One-click Hadoop Cluster Deployment on OpenPOWER Systems

Pradeep K Surisetty

IBM
#Whoami

- Systems & Infrastructure Engineer
- 9 Years + of Linux, Virtualization
- Believe in Open Source Everything
- Virtualization Test Lead/Solution Engineer
- pradeepkumars@in.ibm.com

This is a team work:
Core Team:
• Pradipta Kumar, Pradeep K Surisetty, Ashish Kumar, Yogananth Subramaniyan, Poornima Nayak, Sudeesh John

Acknowledgements:
• Dipankar Sarma, Vaidyanathan Srinivasan, Tarundeep S Kalra, Anbazhagan Mani, Ashish Bille, Akash Gunjal

Join the conversation at #OpenPOWERSummit
Elastic Hadoop on OpenPower Systems

Goal

- Make Deployment & Operation of Hadoop Clusters simple on OpenPower Systems
- Managed by OpenStack.
- Run Hadoop Performance Benchmarks on this cluster.

Key characteristics

- Opensource Hadoop
- OpenStack Native
- Example OpenStack Sahara based Elastic Hadoop on OpenPower Servers
- Benchmark Results
Intro on OpenStack and Sahara

OpenStack

OpenStack core components

- Compute - Nova
- Networking - Neutron
- Object Storage – Swift
- Block Storage – Cinder
- Dashboard - Horizon
- Identity Service - Keystone
- Image Service - Glance

Sahara Project

Sahara project is an initiative to provision Hadoop on top of OpenStack (started by Mirantis, Hortonworks and Red Hat)

Join the conversation at #OpenPOWERSummit
High Level Architecture Overview

Compute + Storage Node
- Nova-compute
- Cinder-volume

Controller Node
- Nova-api
- Cinder-scheduler
- Cinder-api
- Glance
- Neutron
- Horizon
- Sahara

Join the conversation at #OpenPOWERSummit
# Test Environment Details

## Hypervisor

<table>
<thead>
<tr>
<th>Version</th>
<th>PowerKVM-2.1.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kernel</td>
<td>3.10.42 -2015.1.pkvm2_1_1.40.</td>
</tr>
</tbody>
</table>

## VM

<table>
<thead>
<tr>
<th>OS</th>
<th>RHEL7 PPC64</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kernel</td>
<td>3.10.0-123.el7</td>
</tr>
<tr>
<td>VCPU</td>
<td>8</td>
</tr>
<tr>
<td>Memory</td>
<td>40G</td>
</tr>
</tbody>
</table>

## OpenStack

<table>
<thead>
<tr>
<th>Version</th>
<th>Juno</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sahara</td>
<td>Upstream</td>
</tr>
<tr>
<td>Diskimage-builder</td>
<td>Upstream</td>
</tr>
</tbody>
</table>
1. Setup OpenStack Controller with Sahara plugin

2. Add Power/KVM compute nodes to OpenStack controller

3. Create Power arch (ppc64) images for Sahara

   sahara-image-elements/diskimage-create/diskimage-create.sh -p vanilla -v 2.4 -i fedora

4. Register Image with Sahara

5. Create Node Group Templates based on required processes in the nodes.
   - Worker Template having only Data Node
   - Master Template Having Name node, Resource Manager, Node Manager

6. Create Cluster Template as required

7. Launch Cluster based on template

8. Submit jobs to the Cluster

Demo Video: https://www.youtube.com/watch?v=JMprhJAF8FQ
Terasort for 500 GB of workload took 7000 seconds on this environment with 2 Data nodes, 1 Name node.
Upstream Contributions for PPC64 support

1. Ramdisk-image-create: Add support for vmlinux file
   
   https://review.openstack.org/#/c/149045/

2. Add support for using local PowerPC VM image
   
   https://review.openstack.org/#/c/149165/

3. Enable vm element to create PowerPC image
   
   https://review.openstack.org/#/c/153404/

Join the conversation at #OpenPOWERSummit
1. Ramdisk-image-create: Add support for vmlinux file

https://review.openstack.org/#/c/149045/

2. Add support for using local PowerPC VM image

https://review.openstack.org/#/c/149165/

3. Enable vm element to create PowerPC image

https://review.openstack.org/#/c/153404/
Reference & Demo Video

- Hadoop on PowerKVM video:
  https://www.youtube.com/watch?v=JMprhJAF8FQ

- Creating an OpenStack cloud using DevStack and Power8 Compute Nodes
  http://goo.gl/ZHYsot

- Creating Openstack cloud using IBM Cloud Manager and Power8 Compute Nodes
  http://goo.gl/3f46Lv

- Hadoop Releases:
  http://goo.gl/MOTq1x
  http://hadoop.apache.org/releases.html/
Summary

- Hadoop Deployment & Operation can be done seamlessly on OpenPower systems using OpenStack and Sahara
Post your questions here

pradeepkumars@in.ibm.com
bpradipta@in.ibm.com