



SOLUTION SHEET

HORTONWORKS DATA CLOUD FOR AMAZON WEB SERVICES

In the new world of business, enterprises embark on the journey to become truly data-driven. According to Forbes, by 2020, 1.7 megabytes of new information will be created every second for every human being on this planet, with one-third of it passing through the cloud¹. This massive data proliferation is challenging enterprises to explore new ways to effectively monitor and manage their big data strategy. Why is this important? Traditional on-premises data platforms are being challenged to capture and analyze the massive data growth.

KEY CHALLENGES

Cloud computing has been revolutionizing the IT industry by adding flexibility to the way IT resources are consumed, enabling organizations to pay only for the resources and services they use. In an effort to reduce IT capital and operational expenditures, organizations of all sizes are using the cloud to provide the resources required to run their applications. Perhaps more than any other application, Big Data processing can benefit greatly from cloud computing with its unlimited scale and on-demand access to compute and storage capacity. However, there are challenges faced by users today when deploying Apache Hadoop® and Big Data in the cloud.

- Flexibility to scale up or scale down resources for specific use cases and workloads and only pay for resources when they are used
- Automation and orchestration for spinning up resources in the cloud for reduced costs, time to market, reliability and scalability

- Self-service or managed Hadoop-as-a-Service IT infrastructure for simplified cluster management to eliminate extra time and resources devoted to managing nodes, setting up clusters and scaling infrastructure

SOLUTION OVERVIEW

Separately, Big Data and Cloud are each creating new opportunities and efficiencies for the enterprise. But when used together, enterprises can realize greater business value and achieve deeper insight more quickly and with greater flexibility than ever before. To make this combination achieve its full potential, Big Data and Cloud needs an experience that marries ease of use with infrastructure agility so that a user can get their analytics “tool of choice” in their hands exactly when they need it.

¹ Source: Forbes.com, Big Data: 20 Mind-Boggling Facts Everyone Must Read
<http://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#6c1d98166c1d>

INTRODUCING HORTONWORKS DATA CLOUD FOR AWS

Hortonworks Data Cloud for Amazon Web Services (AWS) enables you to spin-up an Apache Hadoop cluster within minutes to start modeling and analyzing your data sets immediately. Instead of manually going through configuration options, you can choose from a set of prescriptive cluster configurations (for example: Apache Spark™ for data processing or Apache Hive™ for data analytics). When you are done with your analysis, you can give the resources back to the cloud, to reduce your costs.

With Hortonworks Data Cloud for AWS, users harness the agility and elasticity of the cloud for powering new workloads and analytic applications. This allows enterprises to analyze data sets to achieve business insights with optimal cost efficiency. It is:

- **Simple:** Easily provision and manage Hortonworks platform capabilities in the public cloud
- **Flexible:** Get started quickly, deploy what you need when you need it

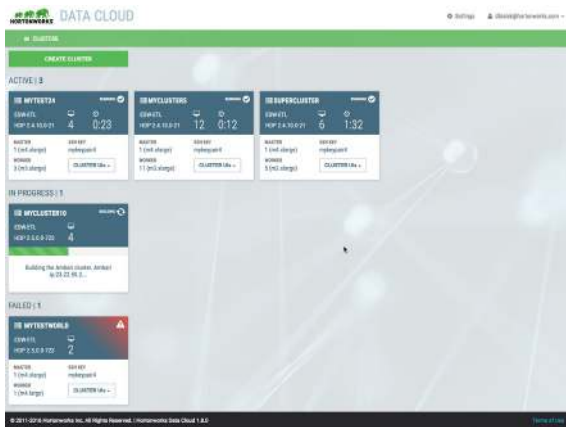


Figure 1: Cloud dashboard

KEY BENEFITS

Designed for AWS

- Delivers the most popular capabilities of Hortonworks Data Platform including Apache Hadoop, Apache Spark, and Apache Hive within an easy to use product instance available on-demand within your existing AWS account
- Optimized for and integrated with AWS services such as EC2, S3, RDS
- Available in US, EU and AP AWS regions
- Optional free community support available from Hortonworks

Focus on your Data

- Prescriptive on-demand experience for the most common use cases
 - Data Science and Exploration
 - ETL and Data Preparation
 - Data Analytics and Reporting

- **Efficient:** Pay for what you use; no upfront hardware costs
- **Enterprise Ready:** Built on the industry leading Hortonworks Data Platform (HDP®)

It addresses the challenges faced by users today.

- Provides flexibility to scale up or down compute and storage resources for specific use cases so customers only pay for what they use
- Enables automation and orchestration required to spin up resources in the cloud, making it easier and faster to get clusters up and running for specific workloads
- Delivers IT and business agility, eliminating the need for extra time and resources devoted to managing nodes, setting up clusters and scaling infrastructure



Figure 2: Cluster template

- Ease of use for developers and data scientists
 - Zeppelin notebook integrated with Spark and Hive
 - Built-in SQL query and file browser UIs
- Flexibility for operators
 - Quickly spin up workload clusters on-demand
 - Manage clusters and workloads, not servers or infrastructure

Interoperable HDP Services

- Hortonworks Data Cloud for AWS running in the Cloud and HDP running in your data center leverage matching HDP services to run your applications where it makes the most sense
- No rework or recoding required
- Provides flexibility and choice

KEY FEATURES

- Prescriptive choice of cluster types optimized and pre-tuned for ephemeral workloads around Apache Spark 2.0, Apache Hive 2 LLAP, and Apache Zeppelin
- Simple, easy to use UI and CLI
- Advanced network, security, storage configurations for power users
- Out of the box support for S3a connectivity and shared Hive Metastore

CRITICAL DIFFERENTIATORS

- Specifically optimized to run in cloud environments for enterprise ephemeral workloads
- The fastest and easiest way to put the power of Apache Hive and Apache Spark to work for you
- Includes a prescriptive set of preconfigured workload clusters that are tuned and optimized to run in Amazon Web Services

TRY NOW – 5 DAY FREE TRIAL AVAILABLE

HORTONWORKS DATA CLOUD FOR AWS

<https://aws.amazon.com/marketplace/pp/B01LXOQBOU>

CLUSTER TYPE	SERVICES
Data Science	✓ Spark 1.6 Zeppelin 0.6.0
EDW-ETL	✓ Hive 1.2.1 Spark 1.6
EDW-ETL	⚠ Hive 1.2.1 Spark 2.0
EDW-Analytics	⚠ Hive 2 LLAP Zeppelin 0.6.0

Table 1: Cluster types

The following configuration classification applies:

- ✓ Stable configurations are the best choice if you want to avoid issues and other problems with launching and using clusters
- ⚠ If you want to use a Technical Preview version of a component in a release of HDP, use these configurations

About Hortonworks

Hortonworks is an industry leading innovator that creates, distributes and supports enterprise-ready open and Connected Data Platforms and Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Hortonworks is focused on driving innovation in open source communities such as Apache Hadoop, Apache NiFi and Apache Spark. Along with its 1,800+ partners, Hortonworks provides the expertise, training and services that allow customers to unlock transformational value for their organizations across any line of business.

Contact

For further information visit
www.hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405

