved utilization of data center resources. With this deployment framework EMC is also able to achieve similar uptimes as in a Real Application Cluster (RAC) deployment, without the expense and administration overhead that an Oracle RAC database brings.

**High Consolidation Ratio Model**

Today, in EMC IT’s Project Propel (SAP), there would be approximately 500 physical servers currently in use. Instead, EMC IT replaces these with a 100% virtual infrastructure of approximately 72 ESXi servers which is a consolidation ratio of about 7:1. This includes the lifecycle of SAP environment; production, 2 x development, 3 x test, patch, performance, etc.

**Model infrastructure**

- The SAP program was deployed on a VCE Vblock, series 700. This includes Cisco UCS high density half width blades with Intel Xeon Nehalem and Westmere processors and 256GB of RAM.

- The environment also makes use of the converged networking architecture inherent in the Cisco UCS design. Redundant components and data center best practices around cabling and switch configuration help to reduce the impact of hardware failures.

- Best of breed storage infrastructure from EMC, including VMAX, Unified (VNX), and VPLEX are key components. EMC Symmetrix VMAX and VMAXe based on Intel Xeon Processors are high-end storage arrays optimized for the virtual data center. Intel Stop and Scream detects poison packets in PCIe, and enables enhanced error isolation in a multi-blade, highly-available environment. This results in shorter downtime; faster problem diagnosis; and simplified repair process, enabling the IT manager to optimize the virtual data center. EMC VNX family utilize the Intel Xeon 5600 series processors, which help make it 2-3x faster overall than its predecessor. The VNX quad-core processor supports demands of advanced storage capabilities such as virtual provisioning, compression, and deduplication. Furthermore, performance of the Xeon 5600 series enables EMC to realize its vision for FAST on the VNX, with optimized performance and capacity, without tradeoffs, in a fully automated fashion.
• vSphere physical raw device mapping (pRDM) storage was used along with EMC array based replication technologies (TimeFinder, SRDF) for fast environment cloning for large databases, efficient disaster recovery and high speed non-disruptive backups.

• VMware vSphere 5.0 is the standard data center “operating system”/hypervisor and provides robust availability and management capabilities along with the support for large VM’s.

• The Cisco UCS blade environment make use of a “service profile” as well as SAN boot to provide a true stateless hardware Vblock.

**Model VMware Infrastructure**

• VMs
  - Installed using a set of standard templates. Each defines a specific class of VM (database, SAP dialog instance, etc)
  - VMware pRDM devices are split across three SCSI controllers for optimum performance utilizing PowerPath/VE for VMware vSphere

• Operating Systems
  - Redhat AS 5.6
    • Oracle did not supporting 6.x at deployment time
    • Internal satellite server to manage releases
  - Microsoft Windows 2008R2
  - SAP code presented via NAS

**Model Oracle components**

• Oracle Cluster Ready Services(CRS) 11.2
  - CRS is protecting:
    • Oracle single instance databases
    • Oracle listeners
    • SAP enqueue replication
    • SAP Central Instances
    • Application virtual IP addresses

• Oracle ASM
  - Modules ECC and BW each contain their own set of diskgroups for SAN based storage replication to the disaster recovery site and backup isolation.
  - Diskgroups are separated into data, redo, archive and temp.
    • Oracle database version 11.2
    • Asmlib – for ease of configuration and troubleshooting
    • RMAN – for database backups
    • Oracle Enterprise Manager – for database monitoring and resolution
Model C (Availability) and Model D (Scalability) - Oracle Mission Critical RAC Multi-Node Deployment Model – VMware 5.0

EMC IT’s mission critical databases require 99.999% (less than 5.26 minutes of unplanned downtime a year) availability because they are revenue-impacting systems. These databases are highly scalable to meet EMC IT’s “End of Quarter” and “End of Year” processing demands. Oracle RAC is used to achieve this kind of availability and scalability requirements.

For these reasons of availability and scalability, these databases require both Model C and D.

Today, EMC mission critical CRM application currently runs on a physical 4-node RAC database (See Figure 10) on a VCE Vblock Reference Architecture 700 platform with plans to move to vSphere 5. Middle tier application servers are already running on a virtualized platform. (See the EMC IT Whitepaper: EMC IT’s “On-Ramp” to the Journey to the Private Cloud)

![Figure 10. EMC IT Current Oracle CRM Deployment Model](image)

Traditionally, if you need to add a new node to a RAC database, procuring and building a new server consumes most of a DBA’s time. It can take months before the new node is ready to be added to the production RAC database. With vSphere, you can build the new VM with help of templates in a matter of minutes to hours. You can then add this new VM to your existing RAC database without any downtime.

