Chapter No. 1
"Installing Vertica"
In this package, you will find:

A Biography of the author of the book
A preview chapter from the book, Chapter NO.1 "Installing Vertica"
A synopsis of the book’s content
Information on where to buy this book

About the Author

**Rishabh Agrawal** is currently working as a senior database research engineer and consultant at Impetus India. He has been tinkering with databases since 2010 and has gained expertise in a variety of NoSQL, massively parallel processing (MPP), and relational databases in the Big Data domain in a short span of time. A MongoDB Certified DBA, he has working knowledge of more than 20 databases, namely Cassandra, Oracle NoSQL, FoundationDB, Riak, Gemfire, Gemfire XD, HBase, Hive, Shark, HP Vertica, Greenplum, SQL Server 2008 R2, and so on. His primary focus areas are research and evaluation of new and cutting-edge database technologies and consulting with clients on the strategic use of diverse database technologies.

When not at work, he revels in photographing vivid subjects, playing badminton, writing poems, and dancing. You can connect with him on LinkedIn at in.linkedin.com/pub/rishabh-agrawal/15/ab4/186.

I would like to acknowledge Impetus (India) that provided me valuable time and resources for creation and completion of this book. I would also like to thank my reviewers and editors (from Packt Publishing) for making sure that this book comes closest to perfect. Last but not least, I will eternally remain indebted to my parents for keeping me inspired and being a pillar of strength in my life.

**For More Information:**

HP Vertica Essentials

Column-oriented databases have emerged as one of the leading solutions for performing analytics on a huge amount of data. Vertica, recently acquired by Hewlett Packard (HP), is one of the flag bearers in the field of distributed-column-oriented databases. Vertica can be easily classified as a Massively Parallel Processing (MPP) database, which allows you to process queries concurrently on a large number of databases. Vertica's distributed architecture not only allows fast query processing, but also a highly fault tolerant architecture. With innovative features such as projections, a replacement for indexes, and views, Vertica has emerged as one of the most sought after relational analytical database solutions.

With more and more clients pressing for Vertica as a solution, we are glad to present a small admin book on Vertica, which will empower DBAs to learn and perform the most essential administration activities on Vertica.

This book has been especially written for Vertica 6.0 (and later). It is assumed that you are a little familiar with the Vertica database. Our references for this book were our experiences with Vertica, the Vertica administration guide, and the Vertica forums. Since this book is just a snapshot of the day-to-day administration activities of Vertica, it is highly advised to look into the official Vertica guides for more information and other administration tasks.

What This Book Covers

*Chapter 1, Installing Vertica*, explains how to install Vertica.

*Chapter 2, Cluster Management*, helps you to learn how to manage a running Vertica cluster.

*Chapter 3, Monitoring Vertica*, elucidates different methods of monitoring various aspects of a Vertica cluster.

*Chapter 4, Backup and Restore*, explains how to create backups of a database and restore them.

*Chapter 5, Performance Improvement*, helps you learn some tricks of the trade to achieve good performance in Vertica.

*Chapter 6, Bulk Loading*, illustrates the working of the bulk loading utility of Vertica.

For More Information:
Massively Parallel Processing (MPP) databases are those which partition (and optionally replicate) data into multiple nodes. All meta-information regarding data distribution is stored in master nodes. When a query is issued, it is parsed and a suitable query plan is developed as per the meta-information and executed on relevant nodes (nodes that store related user data). HP offers one such MPP database called Vertica to solve pertinent issues of Big Data analytics.

Vertica differentiates itself from other MPP databases in many ways. The following are some of the key points:

- **Column-oriented architecture**: Unlike traditional databases that store data in a row-oriented format, Vertica stores its data in columnar fashion. This allows a great level of compression on data, thus freeing up a lot of disk space. (More on this is covered in Chapter 5, Performance Improvement.)

- **Design tools**: Vertica offers automated design tools that help in arranging your data more effectively and efficiently. The changes recommended by the tool not only ease pressure on the designer, but also help in achieving seamless performance. (More on this is covered in Chapter 5, Performance Improvement.)

- **Low hardware costs**: Vertica allows you to easily scale up your cluster using just commodity servers, thus reducing hardware-related costs to a certain extent.

This chapter will guide you through the installation and creation of a Vertica cluster. This chapter will also cover the installation of Vertica Management Control, which is shipped with the Vertica Enterprise edition only. It should be noted that it is possible to upgrade Vertica to a higher version but vice versa is not possible.

