Integrated Data Warehouse Solution with Toad BI Suite and Dell PowerEdge VRTX Enclosure

This reference architecture provides guidelines and best practices to achieve a compact and optimized platform for data warehouse configuration and analytical deployments on the Dell PowerEdge VRTX enclosure.

Dell Global Solutions Engineering

Authors: Anthony Fernandez, Mayura Deshmukh, Deepthi Cherlopalle, Scott Swist, and Hema Sudarsanam
Additional Contributors: Nirmala Sundararajan

February 2014

A Dell Reference Architecture

Revisions

Date	Description
February 2014	Initial release

Table of Contents

Executive Summary

Ever increasing digital data has created a need for complete, scalable, highly available, customizable, affordable data warehouse and business intelligence (DW/BI) solution. This need is fulfilled by the converged Dell BI VRTX solution, which combines the power of the Dell™ PowerEdge™ VRTX system with the Toad Business Intelligence Suite of products for data analysis and reporting.

The BI VRTX solution enables organizations to focus more on getting value from their data while minimizing complexity of the installation of the infrastructure required for analysis. It offers a vertical solution opportunity to meet the IT infrastructure performance needs of a small to midsize business in a single unit.

The key features of the BI VRTX solution are:

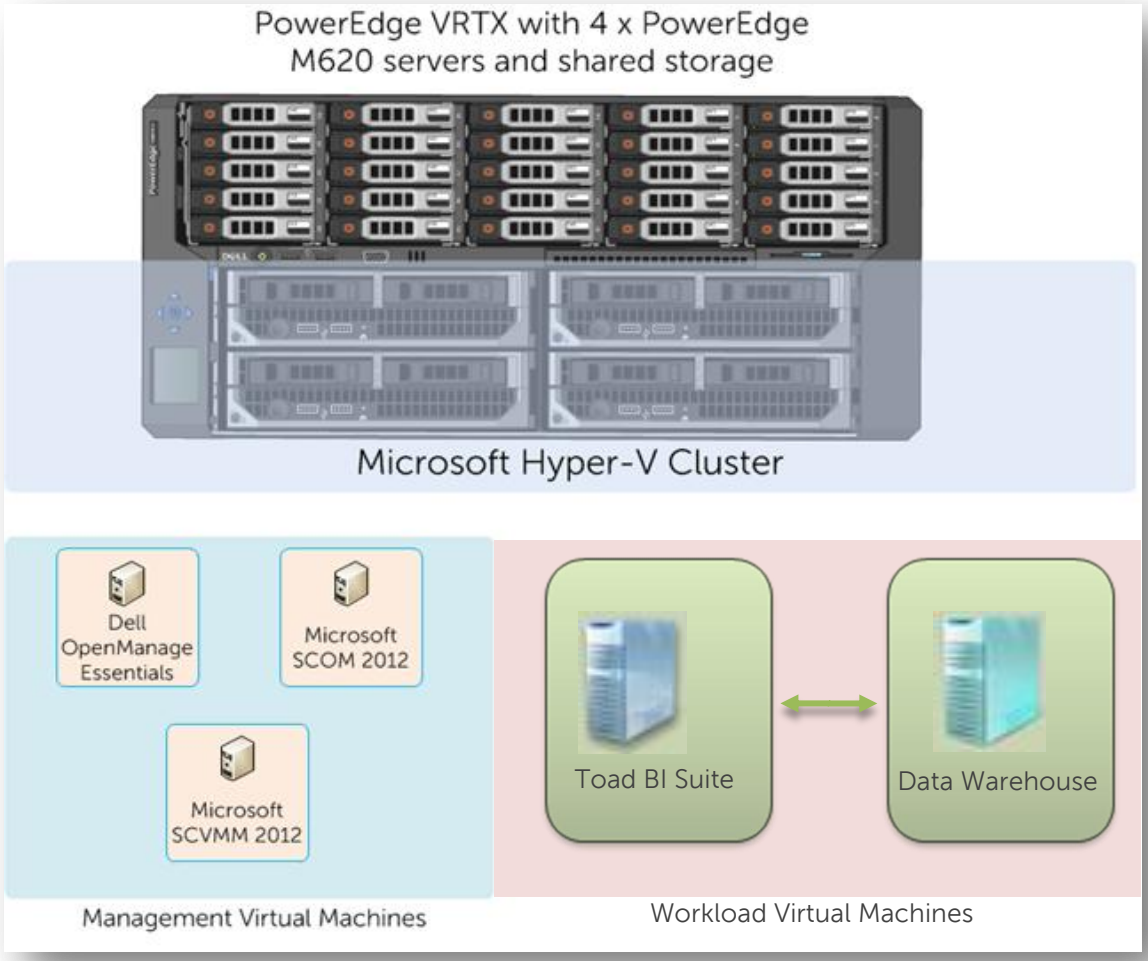
- **Ease of Deployment**: The BI VRTX solution ships with the optional Dell™ OpenManage™ Cluster Configurator (OMCC). OMCC simplifies the Hyper-V cluster configuration steps for networking, shared storage and the server nodes. It enables you to create a 2 or 4 node Microsoft® Hyper-V cluster in less than an hour.
- **Scalability**: Availability of multi-blade platform with either Dell PowerEdge M520 or Dell PowerEdge M620 blade servers. The solution offers an "in-box" raw storage capacity of up to 48TB.
- **DW/BI Integration**: With the optional preinstalled operating system, Dell also provides a preinstalled Virtual Machine (VM) to help you start deploying DW/BI applications quickly. The VM can be used to deploy data warehouse RDBMS or Toad BI Suite analytical software. For the data warehouse, the database engine is open for your preference. Engineering has tested various vendors such as, Microsoft SQL Server, MySQL and Oracle. For more information, see **Table 3-1: Tested and Validated Databases for the Data Warehousing/BI VRTX Solution**.

