RED HAT AND HORTONWORKS: OPEN MODERN DATA ARCHITECTURE FOR THE ENTERPRISE

A Hortonworks and Red Hat whitepaper

WHAT IS HADOOP?

Apache **Hadoop** is an open-source technology born out of the experience of web scale consumer companies such as Yahoo, Facebook and others, who were among the first to confront the need to store and process massive quantities of digital data.

INTRODUCTION

Big data represents an increasingly important opportunity for enterprise business leaders. According to a 2013 study by consulting firm Bain & Company[1], companies that have excelled in developing advanced analytical capabilities are twice as likely to be a top-quartile financial performer within their industries, three times more likely to execute decisions as intended, and five times more likely to make decisions faster. With these kinds of results at stake, it's no surprise that big data deployments have become a top priority for enterprise IT leaders.

When it comes to deploying big data capabilities, the starting place for many enterprises is Apache Hadoop. A popular initial application is the construction of a 'data lake' — an enterprise-wide repository for the volumes of unstructured and semi-structured data that today's businesses are amassing.

With a data lake as the centerpiece of a **modern data architecture**, enterprises are able to take advantage of the efficiency, opportunity, and insight that comes from managing and analyzing their data with a scalable and flexible data platform. But doing so is not without challenges.

Specifically, IT leaders are seeking solutions for the following:

How to integrate Hadoop into existing information infrastructure

How to build robust, data-centric applications with Hadoop

How to operationalize Hadoop and deliver high-quality service as the enterprise increases its dependence on the technology

To help meet the demands of enterprise big data initiatives, Red Hat has partnered with Hortonworks, an industry leader whose Hortonworks Data Platform (HDP) provides a tested, trusted, and open foundation for the modern data architecture.

Together, the companies provide a comprehensive solution that addresses the challenges of building big data applications, integrating information across the enterprise, and operating and supporting an enterprise-grade Hadoop environment.

facebook.com/redhatinc
@redhatnews
linkedin.com/company/red-hat

redhat.com

PROFILE: FINANCIAL SERVICES

A large financial institution took separate data warehouses from multiple departments and combined them into a geographical repository in Hadoop for analysis.

This allows the bank to construct new and more accurate scores of the risk in its customer portfolios. The more accurate scores allowed the bank to manage its exposure and offer customers better products and advice.

THE HORTONWORKS AND RED HAT BIG DATA SOLUTION

Red Hat and Hortonworks are building on a shared, open source approach that addresses the big data requirements of key enterprise stakeholders.

By tightly integrating HDP with open hybrid cloud technologies – including Red Hat® Storage, Red Hat JBoss® Middleware, Red Hat Enterprise Linux®, and Red Hat Enterprise Linux OpenStack® Platform with OpenJDK support – Hortonworks and Red Hat are delivering solutions that enable the next generation of big data applications.
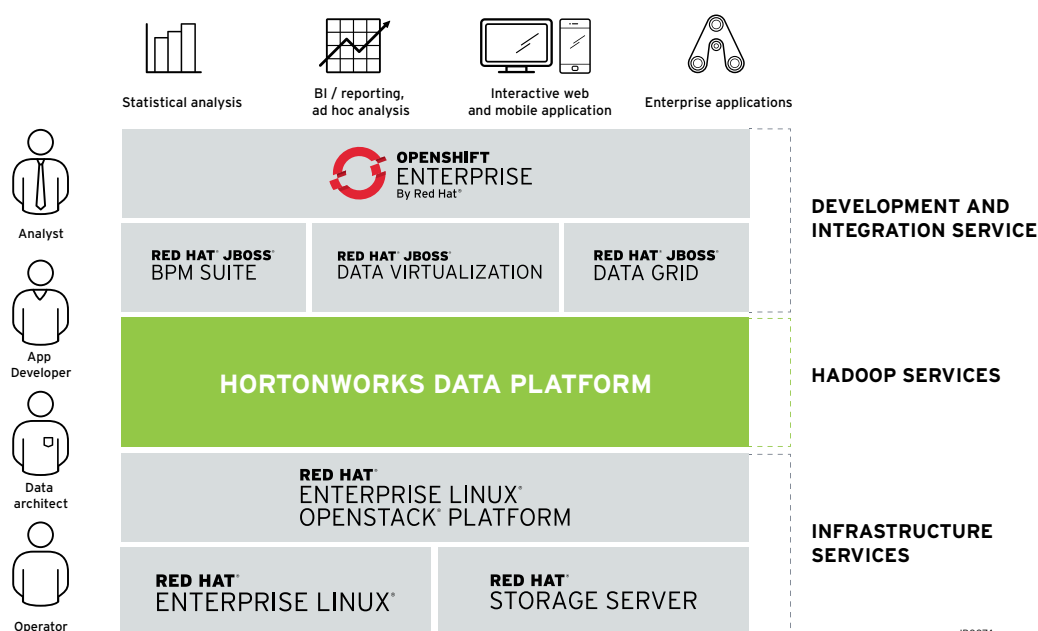
INTEGRATING HADOOP DATA WITH ENTERPRISE INFORMATION

