DELL EMC

OpenStack and Hadoop
Achieving near bare-metal performance for big data workloads running in a private cloud

ABSTRACT

IT organizations are increasingly turning to the open source Apache™ Hadoop® software framework for distributed processing of big data. Hadoop is typically deployed on dedicated physical servers in on-premises data centers. While Hadoop can be deployed in scalable, cost-effective private clouds, many IT organizations have not taken this step because of concerns about performance limitations and deployment complexity. Today, Dell EMC is changing this equation. Dell EMC streamlines the deployment of Hadoop in OpenStack clouds with Ready Bundles that are optimized to deliver near bare-metal performance for big data workloads running in a cloud environment.

October 2017

DELL EMC | intel

TABLE OF CONTENTS

FINDING THE RIGHT CLOUD SOLUTION FOR A HADOOP DEPLOYMENT

In enterprise environments, the benefits of private clouds are now well understood — from IT agility and scalability to better utilization of resources. Today, some innovative IT organizations are looking to bring these and other cloud benefits to their big data workloads by deploying the Apache Hadoop software framework in an OpenStack cloud environment.

For these organizations, the initial challenge is to find the right private cloud solution. Some organizations have tried to get to the cloud quickly by deploying appliances that package hardware and software in pre-integrated systems. The outcomes of these efforts have been mixed. Many organizations have found that their out-of-the box cloud environments lack the flexibility and extensibility needed for a true enterprise-class cloud.

Homegrown cloud solutions, on the other hand, offer flexibility and extensibility but bring their own set of challenges. These solutions, which are typically based on the OpenStack platform, can be difficult to deploy due to the interdependency of software components from multiple open source projects, the rapid evolution of the OpenStack platform and the need for special skillsets.

When building an OpenStack cloud from scratch, integrations with hardware, cloud management software, services and enterprise IT systems are needed to package the platform into a solution that delivers real results to the enterprise. This isn't an easy proposition. As just about anyone in the OpenStack user community can tell you, this work requires significant knowledge and substantial investment, and it can lead to a long-running IT initiative before an environment goes into production.

Today, Dell EMC, Intel and Red Hat offer a better way forward — via a robust Ready Bundle for an agile OpenStack cloud built for data-intensive enterprise-scale workloads. This solution gives you the best of two worlds: the confidence that comes with a proven architecture along with the flexibility and extensibility of a custom cloud solution.

THE DELL EMC READY BUNDLE FOR RED HAT OPENSTACK PLATFORM

A VALIDATED AND TESTED CLOUD STACK

The Dell EMC Ready Bundle for Red Hat OpenStack Platform streamlines your path to a private cloud. It combines a core validated architecture with select extensions to create an adaptive infrastructure that integrates innovations from the OpenStack community with proven technologies from Dell EMC, Intel and Red Hat.

This open, secure, reliable and supported solution is built on validated, integrated components, including Dell EMC™ PowerEdge™ servers with Intel® Xeon® processors. The solution features rapid, automated provisioning capabilities to simplify and speed deployment — so you can take advantage of cloud benefits sooner. Better still, you can deploy your solution with the confidence that comes with knowing that Dell EMC and Red Hat have put all configurations and deployment options through extensive validation and stress testing.

With the Ready Bundle for Red Hat OpenStack Platform, your system administrators can get your private cloud up and running quickly and keep it running reliably — with all the ease and elegance of a fully featured, enterprise-grade cloud manager.

TOOLS FOR RAPID DEPLOYMENT AND MORE

One of the hallmarks of the Dell EMC Ready Bundle for Red Hat OpenStack Platform is a rapid and automated deployment experience. To enable this experience, the Ready Bundle is available with a powerful automation toolkit called JetPack. This toolkit provides a single package for configuring the underlying hardware and the OpenStack software in a fully automated fashion.

JetPack includes a large set of automation tools that are used in conjunction with Red Hat OSP Director and OpenStack Ironic bare metal provisioning to provide a completely automated software deployment experience, as well as an automated update and upgrade experience. Whether you choose to have Dell EMC experts deploy your system or to do it yourself, JetPack gives you access to the same tools and validated deployment guides and the same rapid, reliable and automated processes.

In addition, when you work with Dell EMC, our engineers help you leverage the OpenStack integration test suite, known as Tempest. This software suite includes tests for OpenStack API validation, scenarios and other specific tests for validating an OpenStack deployment. The use of the Tempest test suite gives you the assurance that your OpenStack solution is ready to go into production.

RUNNING HADOOP IN AN OPENSTACK CLOUD

While the benefits of the Dell EMC Ready Bundle for Red Hat OpenStack Platform are compelling, the goal is never better infrastructure. It's better performance for workloads. When you move applications to a private cloud, your business units expect to enjoy the same level of performance they get when their workloads run on dedicated hardware.

For organizations deploying Hadoop in an OpenStack cloud, Dell EMC provides optimized configurations based on proven architectures. These configurations lessen the need to tune your hardware and cloud software to enable the cloud environment to best meet the demands of your Hadoop workloads.

With these optimized configurations for Hadoop workloads running in a Red Hat OpenStack cloud, your organization can achieve near bare-metal performance for Hadoop in a cloud environment — an outcome proven in Dell EMC testing.

PROVEN PERFORMANCE

At the Barcelona OpenStack Summit in October 2016, Dell EMC engineers presented lab testing results showing that Hadoop performance on an OpenStack cloud can almost match that of Hadoop on dedicated hardware. This level of performance was achieved in a proof of concept that leveraged an optimized Red Hat OpenStack configuration with a Hadoop distribution from Cloudera®.

This configuration included:

• Leading-edge hardware, including Dell EMC servers with Intel® Xeon® processors

• A select version of the OpenStack Nova project for compute provisioning

• Use of OpenStack block storage (Cinder) volumes using disk pinning on local storage

• Use of NUMA (Non-Uniform Memory Access) memory nodes

To view the full presentation, see Hadoop on OpenStack Cloud: The Elephant Can Fly.

READY BUNDLES FOR HADOOP

For organizations looking to accelerate the deployment of a Hadoop environment running on dedicated hardware, Dell EMC has an ideal solution: Ready Bundles for Hadoop. These solutions, developed jointly with leading Hadoop distributions, are based on best practices gleaned from Dell EMC's work on Hadoop installations at customer sites around the world. They include all the essential hardware, software, resources and services you need to deploy and manage a robust Hadoop production environment.

KEY TAKEAWAYS

Dell EMC offers everything you need to bring the combined power of OpenStack and Hadoop to your data center — including Intel-based Dell EMC PowerEdge servers, Red Hat OpenStack software and leading distributions of Hadoop. With this combination, you can gain the best of all worlds — the agility, scalability and economies of cloud along with near bare-metal performance for your big data workloads running in a cloud environment.

Learn more at: dell.com/openstack and dell.com/hadoop

DELL EMC   intel