Deploy Big Data Extensions on vSphere Standard Edition

You can deploy Big Data Extensions 1.0 Beta on VMware® vSphere™ Standard Edition for the purpose of experimentation and proof-of-concept projects to explore the potential of running Hadoop in a virtualized environment.

Intended Audience

This document is for vSphere system administrators and developers who want to deploy and use Big Data Extensions 1.0 Beta on vSphere Standard Edition. To successfully work with Big Data Extensions, you should be familiar with Hadoop and VMware vSphere.

Deploying Big Data Extensions on vSphere Standard Edition

To deploy Big Data Extensions on vSphere Standard Edition, you must deploy the Hadoop Template virtual machine and Serengeti Management Server virtual machine separately within a vSphere datacenter object (In contrast, when deploying on vSphere Enterprise or Enterprise Plus, you deploy a single OVA file which, when deployment completes, creates both the Hadoop Template virtual machine and Serengeti Management Server virtual machine.). These instructions assume you are familiar with deploying virtual machines in vSphere using the vSphere Web Client. For information on deploying virtual machines, see the vSphere documentation.

Prerequisites

■ Configure a vSphere Datacenter for use with Big Data Extensions. For information on setting up a datacenter, see Appendix C, "vSphere Settings" in the *VMware vSphere Big Data Extensions Command-Line Interface Guide*.

> NOTE   vSphere Standard Edition does not include Distributed Resources Scheduler (DRS). You can ignore the steps to enable DRS when deploying Big Data Extensions on vSphere Standard Edition.

- Download the Big Data Extensions virtual machine templates for vSphere Standard Edition from the [VMware download site](#).
- Verify that you have at least 40 GB available for the virtual machines. You will need additional storage resources for the Hadoop cluster.

Procedure

1 Using the vSphere Web Client, create a folder in which to store the Hadoop Template virtual machine and Serengeti Management Server virtual machine under a valid vCenter Datacenter object.

2 Right-click the datacenter under which you created the folder, and select **New Virtual Machine**. Following the instructions in the **New Virtual Machine** wizard, deploy the Hadoop Template virtual machine to the folder you created in the previous step.

3 Right-click the same datacenter object as in the previous step (step 2), and select **New Virtual Machine**. Deploy the Serengeti Management Server virtual machine within the same folder in which you deployed the Hadoop Template virtual machine

When the wizard prompts you to type the name of the Hadoop Template virtual machine, use the same name you used to label the Hadoop Template virtual machine in the previous step (step 2). The Management Server uses this Hadoop Template virtual machine to deploy Hadoop clusters.

4 For information on configuring storage and networks for use with Big Data Extensions, see the topic "Deploy the Serengeti vApp" in the *VMware vSphere Big Data Extensions Command-Line Interface Guide*.

NOTE   vSphere Standard Edition does not support the use of resource pools. You can ignore the steps to choose a resource pool in which to deploy the Big Data Extensions vApp when using vSphere Standard Edition.

When deployment completes, two virtual machines are in the datacenter:

- The Serengeti Management Server virtual machine, which is started as part of the Big Data Extensions deployment.
- The Hadoop Template virtual machine, which is not started. Big Data Extensions clones Hadoop nodes from this template when provisioning a cluster. You do not need to start or stop this virtual machine.

What to do next

- If you are using vSphere 5.1 or later, you can install and use the Big Data Extensions graphical user interface to perform management and monitoring tasks. See "Install the Big Data Extensions Plug-in" in the *VMware vSphere Big Data Extensions Administrator's and User's Guide.*

- Access the Serengeti command-line interface (CLI) console to perform additional configuration tasks. See "Access the Serengeti CLI via the Remote CLI Client," in the *VMware vSphere Big Data Extensions Command-Line Interface Guide*.

- If you did not leave the **Initialize Resources** check box selected (checked), you must add a datastore and network for use by the Hadoop cluster you intend to create. See Chapter **Error! Reference source not found.** "Managing vSphere Resources for Hadoop Clusters," in the *VMware vSphere Big Data Extensions Command-Line Interface Guide*.

- Create and deploy a Hadoop cluster using your Hadoop distribution. See Chapter **Error! Reference source not found.**, "Creating Hadoop Clusters," in the *VMware vSphere Big Data Extensions Command-Line Interface Guide*.