The recent emergence and explosion of new types of data has put tremendous pressure on traditional enterprise data repositories. One technology that has emerged as a way to realize the value in big data is Apache Hadoop, with momentum described as 'unstoppable' by Forrester Research in the *Forrester WaveTM: Big Data Hadoop Solutions, Q1 2014*.

Once Hadoop is in place, the challenge for IT departments becomes how to put all that data to good use. To get the most value, unstructured data collected in Hadoop must be integrated with information spread across many applications and systems in the enterprise.

PROFILE: FOOD SERVICE

For food retailers, the fresh food category is important for customer satisfaction. Providing sufficient stock while avoiding food waste makes for happy customers and keeps the retailer profitable.

One retailer uses Hadoop to create a fully automated, data-driven replenishment process based on a variety of internal and external data sources combined with advanced predictive analytics.

Data virtualization is a straightforward approach to integrating enterprise information. It enables business leaders, data scientists, innovators, and operational decision makers to easily obtain the information they need to be effective, without worrying about the details of accessing, storing, converting, integrating, copying, moving, securing, and distributing data.

To support big data, an organization's integration technology must work with a range of underlying data sources, including conventional databases, flat files, Hadoop, in-memory computing, and other legacy and emerging technologies. It must tap into streams of data, and be able to harvest data from transactional systems, business applications, and enterprise data warehouses.
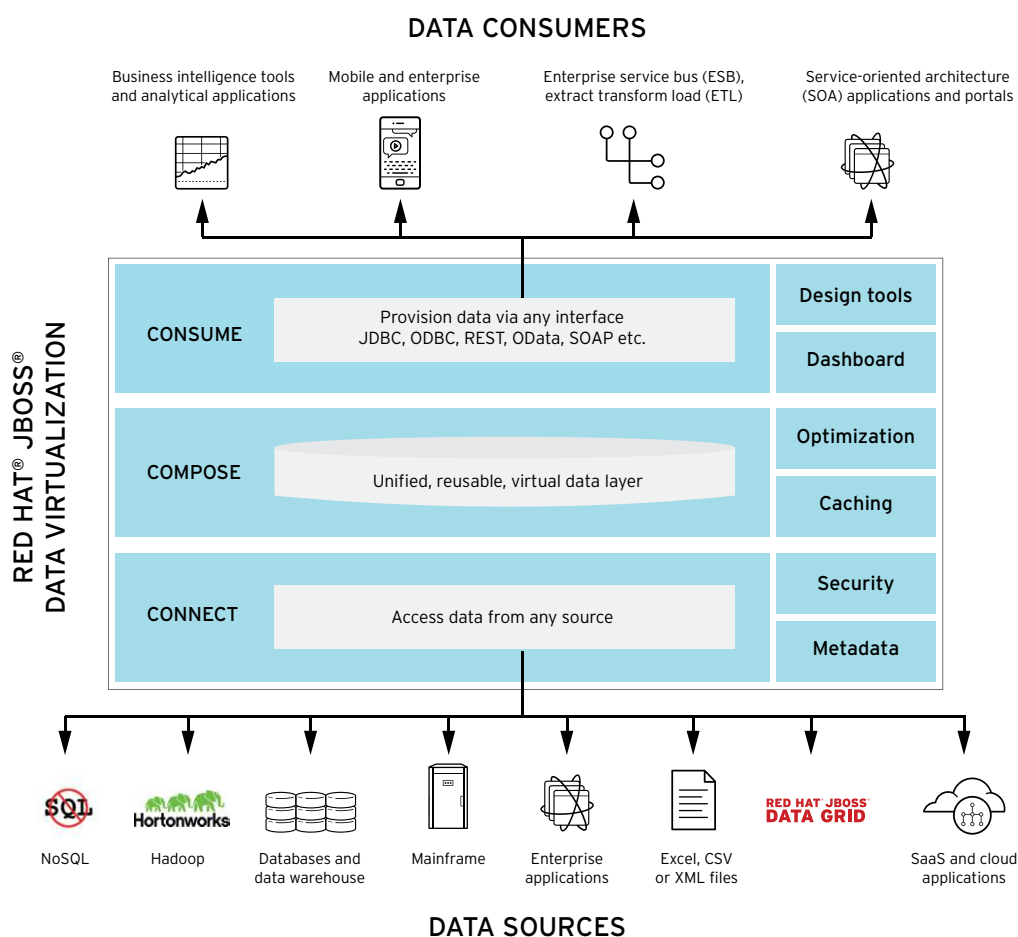
JB0041-2

Red Hat JBoss Data Virtualization meets all the requirements of big data integration. It sits in front of multiple data sources and allows them to be treated like a single source, delivering the desired data – in the required form and at the right time – to any application or user.