Additionally, if you have hardware capacity in your ESX/ESXi cluster, you can horizontally scale your RAC database in matter of hours. This is very useful when you need additional compute power. Running Oracle RAC on vSphere drastically reduces turnaround time for adding new nodes.

Models A and B work great when you can run your workload on one server. When you need more resources than one server to run your workload, you need to scale your database horizontally. Oracle RAC helps in this case. Oracle RAC database scales across multiple servers providing enough horsepower to satisfy your workload.

This mission critical database will be deployed on a virtualized 4 node RAC database deployed on VMware vSphere 5.0 (See Figure 11). Previously, EMC IT was not able to accomplish this on vSphere 4.1 because of 8 vCPUs per VM limitation. With vSphere 5.0, each VM can have up to 32 vCPUs and 1 TB of memory. This enables EMC IT to virtualize any mission critical database, which requires additional compute power.

This will be deployed in the future using a dedicated vSphere 5.0 cluster on the same Vblock Reference Architecture infrastructure. Based on Oracle licensing requirements, this cluster will be reserved for RAC databases, where we will be able to add other RAC databases on the same physical hardware that was used to host this database.
Figure 11. Virtualized Deployment Model C (Availability) Model D (Scalability)
ORACLE VIRTUAL DEPLOYMENT MODELS BENEFITS

The following section details the benefits of the virtualized Oracle deployment models (RAC, Single Instance) and the ability to migrate from an Oracle Enterprise Edition to a Standard Edition model.

Oracle Single Instance

Oracle single instance, non-RAC, databases benefit greatly from deploying on VMware vSphere environments. The most obvious of these is high availability. Since the OS is no longer bound to a physical server, the database itself is also not bound. In the event of any type of hardware failure, the database is simply started on another node. This feature is referred to as VMware HA and automatically adds high availability to any database or application.

With vSphere 5.0, another level of HA capability was added. vSphere now has the ability to monitor services, called VMware GuestAppMonitor. This feature has the ability to monitor services, such as listener and databases, and in the event that the service is no longer responding, vSphere can restart the VM to resolve the issue.

Oracle database servers typically have very low utilization. Virtualizing within a VMware cluster allows EMC to deploy more databases into fewer physical servers to more efficiently utilize the unutilized resources on a server. EMC can also consolidate more on these databases together due to VMware DRS technology. Both of these result in significant savings on hardware costs as well as license and support costs for the databases running on them.

Distributed Resource Scheduler is a capability that utilizes vMotion to balance resource usage among all of the servers within the DRS resource group, by default all servers in the cluster. This means that if one or more database servers require more resources than are available on the physical ESX server, vSphere DRS will non-disruptively migrate one or more VMs to other, lesser-utilized ESX servers. This technology, vMotion, is also available to administrators to migrate running VMs from one ESX server to another. This technology, vMotion, is also available to administrators to migrate running VMs from one ESX server to another. This vMotion migration occurs with no downtime to the services running on that VM and almost no noticeable overhead to the transactions occurring on that VM.

Another add-on, separately licensed feature is VMware vCenter Site Recovery Manager (SRM). SRM, in conjunction with EMC replication like SRDF and Recover Point technology, is a feature that allows a fully automated disaster recovery environment with little to no scripting required. It provides a repository for DR run plans that are self-documenting and centrally located. It also allows complete DR testing without impacting the production system by performing a simulated failover while fencing the environment from actual production systems.

Oracle RAC

In addition to the many benefits of virtualizing Oracle on vSphere, RAC has one more. With RAC on physical servers, there needs to be at least one extra node in the CRS cluster to handle the load, in the event of a server failure. With vSphere, this requirement does not exist, since the VM will be immediately restarted on another ESX server in the cluster.

Oracle RAC to Single Instance Use Case

In many cases EMC IT had deployed Oracle RAC in support availability requirements. This was especially true in the mission critical databases. This significantly complicates the administration of the database, and in many cases actually was the root cause of outages EMC incurred. By migrating these databases to VMware’s vSphere product, EMC IT is able to significantly simplify EMC’s Oracle database landscape. In many cases, this will result in better uptimes than EMC had with RAC. It can also significantly reduce your operating expenses in license, labor and hardware costs.

Take, for example, a small two node RAC supporting a mission critical application. By moving this to a VM, you are eliminating/removing the following:

1. One full oracle license
2. Oracle RAC license
3. Reduced Oracle support costs
4. A complete Oracle RAC server
5. A dedicated VLAN for the Oracle RAC interconnect
6. Reduce administration costs and complexity
Multiply the above cost saving (Savings 1-6) to your other Oracle instance environments such as Development, Test, and Patch; the monetary benefits are significant.

