

VMware vSphere Big Data Extensions

Extending the vSphere Platform to Support Big Data and Hadoop
www.vmware.com/bde

AT A GLANCE

VMware vSphere Big Data Extensions, or BDE, is a feature within vSphere to support Big Data and Hadoop workloads. BDE provides an integrated set of management tools to help enterprises deploy, run, and manage Hadoop on the vSphere platform. Through the vSphere vCenter user interface, enterprises are able to manage and scale Hadoop seamlessly.

KEY BENEFITS

- Achieve Operational Simplicity with Performance
- Maximize Resource Utilization on New or Existing Hardware
- Architect Scalable and Flexible Big Data Platform

What is Big Data?

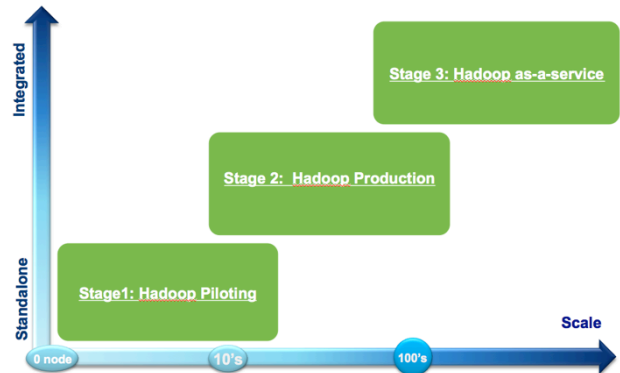
Data growth is exploding in the Enterprise. This explosion is due to a combination of market trends and technological advances. Data growth has been driven by a new generation of scale-out and real-time applications, increasing usage of mobile devices resulting in employees spending more time online, and the emergence of new application platforms and frameworks that has accelerated the time to create an application. This confluence of events has resulted in an explosion of unstructured data that is expensive to store within a relational database. These developments have caused IT departments within enterprises to look for new ways to process, analyze and store data more effectively.

Apache Hadoop

The Hadoop technology was open-sourced by Yahoo! in June 2009 to the Apache Software Foundation. Since that time, Hadoop has emerged as a powerful new technology that provides users with an inexpensive and efficient way of storing and processing large amounts of unstructured data. A rich open-source ecosystem has emerged around the technology, building adjacent technologies to extend the power of Hadoop and make it easier to run Hadoop workloads. However, enterprise adoption has been slowed by a skills gap and a lack of enterprise-grade tools to make it easier for IT departments to run the technology for business-critical workloads. Moreover, concerns persist

around security, cost and infrastructure flexibility. The result is that many enterprises are still in the early stages of testing or piloting the Hadoop technology.

Figure 1: Hadoop Adoption in the Enterprise



VMware vSphere Big Data Extensions

VMware introduced Big Data Extensions, or BDE, as a commercially supported version of Project Serengeti designed for enterprises seeking VMware support. BDE enables customers to run clustered, scale-out Hadoop applications on the vSphere platform, delivering all the benefits of virtualization to Hadoop users. BDE delivers operational simplicity with an easy-to-use interface, improved utilization through compute elasticity, and a scalable and flexible Big Data platform to satisfy changing business requirements. VMware has built BDE to support all major Hadoop distributions and associated Apache Hadoop projects such as Pig, Hive, and HBase.

How it works

BDE is a downloadable virtual appliance integrated as a plug-in to vCenter server. BDE requires a vSphere 5.0 or later license and an Enterprise or Enterprise Plus license. The Serengeti virtual appliance runs on top of vSphere and includes two virtual machines: Serengeti Management Server and the Hadoop Template Server. The Serengeti Management Server handles creation of the cluster, including creation and configuration of the virtual machines and assignment of Master node and Slave node roles. Once the cluster is created, the Serengeti Management Server then clones the Hadoop template to create and scale out the cluster. Once this is complete, the Serengeti Management Server starts Hadoop. BDE is managed and controlled through vCenter server.



VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com

Copyright © 2013 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at <http://www.vmware.com/go/patents>. VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

VMware vSphere Big Data Extensions

Extending the vSphere Platform to Support Big Data and Hadoop
www.vmware.com/bde

Features

vCenter integration to rapidly deploy Hadoop clusters

- Deploy clusters with HDFS, MapReduce, HBase, Pig, and Hive Server with flexible cluster design
- Automate the deployment and scaling of Hadoop clusters

Hadoop-as-a-Service

- Enable self-service provision of Hadoop clusters in the private cloud with vCloud Automation Center
- Remove dependency and potential bottleneck associated with IT infrastructure management

Management Tool Integration

- Use BDE to manage infrastructure and leverage Hadoop distributor management tools to perform Hadoop installation and subsequent monitoring
- Integration with Cloudera Manager and Hortonworks Ambari

Elastic Scaling and True Multi-tenancy

- Elastically scale compute and data separately
- Preserve data locality to improve performance

Architectural Flexibility

- Gain platform flexibility with support from major Hadoop distributions
- Select from hybrid, local storage, and shared storage options

Benefits

Achieve Operational Simplicity with Performance

Automated deployment and management – Eliminate manual deployment processes with intelligent automation and proactively monitor the health of Hadoop clusters.

Self-service provisioning – Self-service provision capabilities empower users and remove IT management bottlenecks.

No degradation in performance – Hadoop TeraSort Benchmark tests show comparable performance results of physical versus virtualized configuration.

Maximize Resource Utilization on New or Existing Hardware

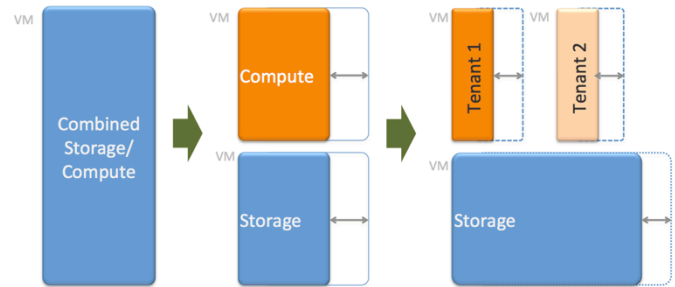
Elastic Scaling – Separating data from compute allows for independent scaling of the compute layer while keeping data persistent and safe. Commission/decommission compute

nodes according to priority and available resources.

True Multi-tenancy – Deploy separate compute clusters for different tenants sharing HDFS. Users can run mixed workloads simultaneously on a single physical host.

VM-based isolation – Achieve resource, version, and security isolation, protecting data and utilization requirements.

Figure 2: Separating data and compute improves resource utilization



Architect Scalable and Flexible Big Data Platform

Choice of Hadoop distribution – BDE works in conjunction with multiple distributions: Cloudera, Pivotal, Hortonworks, MapR, and Apache Hadoop. Cloudera Manager or Hortonworks Ambari can perform installation and monitoring. BDE also supports Hadoop 2.0 and YARN.

Choice of Storage Options – IT departments can design a Big Data platform that leverages trusted and proven external storage.

Scale Effectively – Scale with confidence to meet changing business requirements.

How Can I Purchase Big Data Extensions?

VMware vSphere Big Data Extensions is packaged as a feature in VMware vSphere 5.5 Enterprise and Enterprise Plus editions. Customers with current entitlements to vSphere Enterprise and Enterprise Plus licenses will receive commercial support for BDE. BDE is not sold as a separate product.

For information or to purchase VMware products, call 1-877-4VMWARE (outside of North America dial +1-650-427-5000), visit <http://www.vmware.com/bde> or search online for an authorized reseller.



VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com

Copyright © 2013 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at <http://www.vmware.com/go/patents>. VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.