CUSTOMER

Getting Started with SAP HANA Vora 1.2

Table of Contents

1   Solution Information

This guide provides general information you need to use the solution **SAP HANA Vora 1.2**. For more detailed information refer to the product installation guide.

Installed Components

The packages contain the following components, which you need to install and deploy on the cluster:

Number	Description
Vora Base	All libraries and binaries required by SAP HANA Vora
Vora Discovery	Hashicorp's Discovery Service, manages service registration
Vora Distributed Log	A distributed log manager providing persistence for the SAP HANA Vora catalog
Vora Catalog	A distributed metadata store for SAP HANA Vora
Vora V2Server	SAP HANA Vora engine
Vora Thriftserver	A gateway compatible with the Hive JDBC Driver
Vora Tools	A web-based user interface with a SQL editor and OLAP modeler

Additional Components

The solution comprises the following component versions:

Name	Release
Spark	1.5.2.2.3
Zeppelin	0.5.7
PostgreSQL	8.3.23
Ambari	2.2.0
HDP	2.3.4
Zookeeper	*3.4.6.2.3*
MapReduce2	*2.7.1.2.3*
HDFS	*2.7.1.2.3*
YARN	*2.7.1.2.3*

Ambari Users

Name	Description
admin	<master password> when accessing Ambari on your instance under the following URL: http://<instance ip>:8080/

OS Users and Groups

Name	Description
ec2-user	Amazon WS user for SSH connectivity. Authentication via instance private key
vora	<master password> as specified during instance creation in SAP Cloud appliance library
cluster_admin	Used by cluster manager tool to manage the node(s) with private key authentication used by Ambari

2 Licenses

Developer License Agreement

This solution could be used for development purpose under the perpetual Developer License Agreement. No license keys required.

3 Security Aspects

Be aware that creating your instances in the public zone of your cloud computing platform is convenient but less secure. Ensure that only port 22 (SSH) is opened when working with Linux-based solutions. In addition, we also recommend that you limit the access to your instances by defining a specific IP range in the Access Points settings, using CIDR notation. The more complex but secure alternative is to set up a virtual private cloud (VPC) with VPN access, which is described in this tutorial on SCN.

The list below describes the ports opened for the security group formed by the server components of your solution instance:

To access back-end servers on the operating system (OS) level, use the following information:

Protocol	Port	Description
SSH	22	Used for SSH connection to Linux-based servers

Additional ports are open for web-based access to various tools.

Protocol	Port	Description
HTTP	8080	Ambari Management UI
HTTP	9099	Zeppelin UI
HTTP	9225	SAP HANA Vora Web Tools UI

Be aware that the ports for "SAP HANA Vora Web Tools UI" and for "Zeppelin UI" are accesses without authorization and everyone can have access to them.

Following ports are configured but not open by default. It is recommend to restrict access to these ports only to the IP addresses from which these will be accessed:

Protocol	Port	Description
HTTP	50070	HDFS Namenode UI
HTTP	50075	HDFS Datanode UI
HTTP	8088	Yarn Resource Manager UI
HTTP	8042	Yarn Node Manager UI

4 Solution Provisioning in SAP Cloud Appliance Library

If you have a user in SAP Cloud Appliance Library, you need to meet the following prerequisites before starting to use the SAP Cloud Appliance library:

- Cloud Provider Configurations

You have a valid account in one of the cloud providers supported by SAP Cloud Appliance Library. If you already have an active cloud provider account, you can proceed directly with the next section. Otherwise, navigate to the cloud provider home page and sign up.

For more information about the supported cloud providers, see the FAQ page.

- Navigate to SAP Cloud Appliance Library

Open the SAP Cloud Appliance Library in your Web browser using the following link: https://cal.sap.com

For more information about how to use solutions in SAP Cloud Appliance Library, see the official documentation of SAP Cloud Appliance Library (choose *Support* → *Documentation* link and choose ⊞ (expand all) button to see all documents in the structure). You can also use the context help in SAP Cloud Appliance Library by choosing the *Help* panel from the right side.

5   System Configurations

Once the components are installed, additional configurations can be performed to use optimal system resources:

In order to handle the variety of workloads related with intense CPU usage, YARN has introduced a new concept called "vcores" (short for virtual cores). A vcore, is a usage share of a host CPU which YARN Node Manager allocates to use all available resources in the most efficient possible way. YARN hosts can be tuned to optimize the use of vcores by configuring the available YARN containers as the number of vcores has to be set by an administrator in yarn-site.xml on each node. The decision of how much it should be set to is driven by the type of workloads running in the cluster and the type of hardware available. The general recommendation is to set it to the number of physical cores on the node, but administrators can bump it up if they wish to run additional containers on nodes with faster CPUs. In order to enable CPU scheduling, there are some configuration properties that administrators and users need to be aware of:

- *yarn.nodemanager.resource.cpu-vcores*: Set to the appropriate number in yarn-site.xml on all the nodes. This is strictly dependent on the type of workloads running in a cluster, but the general recommendation is that admins set it to be equal to the number of physical cores on the machine.
- *yarn.scheduler.minimum-allocation-vcores*: This is the minimum allocation for every container request at the Resource Manager, in terms of virtual CPU cores. Requests lower than this won't take effect, and the specified value will get allocated the minimum.
- *yarn.scheduler.maximum-allocation-vcores*: This is the maximum allocation for every container request at the Resource Manager, in terms of virtual CPU cores. Requests higher than this won't take effect, and will get capped to this value.

"yarn.scheduler.maximum-allocation-vcores" controls the maximum vcores that any submitted job can request. "yarn.nodemanager.resource.cpu-vcores" controls how many vcores can be scheduled on a particular NodeManager instance. So "yarn.nodemanager.resource.cpu-vcores" can vary from host to host (NodeManager to NodeManager), while "yarn.scheduler.maximum-allocation-vcores" is a global property of the scheduler.

www.sap.com/contactsap