**Oracle Enterprise Edition to Standard Edition Use Case**

In many cases, a RAC license would also bump a database requirement from standard edition (4 sockets) to enterprise edition, since you need to have reserve capacity in your RAC cluster to support a node failure. With the removal of the RAC requirement, each machine is measured separately and I can put many standard edition VMs on a single server with 4 or fewer sockets.
EMC IT ORACLE VIRTUAL DEPLOYMENT MODELS LESSONS LEARNED

The following section contains some of the highlights of the lessons learned regarding Oracle from this migration to Oracle virtual deployment models.

EMC IT performed extensive testing within EMC IT’s Lab prior to promoting to production these new technologies.

In this lab EMC IT is able to test all of EMC’s theories, other’s best practices and common tuning opportunities prior to deploying to EMC’s production environment.

Among EMC’s test environments is a production copy of EMC IT’s 8 TB Oracle eBusiness Suite application.

This instance is a copy of EMC’s production environment where EMC IT has tested various releases of vSphere, various disk layouts and cloning techniques as well as countless performance tuning tests.

Based on the lessons learned in the lab and then tested in EMC’s production databases, EMC IT has put together some VMware vSphere best practices for Oracle databases. Many of these practices also work with other types of VMs and other databases, but the following list is intended and tested specifically with Oracle databases.

NUMA and the Oracle Database

Oracle Databases up to and including 11g have tried to make use of local access to memory on systems that support NUMA. The implementations however, were not efficient and almost always result in performance degradation and not an improvement. In Linux, the default implementation does not properly link into the NUMA shared library, so the NUMA “optimizations” are not actually working. In testing in the lab it was confirmed that enabling NUMA caused performance to either get worse or at best stay the same. To be sure that Oracle NUMA is not used it should be explicitly disabled with _enable_NUMA_optimization=FALSE and _db_block_numa=1.

The operating system also can be NUMA aware and will attempt to schedule a process to a CPU on a socket that has the memory a process is using. In the case of a database process, the OS might fail to successfully schedule the process on the same socket as the needed memory, but fail very quickly and the database process will access the local memory randomly.

In order for the OS to be NUMA aware and attempt to schedule processes with local NUMA node memory, NUMA must be enabled in the hardware BIOS. This is the default on most systems. In Linux the command numactl — hardware shows the number of NUMA nodes as well as the distance. If the results of numactl — hardware show only one node, then NUMA is not enabled.

In order to improve the process’ chance of hitting that local memory, the database VM should utilize the fewest number of sockets possible.

For example, a database VM is using 20 vCPUs with a utilization of 40% and is running on an Intel Nehalem based ESX server (8 cores per socket with 4 sockets). By using 20 vCPUs the VM is spread across 3 sockets. Consider running the database with 16 vCPUs, which will reduce the number of sockets used to two. The increased chances of access to local memory will help reduce the overall CPU requirement as the threads accessing memory will complete much faster, about 20% in EMC’s testing.

The ability to influence the VM’s likelihood of hitting that local memory can be significantly increased, in vSphere 5.0 by setting the parameter numa.vcpu.preferHT=TRUE. This can be done in the .vmx file and will instruct vSphere to schedule the vCPUs on the fewest sockets as possible by utilizing both the core and its associated hyperthread. As the parameter indicates, the VM will prefer to use HyperThreads over spreading across as many physical cores as possible. In the previous example, this would put the entire 16 vCPUs VM into one socket and all memory access would be local.

Memory Access

Virtual memory, within vSphere, is only given to the VM when it is actually used. This is a very good feature as it allows the unused memory to be allocated to other VMs that actually need it. The downside, for Oracle databases, is that the database does not allocate all of its memory at startup, unless you use the init.ora parameter PRE_PAGE_SGA=YES. This can lead to VMs having to be moved, with Distributed Resource Scheduler (DRS), as the database memory utilization grows. In an even worse scenario, it could lead to portions of the Oracle SGA memory being swapped to disk if memory resources become too limited.
To avoid this, set memory reservations for database VMs to the size of the SGA, PGA and kernel combined that the database will be using during its normal use. A good estimate would be the size of the SGA + 2x pga_aggregate_target + 500MB for Linux. The VM can be set larger to account for the unusual loads. Memory will still be given to the database VM only when it needs it, the reservation amount will be set aside on the ESX server for that VM.