For More Information:
Before installing Vertica, you should bear in mind the following points:

- Only one database instance can be run per cluster of Vertica. So, if you have a three-node cluster, then all three nodes will be dedicated to one single database.
- Only one instance of Vertica is allowed to run per node/host.
- Each node requires at least 1 GB of RAM.
- Vertica can be deployed on Linux only and has the following requirements:
  - Only the root user or the user with all privileges (sudo) can run the install_vertica script. This script is very crucial for installation and will be used at many places.
  - Only ext3/ext4 filesystems are supported by Vertica.
  - Verify whether rsync is installed.
  - The time should be synchronized in all nodes/servers of a Vertica cluster; hence, it is good to check whether NTP daemon is running.

Understanding the preinstallation steps
Vertica has various preinstallation steps that are needed to be performed for the smooth running of Vertica. Some of the important ones are covered here.

Swap space
Swap space is the space on the physical disk that is used when primary memory (RAM) is full. Although swap space is used in sync with RAM, it is not a replacement for RAM. It is suggested to have 2 GB of swap space available for Vertica. Additionally, Vertica performs well when swap-space-related files and Vertica data files are configured to store on different physical disks.

Dynamic CPU frequency scaling
Dynamic CPU frequency scaling, or CPU throttling, is where the system automatically adjusts the frequency of the microprocessor dynamically. The clear advantage of this technique is that it conserves energy and reduces the heat generated. It is believed that CPU frequency scaling reduces the number of instructions a processor can issue. Additional theories state that when frequency scaling is enabled, the CPU doesn't come to full throttle promptly. Hence, it is best that dynamic CPU frequency scaling is disabled. CPU frequency scaling can be disabled from Basic Input/Output System (BIOS). Please note that different hardware might have different settings to disable CPU frequency scaling.

For More Information:
Understanding disk space requirements

It is suggested to keep a buffer of 20-30 percent of disk space per node. Vertica uses buffer space to store temporary data, which is data coming from the merge out operations, hash joins, and sorts, and data arising from managing nodes in the cluster.

Steps to install Vertica

Installing Vertica is fairly simple. With the following steps, we will try to understand a two-node cluster:

1. Download the Vertica installation package from http://my.vertica.com/ according to the Linux OS that you are going to use.
2. Now log in as root or use the sudo command.
3. After downloading the installation package, install the package using the standard command:
   - For .rpm (CentOS/RedHat) packages, the command will be:
     
     ```bash
     rpm -Uvh vertica-x.x.x-x.x.rpm
     ```
   - For .deb (Ubuntu) packages, the command will be:
     
     ```bash
     dpkg -i vertica-x.x.x-x.x.deb
     ```

Refer to the following screenshot for more details:

```
[impetus@centos64a setups]$ rpm -Uvh vertica-6.1.3-0.x86_64.RHEL5.rpm
error: can't create transaction lock on /var/lib/rpm/.rpm.lock (Permission denied)
[impetus@centos64a setups]$ sudo rpm -Uvh vertica-6.1.3-0.x86_64.RHEL5.rpm
[sudo] password for impetus:
Preparing... #00000000000000000000000000000000[ 100%]
#00000000000000000000000000000000[ 100%]

Vertica Analytic Database V6.1.3-0 successfully installed on host centos64a
```

Important Information

If you are upgrading from a previous version, you must backup your database before continuing with this install. After restarting your database, you will be unable to revert to a previous version of the software.

To download the latest Vertica documentation in zip or tar format please visit the myvertica web site.

To complete installation and configuration of the cluster, run: /opt/vertica/sbin/install_vertica

Running the Vertica package

For More Information:

4. In the previous step, we installed the package on only one machine. Note that Vertica is installed under /opt/vertica. Now, we will set up Vertica on other nodes as well. For that, run the following command on the same node:

```
/opt/vertica/sbin/install_vertica -s host_list -r rpm_package -u dba_username
```

Here, -s is the hostname/IP of all the nodes of the cluster, including the one on which Vertica is already installed. -r is the path of the Vertica package and -u is the username that we wish to create for working on Vertica. This user has sudo privileges. If prompted, provide a password for the new user. If we do not specify any username, then Vertica creates dbadmin as the user, as shown in the following example:

```
[impetus@centos64a setups]$ sudo /opt/vertica/sbin/install_vertica
-s 192.168.56.101,192.168.56.101,192.168.56.102 -r "/ilabs/setups/vertica-6.1.3-0.x86_64.RHEL5.rpm" -u dbadmin
```

Vertica Analytic Database 6.1.3-0 Installation Tool
Upgrading admintools meta data format..
scanning /opt/vertica/config/users
Starting installation tasks...
Getting system information for cluster (this may take a while)....

Enter password for impetus@192.168.56.102 (2 attempts left):
backing up admintools.conf on 192.168.56.101
Default shell on nodes:
192.168.56.101 /bin/bash
192.168.56.102 /bin/bash

Installing rpm on 1 hosts....
installing node.... 192.168.56.102

NTP service not synchronized on the hosts: ['192.168.56.101', '192.168.56.102']

Check your NTP configuration for valid NTP servers.