Through the Red Hat and Hortonworks alliance, Red Hat JBoss Data Virtualization has been integrated with Hortonworks Data Platform. Red Hat JBoss Data Virtualization facilitates and improves the use of HDP in the enterprise by:

- Abstracting Hadoop and other data sources into relational-like views.
- Integrating HDP with existing enterprise data sources.
- Adding the ability to query HDP data in real-time.
- Providing access to HDP data via web services, JDBC and ODBC, and other standard interfaces.
- Adding security and governance to the big data infrastructure.
- Flattening data silos through a unified data layer.

These capabilities allow companies deploying Red Hat JBoss Data Virtualization in conjunction with HDP to experience a significant reduction in the time and effort required to turn raw data into actionable insights.

BUILDING BIG DATA APPLICATIONS WITH HADOOP

With a strategy in place for integrating enterprise information, an organization's architects and developers must be able easily incorporate the data and capabilities of Hadoop and other systems into the applications and systems that they create.

Robust enterprise applications typically depend on the careful orchestration of a variety of middleware services. Red Hat and Hortonworks are working together to ensure interoperability between HDP and the Red Hat JBoss Middleware portfolio, including:

Red Hat JBoss BRMS

Red Hat JBoss BRMS is a comprehensive business decision automation platform that combines business rules management and complex event processing (CEP) into a single open source solution.

Today's enterprise operates in an environment of ever-increasing business, market, and regulatory complexity. Traditionally, the rules required to navigate this complexity are implicitly encoded in and scattered across the enterprise application portfolio.

Red Hat JBoss BRMS helps the enterprise:

- Increase agility by externalizing business rules so that they may be more easily managed and updated.
- Reduce time-to-market by incorporating existing rules into new applications.
- Enhance responsiveness by applying rules to operational decisions in both batch and real-time.
- Drive cohesion by allowing business, IT, and analytics professionals to work together to develop and apply analytics.

Using Red Hat JBoss BRMS in conjunction with Red Hat JBoss Data Virtualization allows IT organizations to connect enterprise decision services to Hadoop and other enterprise data sources. Seamless access to all organizational data sources—without time-consuming and costly integration and custom coding—means organizations can improve data-driven decision-making and optimize business outcomes.

Red Hat JBoss Data Grid

Red Hat JBoss Data Grid is a distributed NoSQL in-memory data grid that provides high-performance, low-latency access to big data.

As the amount of data enterprise applications must deal with grows, reads and writes to traditional back-end data stores become a major performance bottleneck. Because the cost of memory has continued to decrease, storing data in-memory is becoming mainstream.

Red Hat JBoss Data Grid provides:

- **Improved application performance** by allowing data to be quickly accessed from memory.

- **Greater flexibility** by storing different types of objects without a fixed data model, and allowing data to be access via a variety of APIs.

- **Elastic scaling** allowing grid nodes to be added simply and non-disruptively.

- **High availability,** by replicating data across multiple distributed nodes.

Using Red Hat JBoss Data Grid with Red Hat JBoss Data Virtualization, an organization can easily combine historical information from Hadoop and other sources with real-time data stored in the data grid. This data mashup gives enterprises greater visibility into operational conditions and allows them to more rapidly react to changing market conditions.

As a result of the partnership between Red Hat and Hortonworks, data from HDP is easily integrated into traditional enterprise-grade applications built around Red Hat JBoss middleware.

OPERATIONALIZING ENTERPRISE HADOOP

As enterprise big data programs mature and Hadoop becomes more deeply embedded in critical operations, the ability to support and operate it efficiently and reliably becomes increasingly important.

To aid in operating a modern data architecture at scale, Red Hat and Hortonworks have collaborated to integrate HDP with Red Hat's platform technologies. As a result of the companies' efforts:

Red Hat Enterprise Linux with OpenJDK now provides an open and flexible operating and development environment for Hadoop and enterprise analytical applications. Deploying HDP on trusted and familiar Red Hat Enterprise Linux assures long-term supportability, speedy issue remediation, and reduced operating costs.