1 Introduction

The BI VRTX design described in this white paper leverages the solution detailed in Dell VRTX Reference Architecture Virtualization Solution using Microsoft Hyper-V 2012 . The goal of this architecture is to provide an efficient small and medium business (SMB) infrastructure solution. This reference architecture reduces the management and support overheads while retaining the virtualized and shared infrastructure best practices that benefit enterprise class Business Intelligence and data applications. The VRTX reference architecture also intends to provide comprehensive design information on the core infrastructure services that are based on Microsoft® Windows Server® 2012.

The figure below shows a representation of management and workload VMs.

Figure 1-1: VRTX Reference Architecture



2 Dell Data Warehouse Reference Architecture Offerings

Dell provides various data warehouse reference architectures configuration options using the Dell PowerEdge VRTX system. Depending on your requirement for the type of blade and the number of blades in the chassis, the table below summarizes the pre-configured bundles with or without the optional OMCC software and the blade model options.

Table 2-1: Dell Reference Architectures for VRTX Servers

Config ID	Servers
11287703	Two M620 Blades with OMCC
11719974	Two M620 Blades
11848239	Two M520 Blades with OMCC
11848248	Two M520 Blades

;

2.1 BI VRTX Architecture

The BI VRTX configuration is offered in tower mount and rack mount option for office or data center installations.

Figure 2-1: Dell VRTX Form Factor

The VRTX platform supports the Dell PowerEdge M520 blades and M620 blades. The table below summarizes the base configurations for each blade model. The blade components can be modified to meet different workload demands.

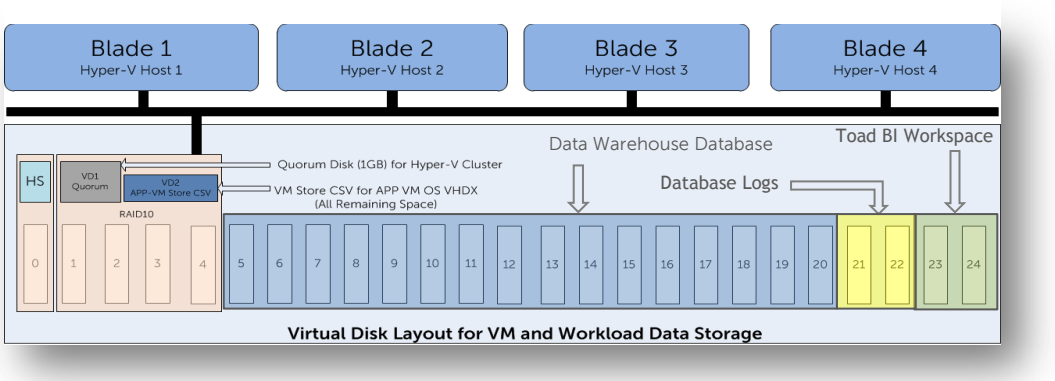
Table 2-2: Dell Reference Architecture Component Details for VRTX Servers