**Huge Pages in Linux**

Using Huge Pages is another method to efficiently allocate memory to the VM that the database will use. With Huge Pages, the Linux VM server will actually allocate that memory to the VM at startup and will not have to give it to the VM as the database is warming up. There are many other benefits of using Huge Pages for Oracle databases and they are covered extensively in other whitepapers. (See the EMC IT Whitepaper: [EMC IT's "On-Ramp" to the Journey to the Private Cloud](#))

**Transparent Pages**

Transparent page sharing is another unique feature of VMware vSphere and is very efficient at minimizing each VMs use of memory. Essentially, the ESX server can recognize that more than one VM are using identical memory pages and can map both VM’s to use the same physical memory. This can significantly reduce the actual memory consumption of individual VMs. For Databases and some other workloads, this can degrade performance if that memory actually changes frequently, as is true with Oracle memory regions, such as buffer caches. Memory reservations prevent this feature from activating, so it is safe and recommended to leave this enabled.

**Dynamic Coalescing**

Dynamic coalescing is a feature where vSphere will wait a very short time before forwarding a network packet to its destination in an attempt to improve network utilization. Normally this is a good feature as this reduces network traffic, which can reduce latencies. With Oracle RAC, this is not a good thing. RAC databases are especially sensitive to interconnect latencies and the network is generally segregated and does not benefit from the coalescing. In this case, for the interconnect network card, this feature should be disabled. This can be done with EthernetX.coalesceScheme = ‘disabled’ in the VMs .vmx file.

**Latency**

To reduce this latency even further, set the Interrupt rate on Network Adapter to 30K within the VM network adapter settings. This will significantly increase the number of times the network card is polled to see if there is work to do. There is a slight increase in CPU consumption, but the performance improvement is worth the overhead.

**BIOS settings**

The additional best practices in the hardware’s BIOS are the following:

- Disabling the BIOS power savings
- Enabling hyperthreading
- Enable processor and memory virtualization off load

**Storage – Virtual Luns (VMDK)**

Most storage arrays will perform best with sequential workloads. The exception to this are the newer all Flash arrays. Of course these arrays are much more expensive, so this capability comes at a steep price. Many of the newer arrays support thin provisioning. Thin provisioning is a great new feature that can save large amounts of disk space in the storage array. Virtual thin provisioning however, can easily introduce fragmentation, since the underlying storage array is unaware that the virtual layer is providing this service.

It is also important to understand that, as of vSphere 5.1 there is a limitation of 256 luns per ESX cluster. This number is important as it is part of the equation when deciding how to layout your virtual luns. With vSphere you have three choices when creating Virtual luns (vmdk):

1. **Thin** – Storage is only allocated when it is used.
2. **Zero Thick** – Storage is only allocated when it is used, but total space is reserved on the datastore. Works with storage array thin pools.
3. **Eager Zero Thick** – Storage is fully allocated when virtual lun is created.
Keeping the above in mind, we follow these best practices

- Redo log luns should always be provisioned eagerzerothick, minimum 2 luns.
- Datafile luns should be provisioned as either eagerzerothick or on a datastore that is dedicated to that virtual lun, minimum of 4 luns when using ASM. Dedicated luns are more storage friendly, but use significantly more luns in the ESX cluster.
- Archive luns should always be provisioned as eagerthick. Archive logs are created at once and therefore generate their own full provisioning.
CONCLUSION

This paper showed EMC IT’s approach to its “Oracle Virtual Deployment Framework” powered by Intel Xeon, which answers the questions of when and what to deploy virtually using EMC IT’s virtual deployment models based upon two key business service levels; availability (99.9 to 99.999 %) and scalability (from 1 CPU to N CPUs) in their Data Center.

EMC is using these virtual deployment models to deploy its mission critical applications.

The Virtual Deployment Models benefits show the following:

- **Model A** – VMware only solution – higher consolidation (physical to virtual servers), reduced CAPEX and OPEX, improved agility; deploy Oracle in minute versus months
- **Model B** – VMware with Clusterware – Movement away from Oracle RAC model, significant CAPEX savings (No Oracle RAC needed) and OPEX (Simplified Oracle Database Management)
- **Model C** – Availability VMware with RAC – higher (dense) consolidation, reduced CAPEX/OPEX
- **Model D** – Scalability VMware with RAC - improved resource utilization, become hardware independent

The EMC IT Oracle Virtual Deployment Models documents lessons learned and important information you can use in your successful journey to a Virtual Data Center on the following topics:

- NUMA and Oracle
- Database Memory Access
- Huge pages in Linux
- Transparent Pages
- Dynamic Coalescing
- Latency
- Hardware settings
- Virtual Luns
REFERENCES
EMC IT: http://www.emc.com/microsites/emc-it-proven/index.htm
VCE Vblock: http://vce.com/

Acknowledgments
The author would like to thank the EMC IT teams for assistance in the creation of this white paper.