Red Hat Enterprise Linux OpenStack Platform now delivers elastic Hadoop services on a secure, private cloud infrastructure. Deploying HDP on a private cloud helps enterprises reduce costs while providing the ability to quickly scale the Hadoop environment to meet business demands, thus improving agility.

OpenShift Enterprise by Red Hat, the company's on-premise Platform-as-a-Service (PaaS) offering, gives developers the ability to rapidly develop and deploy entire application stacks on Hortonworks Data Platform. This allows developers to standardize their workflow and create repeatable processes for application delivery, streamlining application development while tapping into the insight and value from data stored in Hadoop.

Red Hat Storage Server, an open, software-defined storage platform built on the robust GlusterFS distributed file system, can now be used with HDP to provide a secure and resilient general-purpose storage system with multiple APIs, making all data simultaneously available via Hadoop, POSIX, and web services interfaces.

Through a joint engineering effort, Red Hat and Hortonworks have created a tightly integrated, easily operable, and highly supportable platform for next-generation data applications and services.

CONCLUSION

The strategic alliance between Hortonworks and Red Hat is making it easier for enterprises to harness the power of big data.

Working together, the companies are building on a shared, community-oriented, open source approach that addresses the growing big data requirements of key enterprise stakeholders:

Integrating Hadoop into existing information infrastructure.

Building enterprise-grade, data-centric applications with Hadoop.

Operationalizing Hadoop and delivering high-quality services around it.

With an enterprise Hadoop platform tightly integrated with open hybrid cloud technologies, Hortonworks and Red Hat are delivering open, secure, cloud-ready, and cost-effective solutions that help customers more easily harness the power of Hadoop.

For more information on the Red Hat and Hortonworks partnership, visit http://hortonworks.com/ partner/redhat/. For details on the open engineering collaboration, visit the Red Hat labs page.

WHY HORTONWORKS FOR HADOOP?

Founded in 2011 by 24 engineers from the original Yahoo! Hadoop development and operations team, Hortonworks has amassed more Hadoop experience under one roof than any other organization. Hortonworks team members are active participants and leaders in Hadoop development; designing, building, and testing the core of the Hadoop platform. The company has years of experience in Hadoop operations and is best suited to support your mission-critical Hadoop project. Read more at http://hortonworks.com/why.

OPEN LEADERSHIP	ECOSYSTEM ENDORSEMENT	ENTERPRISE RIGOR
Hortonworks has a singular focus and commitment to drive innovation in the open exclusively via the Apache Software Foundation process.	Hortonworks is focused on the deep integration of Hadoop with existing data center technologies and team capabilities.	Hortonworks has a world-class enterprise support and services organization with vast experience in the largest Hadoop deployments.
Hortonworks is responsible for the majority of core code-base advances to deliver Apache Hadoop as an enterprise data platform.	Hortonworks has secured strategic relationships with trusted data center partners including Microsoft, SAP, Teradata, Rackspace, and many more.	Hortonworks engineers and certifies Apache Hadoop with the enterprise in mind, all tested with real-world rigor in the world's largest Hadoop clusters.

For an independent analysis of Hortonworks Data Platform, download the *Forrester Wave™: Big Data Hadoop Solutions, Q1 2014* from Forrester Research.

ABOUT HORTONWORKS

Hortonworks develops, distributes, and supports the only 100% open source Apache Hadoop data platform. Our team comprises the largest contingent of builders and architects within the Hadoop ecosystem who represent and lead the broader enterprise requirements within these communities. The Hortonworks Data Platform provides an open platform that deeply integrates with existing IT investments and upon which enterprises can build and deploy Hadoop-based applications. Hortonworks has deep relationships with the key strategic data center partners that enable our customers to unlock the broadest opportunities from Hadoop.

WHITEPAPER   Red Hat and Hortonworks: Open modern data architecture for the enterprise

ABOUT RED HAT

Red Hat is the world's leading provider of open source solutions, using a community-powered approach to provide reliable and high-performing cloud, virtualization, storage, Linux, and middleware technologies. Red Hat also offers award-winning support, training, and consulting services. Red Hat is an S&P company with more than 80 offices spanning the globe, empowering its customers' businesses.

NORTH AMERICA	EUROPE, MIDDLE EAST, AND AFRICA	ASIA PACIFIC	LATIN AMERICA
1 888 REDHAT1	00800 7334 2835 europe@redhat.com	+65 6490 4200 apac@redhat.com	+54 11 4329 7300 info-latam@redhat.com