Component	Details	
Blade Servers	M520	M620
CPU per blade	Two Intel® Xeon® E5-2407 v2 @ 2.5GHz	Two Intel® Xeon® E5-2637 v2 @ 3.5GHz
Number of cores per socket	4 core	4 core
Memory per blade	12 x 16GB 1600MHz	12 x 16GB 1866MHz
Blade Internal Drives	2 x 300GB 10K 2.5" SAS	2 x 300GB 10K 2.5" SAS
Hard Drives	25 x 600GB 10K 2.5" SAS	25 x 900GB 10K 2.5" SAS
Operating System	Microsoft Windows 2012 Standard	Microsoft Windows 2012 Standard
VM OS	Microsoft Windows 2012	Microsoft Windows 2012

The following figure depicts a scenario with maximum data warehouse storage allocation. Depending on the size of the data warehouse and other VM storage requirements, you can adjust the disk space to ensure a balanced distribution of resources. For environments that require additional storage, install PCIe cards to access an array of storage solutions.

Contact your Dell representative for additional storage requirements.

Figure 2-2: Storage Layout for Dell BI/Data Warehouse VRTX Solution



Both the tower and rack configurations utilize the 2.5" chassis with up to 25 hard drives. More hard drive spindles allow the workload to be spread. You can choose between various RAID settings.

For write intensive applications, it is recommended to use RAID 1 or RAID 10 configuration. In RAID 1, the storage can be configured with a set of eight RAID 1 disk groups which provide raw storage capacity of 6.5TB. For a higher capacity configuration that does not incur a high number of writes, RAID 5 can be a suitable solution with 1.2TB 10K hard drives. For a large data warehouse application, 16 disks in RAID 5 configuration with raw storage of 16.3TB is a good option. Depending on the RDBMS used and the structure of the data, compression may yield higher data set capacity.

The following table describes two optional configurations for data warehouse using 900GB hard drives and 1.2TB hard drives in nine RAID 1 disk groups or one large RAID 5 disk group and the associated raw capacities.

Table 2-3: Storage Layout and Capacity for Data Warehouse

Performance Based 2.5"Chassis 10K HDDs	**Raw Max 900GB HDDs**	**Raw Max 1.2TB HDDs**
8 x R1 Pairs	6.5TB	8.7TB
1 x R5 Group	12.2TB	16.3TB

The following table describes two optional configurations for Toad BI VM using 900GB hard drives and 1.2TB hard drives in nine RAID 1 disk groups or one large RAID 5 Disk group and the associated raw capacities.

Table 2-4: Storage Layout and Capacity for Toad BI Suite

Performance Based 2.5"Chassis 10K HDDs	**Raw Max 900GB HDDs**	**Raw Max 1.2TB HDDs**
1 x R1 Pairs	900GB	1.2TB

Depending on the expected number of VMs that the system hosts in the VM store Disk Group in Virtual Disk 2 (VD2) described in **Table 2-2: Dell Reference Architecture Component Details for VRTX Servers** and the capacity of hard drives selected, you can choose to configure VD2 in a minimum of two hard drives in RAID 1 and provision the remaining disks to the data warehouse, logs or BI application virtual disks. Similarly, the data warehouse can be split in various virtual disks if another database or temporary landing space is needed.

It is recommended to have a dedicated virtual disk for the data warehouse to minimize any IO contentions with other databases or I/O intensive files.