For More Information:
Vertica recommends that you keep the system clock synchronized using NTP or some other time synchronization mechanism to keep all hosts synchronized. Time variances can cause (inconsistent) query results when using Date/Time Functions. For instructions, see:

- http://kbase.redhat.com/faq/FAQ_43_2790.shtm

Info: the package 'pstack' is useful during troubleshooting. Vertica recommends this package is installed.

Checking/fixing OS parameters.....
Setting vm.min_free_kbytes to 37872 ...
Info! The maximum number of open file descriptors is less than 65536
Setting open filehandle limit to 65536 ...
Info! The session setting of pam_limits.so is not set in /etc/pam.d/su
Setting session of pam_limits.so in /etc/pam.d/su ...
Detected cpufreq module loaded on 192.168.56.101
Detected cpufreq module loaded on 192.168.56.102
CPU frequency scaling is enabled. This may adversely affect the performance of your database.
Vertica recommends that cpu frequency scaling be turned off or set to 'performance'

Creating/Checking Vertica DBA group

Creating/Checking Vertica DBA user

Password for dbadmin:
Installing/Repairing SSH keys for dbadmin

Creating Vertica Data Directory...

Testing N-way network test. (this may take a while)
Installing Vertica

All hosts are available
Verifying system requirements on cluster.
IP configuration
IP configuration
Testing hosts (1 of 2)....

Running Consistency Tests
LANG and TZ environment variables
Running Network Connectivity and Throughput Tests...
Waiting for 1 of 2 sites... ...

Test of host 192.168.56.101 (ok)
====================================
Enough RAM per CPUs (ok)
--------------------------------
Test of host 192.168.56.102 (ok)
====================================
Enough RAM per CPUs (FAILED)
--------------------------------
Vertica requires at least 1 GB per CPU (you have 0.71 GB/CPU)
See the Vertica Installation Guide for more information.

Consistency Test (ok)
======================
Info: The $TZ environment variable is not set on 192.168.56.101
Info: The $TZ environment variable is not set on 192.168.56.102

Updating spread configuration...
Verifying spread configuration on whole cluster.

For More Information:
Creating node node0001 definition for host 192.168.56.101
... Done
Creating node node0002 definition for host 192.168.56.102
... Done
Error Monitor 0 errors 4 warnings
Installation completed with warnings.
Installation complete.

To create a database:
1. Logout and login as dbadmin.**
2. Run /opt/vertica/bin/adminTools as dbadmin
3. Select Create Database from the Configuration Menu

** The installation modified the group privileges for dbadmin.
If you used sudo to install vertica as dbadmin, you will
need to logout and login again before the privileges
are applied.

5. After we have installed Vertica on all the desired nodes, it is time to create
da database. Log in as a new user (dbadmin in default scenarios) and connect
to the admin panel. For that, we have to run the following command:
 /opt/vertica/bin/adminTools

6. If you are connecting to admin tools for the first time, you will be prompted
for a license key. If you have the license file, then enter its path; if you want
to use the community edition, then just click on OK.

License key prompt

For More Information:
7. After the previous step, you will be asked to review and accept the End-user License Agreement (EULA).

Prompt for EULA

After reviewing and accepting the EULA, you will be presented with the main menu of the admin tools of Vertica.

Admin tools main menu

For More Information:
8. Now, to create a database, navigate to Administration Tools | Configuration Menu | Create Database.

![The Create Database option in the configuration menu](image)

9. Now, you will be asked to enter a database name and a comment that you would like to associate with the database.

![Database name and comments](image)

For More Information:
10. After entering the name and comment, you will be prompted to enter a password for this database.

![Password for the new database](image)

11. After entering and re-entering (for confirmation) the password, you need to provide pathnames where the files related to user data and catalog data will be stored.

![Catalog and data pathnames](image)
After providing all the necessary information related to the database, you will be asked to select hosts on which the database needs to be deployed. Once all the desired hosts are selected, Vertica will ask for one final check.

12. Now, Vertica will create and deploy the database.

13. Once the database is created, we can connect to it using the VSQL tool or perform admin tasks.

For More Information:
Summary
As you can see, Vertica installation is simple. You can perform further checks by creating sample tables and performing basic CRUD operations.

For a clean installation, it is recommended to serve all the minimum requirements of Vertica. It should be noted that installation of client API(s) and Vertica Management Console needs to be done separately and is not included in the basic package.

In the next chapter, you will learn some tricks relating to cluster management.
Where to buy this book


Free shipping to the US, UK, Europe and selected Asian countries. For more information, please read our shipping policy.

Alternatively, you can buy the book from Amazon, BN.com, Computer Manuals and most internet book retailers.