# HORTONWORKS HDP WITH ECS REFERENCE ARCHITECTURE

#### **ABSTRACT**

This white paper describes a reference architecture for using Hortonworks HDP with  ${\ensuremath{\sf EMC}}$  ECS.

August, 2015





To learn more about how EMC products, services, and solutions can help solve your business and IT challenges, <u>contact</u> your local representative or authorized reseller, visit <u>www.emc.com</u>, or explore and compare products in the <u>EMC Store</u>

Copyright © 2015 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided "as is." EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

VMware and <insert other VMware marks in alphabetical order; remove sentence if no VMware marks needed. Remove highlight and brackets> are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions. All other trademarks used herein are the property of their respective owners.

Part Number HXXXXX <required, see Part numbers below for more info>

## **TABLE OF CONTENTS**

PRODUCT VERSIONS	4
DEPLOYMENT OPTIONS	4
CLUSTER TOPOLOGY	
CLUSTER COMPONENT LIST	
REFERENCES	6

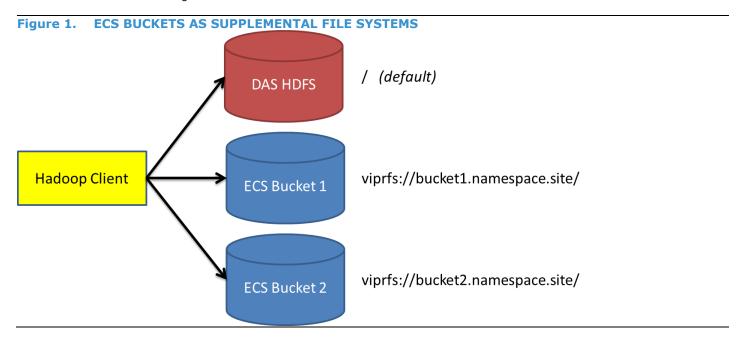
#### **PRODUCT VERSIONS**

This document applies to following product versions.

Table 1 PRODUCT VERSIONS		
PRODUCT	VERSION	
Hortonworks HDP	2.2.6.0-2800	
Ambari	2.0.1	
CentOS	6.6	
ECS	2.0.1.0-427.6d6535a	
ViPRFS Client	1.2.0.0-hadoop-2.3	

#### **DEPLOYMENT OPTIONS**

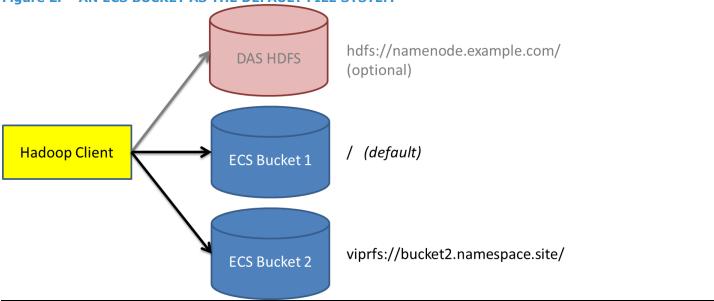
ECS buckets can be used as supplemental or auxiliary file systems. In this scenario, fs.defaultFS in core-site.xml remains as the default and references the HDFS NameNode. Applications can access data in any ECS bucket by using fully-qualified paths in the format "viprfs://bucket.namespace.site/". Jobs and applications may need to be reconfigured or updated to use a fully-qualified path. This is the recommended configuration and is what this document describes.



An ECS bucket can be used as the default file system in Hadoop, in which all Hadoop paths without an explicit URI prefix such as "hdfs://" refer to objects in a single ECS bucket. In this scenario, fs.defaultFS in core-site.xml is set to

"viprfs://bucket.namespace.site/". Data on the traditional HDFS NameNode and DataNodes will be accessible by using the fully-qualified URI such as "hdfs://namenode.example.com/". Some applications that require the Apache HDFS implementation will not work in this Hadoop cluster.

Figure 2. AN ECS BUCKET AS THE DEFAULT FILE SYSTEM



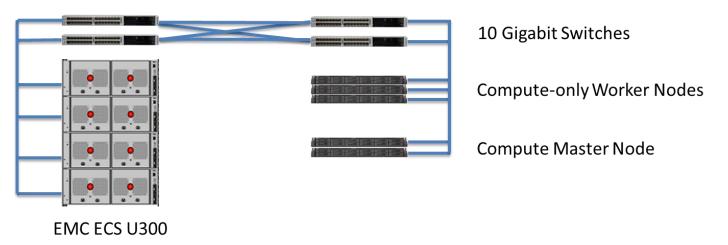
In this reference architecture document, it is assumed that any DAS HDFS usage is minimal. If DAS HDFS usage is significant, then refer to a DAS-orientied reference architecture document.

#### **CLUSTER TOPOLOGY**

The diagram below shows a simplified topology diagram for a cluster.

All ECS, worker nodes, and master nodes have dual 10-Gigabit NICs for maximum performance and redundancy.

Figure 3. Cluster Toplogy Diagram



Caption – Add caption. Delete row if not needed.

## **CLUSTER COMPONENT LIST**

**Table 2** Cluster Component List

COMPONENT	CONFIGURATION	NOTES	QUANTITY
Compute-only Worker Node		3 is the recommended quantity unless the workload is primarily in-memory, in which case it can be increased.	3
Form Factor	Rack-mount or Blade		<u> </u>
CPU	2 sockets		<del>_</del>
	6+ cores/socket		
	2+ GhZ		
Memory	256 GiB		<u> </u>
Disks (OS, applications)	2 SATA/SAS disks		<u> </u>
Disks (Intermediate data)	8-12 SATA/SAS disks or 1-2 SSDs	Consider the amount of intermediate data that may be produced during job execution. SSDs are good options for blade servers.	_
Compute Master Node			2
CPU	2 sockets		<del>_</del>
	6+ cores/socket		
	2+ GhZ		_
Memory	128 GiB		_
Disks	2 SATA/SAS disks		
Ethernet Switch for Compute	16 port 10-Gigabit Ethernet		2
Nodes			
EMC ECS U300	(4) Nodes		1
	(15) drives per node		
	(2) 10 GbE Arista 7124 switch		

Caption – Add caption. Delete row if not needed.

### **REFERENCES**

Hortonworks HDP with ECS Configuration and Best Practices Guide