3 Setup and Configuration of the BI VRTX Solution

The BI VRTX solution comes preinstalled with Microsoft Windows Server 2012. Additionally, you can choose configuration with Open Manage Cluster Configurator (OMCC) or configuration with Microsoft Windows Server 2012 and Hyper-V feature enabled.

To install a Hyper-V cluster manually without the OMCC option, see http://technet.microsoft.com/en-us/library/jj863389.aspx.

OMCC is a one-time initial configurator of Microsoft Failover Cluster for Hyper-V. On first boot of each server after the first logon, the OMCC wizard tool is activated to begin the process of setting up a failover cluster for Hyper-V.

For more information for OMCC installation, see http://en.community.dell.com/techcenter/systems-management/w/wiki/4848.openmanage-cluster-configurator-for-windows-server-2012.aspx.

3.1 Storage Configuration

The BI VRTX solution uses the Dell VRTX Chassis Management Controller (CMC) to create and assign virtual disks to the nodes.

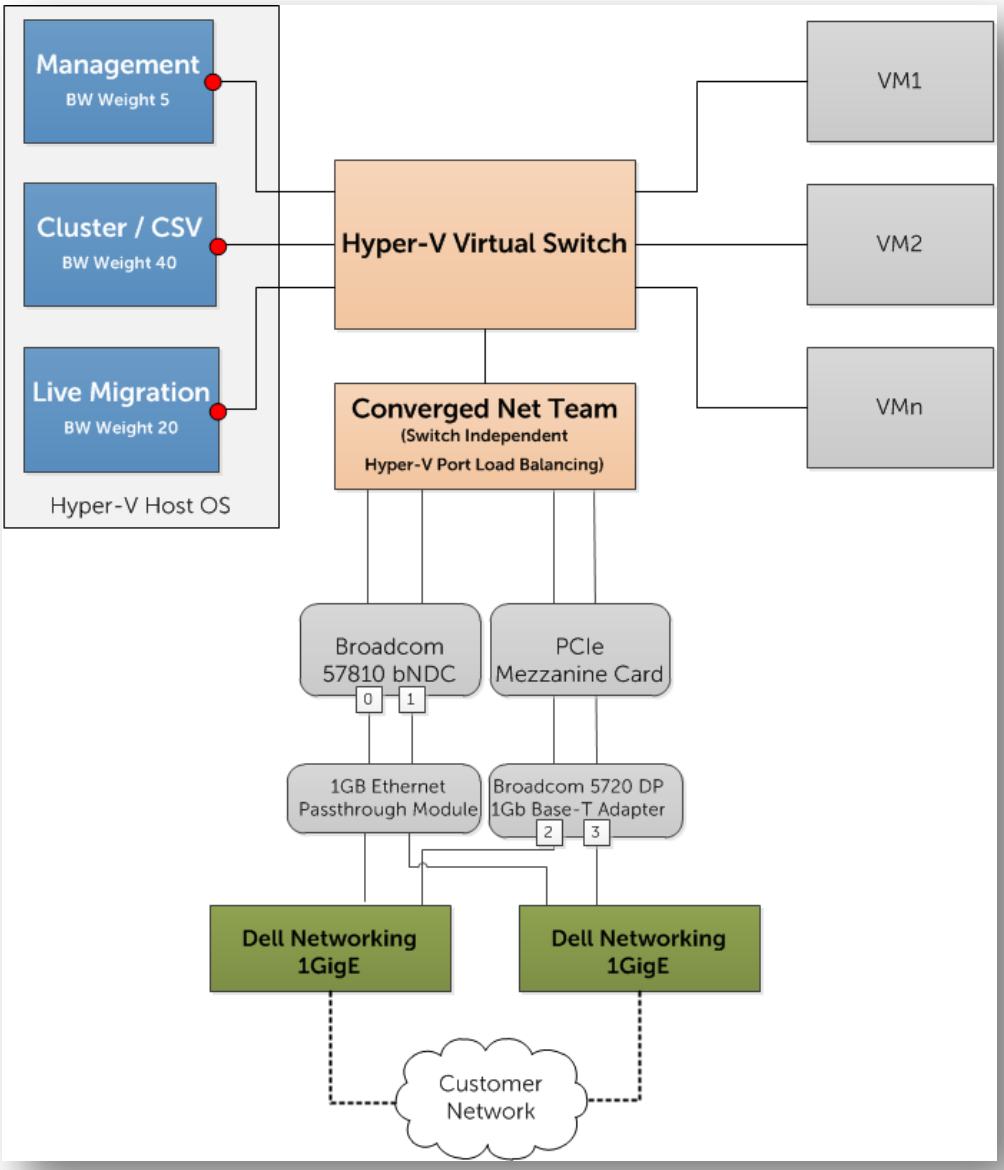
Refer to the Dell PowerEdge VRTX Shared Infrastructure Platform link to configure your storage as shown in **Dell Reference Architecture Component Details for VRTX Servers**.

3.2 Network Architecture

The Dell PowerEdge VRTX enclosure offers two options for the network fabric: a pass-through module and an Ethernet switch module. To build high availability, the host ports are mapped to redundant Dell Networking 1GigE top of the rack switches to support management, live migration, and cluster interconnects traffic. The traffic types are logically separated using VLANs and QoS settings. The two switches are lagged together through an inter-switch-link (ISL), which provides a 20 GB bandwidth between the two switches. The solution provides four 1 GigE uplinks from each switch to link into an existing core network infrastructure.

Each Dell PowerEdge M620 server is configured with a Broadcom BCM57810 blade Network Daughter Card (NDC) providing two 1GbE ports (as the Network Fabric A bandwidth supports only 1 GB network). There is another PCIe Network Broadcom BCM 5720 Dual Port adapter configured to provide additional bandwidth and high availability for the Hyper-V cluster. All the four network ports (two ports from the Network pass through switch and two from the PCIe adapter) are connected to the redundant Dell Networking switches placed outside of the PowerEdge VRTX enclosure as shown in **Error! Reference source not found.**.

Figure 3-1: Logical Network Architecture for PowerEdge VRTX



On the Microsoft Hyper-V host, a converged network design using Microsoft NIC teaming is used to provide network connectivity to the application VMs.

For more details on the configuration of this Converged Network design, see the Remote office Infrastructure - Reference Architecture on Dell PowerEdge VRTX using Microsoft Hyper-V 2012.

3.3 Virtual Machines

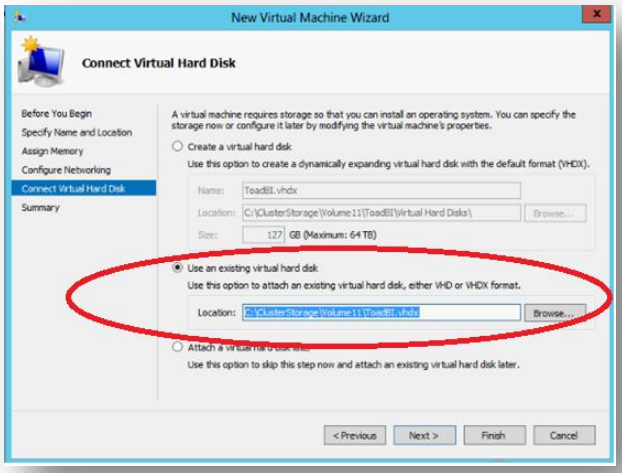
As part of the VRTX BI configuration, Dell provides one virtual hard disk file (VHDx) with a pre-installed operating system within the VM. This can be used to install the Toad BI Suite or the data warehouse RDBMS software. Additional VMs have to be manually configured and deployed as and when required.

This VHDx file is located at `C:\Dell_OEM\VHD\` directory on either of the nodes. In case of cluster implementation, this file should be copied to shared storage dedicated to host the VM files for the cluster located at `C:\ClusterStorage\` directory.

To activate the Dell provided VMs, use the virtual product key on the certificate of authenticity (COA) sticker affixed on the system. You can perform security updates using standard methods before placing the system into production.

To deploy the VM use the **New Virtual Machine Wizard** which can be accessed through the Failover Cluster Manager or the Hyper-V Manager and point to the VHDx file provided by Dell.

Figure 3-2: Deploying the VM



3.3.1 Toad BI Suite Virtual Machine

Toad BI Suite of products complements the existing IT and BI architectures and facilitates data access, data integration, and data analysis tasks. It enables sophisticated and or vertical data analysis. The users can extract insight using simple reporting and dash boarding.

Using Toad Intelligence Central, data from disparate data sources can be moved into the data warehouse. The data is scrubbed, transformed and then used for analysis. Toad Intelligence Central enables users to share and view the data.

Using Toad Data Point, analysts can query the data and develop automation routines which simplify data analysis, data provisioning and data reporting processes.

Using Decision Point, users can access, consolidate, prepare, browse and visualize data from nearly any source.

- Click on the link to download the trial version of Toad BI Suite.
- To setup Toad BI Suite and to connect it to the data warehouse, refer to the white paper Business Intelligence on Dell Quickstart Data Warehouse Appliance Using Toad Business Intelligence Suite.
- For using the Toad BI Suite and examples on building reports and analyzing patent data, refer to the white paper Analyzing Patent Data in Dell Quickstart Data Warehouse Appliance Using Toad Business Intelligence Suite.

3.3.2 RDBMS Virtual Machine

The converged BI VRTX solution is flexible and functional for all environments. The database engine may be SQL Server 2012, MySQL or Oracle, based on your preference. The table below describes the database configuration tested and validated for the data warehouse setup. It also includes the features that need to be enabled and the software that needs to be installed for different databases.

Table 3-1: Tested and Validated Databases for the Data Warehousing/BI VRTX Solution

Database	Features to be enabled on Server	Software Version	Additional Software Required to be Installed
SQL Server	.Net 3.5 and .Net 4.5	SQL Server 2012 Enterprise Edition	None
Oracle	None	Oracle 11g R2 11.2.0.3.0 Enterprise Edition	None
	None	Oracle 12c R1 12.1.0.1.0 Enterprise Edition	None
MySQL	.NET 4.5 (.NET Framework 4 Client Profile is required for Notifier 1.1.4)	Mysql 5.5.14.0 Community Server	Microsoft C++ Redistribution for MySQL workbench 6.0

To use the VRTX as part of transactional databases, refer to Microsoft SQL Server 2012 Solutions for OLTP Applications on Dell PowerEdge VRTX.

A References

A.1 TOAD BI Documentation

- TOAD Decision Point
  http://software.dell.com/products/toad-decision-point/

- TOAD Data Point
  http://software.dell.com/products/toad-data-point/

A.2 Dell PowerEdge VRTX and Microsoft Documentation

- Open Manage Cluster Configurator for Windows Server 2012.
  http://en.community.dell.com/techcenter/systems-management/w/wiki/4848.openmanage-cluster-configurator-for-windows-server-2012.aspx

- Dell PowerEdge VRTX Shared Infrastructure Platform
  http://en.community.dell.com/techcenter/extras/w/wiki/4744.dell-poweredge-vrtx-shared-infrastructure-platform.aspx

- Reference Architecture on Dell PowerEdge VRTX using Microsoft Hyper-V 2012
  http://en.community.dell.com/techcenter/extras/m/white_papers/20395858.aspx

- Microsoft SQL Server 2012 Solutions for OLTP Applications on Dell PowerEdge VRTX
  http://www.dell.com/learn/us/en/04/business~solutions~whitepapers~en/documents~sql-server-vrtx-reference-architecture.pdf

- Dell Support
  www.dell.com\support

- Microsoft – Deploy a Hyper-V Cluster
  http://technet.microsoft.com/en-us/library/jj863389.aspx

A.3 Blogs

- Analyzing Patent Data in Dell Quickstart Data Warehouse Appliance Using Toad Business Intelligence Suite
  http://en.community.dell.com/techcenter/enterprise-solutions/m/bi_converge_appliance_media/20436696/download.aspx

- Business Intelligence on Dell Quickstart Data Warehouse Appliance Using Toad Business Intelligence Suite
  http://en.community.dell.com/techcenter/enterprise-solutions/m/bi_converge_appliance_media/20347856/download.aspx