



Rackspace® Managed Big Data Platform

Hadoop is hard. Rackspace makes it easy.

Table of Contents

1. Introduction	1
2. Challenges of Managing Hadoop	2
3. Managed Hadoop	3
4. Optimized Performance	5
5. Flexibility	6
6. Companies With Security and Compliance Requirements	7
7. Rackspace Managed Big Data vs. Unmanaged Hosting	8
8. Use Cases	9
9. Conclusion	11

1. Introduction

Apache™ Hadoop® is an open-source software framework that allows for distributed processing of large data sets across clusters of computers. Using simple programming models, Hadoop enables companies to derive business value from large volumes of unstructured, aggregated data. This opens up a wide range of cutting-edge use cases, such as fraud detection and social-media sentiment analysis. However, Hadoop is notoriously difficult to deploy and requires expensive, hard-to-find expertise.

The Rackspace Managed Big Data Platform is an enterprise-ready service based on the Hortonworks Data Platform (HDP) for Hadoop. The Rackspace Managed Big Data Platform offers the powerful benefits of big-data analysis without the challenges of configuring clusters, maintaining the platform or troubleshooting jobs.

Every Hadoop environment is unique, so Rackspace custom-designs each Managed Big Data Platform environment to match the precise requirements of customers' data initiatives. Solutions come with all application components installed and optimized, and Rackspace offers multiple reference architectures and storage mechanisms based on industry best practices. This helps to create an optimal fit for customer-query performance.

By combining a purpose-built platform with support for Hadoop, the Rackspace Managed Big Data Platform provides:

1. Managed Platform for Hadoop
2. Optimized Performance
3. Flexibility

The Rackspace Managed Big Data Platform is backed by Rackspace's unique Fanatical Support®. That means that our experienced Hadoop specialists are available 24x7x365 to patch, maintain and troubleshoot your environment.

The Rackspace Managed Big Data Platform is ideal for customers seeking all the benefits of big data without the hassles of managing and maintaining Hadoop clusters. The service is designed to help businesses planning to stand up Hadoop in the near future — and businesses that have already tried and struggled to stand it up successfully.

2. Challenges of Managing Hadoop

Hadoop's ability to efficiently store and process large, diverse data sets is attractive to business leaders looking to make informed, data-driven decisions. Companies that attempt to deploy Hadoop on their own, however, face several daunting challenges.

A Vicious Cycle

Hadoop is unusual: As a distributed system, it requires extensive management and offers no centralized storage. Experienced architecture experts are hard to find, and their expertise has a brief shelf life because the technology evolves so fast — each point release adds new tools and brings notable changes to the architecture. Together, these complexities create a cycle of chronic challenges. Even if businesses can find experts, they may reach the end of a months-long procurement process only to discover that the architecture of Hadoop has already changed. That sends architects back to the drawing board and leaves companies far behind schedule.

Hadoop Security Is Limited

Enterprise deployments of Hadoop have long been held back by the lack of default security solutions and the difficulty of maintaining affordable, compliance-grade protections. Open-source projects have helped improve Hadoop's out-of-the-box security, but businesses still need to add layers of protection to adequately monitor and safeguard valuable customer data.

Capital and Operating Costs Can Be Prohibitive

Implementing and managing a Hadoop-based big data solution can be extremely expensive. The hardware required to meet big data demands is a major investment for enterprises without access to large volumes of commodity machines. And then there are Hadoop's often-overlooked operational costs: Leveraging Hadoop frequently requires businesses to re-train existing staff and hire new employees to monitor and maintain the system.

Non-optimized Hardware Causes Delays

Running Hadoop on non-optimized hardware can yield slower queries, leaving end users with sluggish visibility into the data they need to make better decisions and build profitable solutions. As Hadoop evolves to address new streaming and interactive workloads, query performance variations of even a couple of seconds will make or break the end-user experience.

3. Managed Hadoop

The Rackspace Managed Big Data Platform can greatly simplify the process of implementing, optimizing and maintaining Hadoop. That allows practitioners to fully focus on deriving insights from their data; IT resources can be reallocated to initiatives that differentiate the business and build a competitive advantage.

Rackspace handles management throughout the Hadoop lifecycle of design, configuration, deployment and maintenance, so internal teams don't have to. Rackspace works directly with Hortonworks, so customers have a single point of contact. Experts from both companies maintain single-pane visibility into the infrastructure and platform/application layers, sparing customers the multi-vendor complexity that often arises when infrastructure and platform providers interact. This helps customers get the answers they need in less time. Users get the combination of infrastructure expertise and Hadoop application expertise that is critical to maintaining a healthy environment.

Rackspace stands behind every part of the service with **Fanatical Support**. Rackspace support staff and engineers are available 24x7x365 for everything from architecture design to query optimization.

Services Included With Hadoop

Deployment	<ul style="list-style-type: none"> Architecture design Installation and configuration Ingesting data to Rackspace Managed Big Data Platform Security configuration
Maintenance	<ul style="list-style-type: none"> Advanced administration, monitoring and alerting Managed patching and updates Upgrades of Hadoop versions Cluster management Job execution Diagnosis of performance issues Diagnosis of data loading, processing and query issues Application development advice Direct communication and escalation to Hortonworks Consultation, recommendation and implementation to solve scaling challenges
Optimization	<ul style="list-style-type: none"> Performance tuning and issue diagnosis Query optimization Compactions
Backups and DR	<ul style="list-style-type: none"> Rebooting of name node and master services Disaster recovery (DR) and business continuity Recommend and plan replication to DR site Participate in DNS management for failover to DR site Participate in DR RTO/RPO requirements Restores from backups

Supported Tools

Hadoop's catalogue of tools is continually expanding — and every tool requires specialized expertise. Rackspace can help businesses manage that complexity by extending **Fanatical Support** to key Hadoop tools, including:

- Apache Hadoop in HDP (HDFS, YARN, MapReduce)
- Apache Tez in HDP
- Apache Hive in HDP
- Apache Pig in HDP
- Apache Sqoop in HDP
- Apache Flume in HDP
- Apache Mahout in HDP
- Apache Ambari in HDP
- Apache Oozie in HDP
- Apache Falcon in HDP
- Apache Knox in HDP
- Apache HBase in HDP
- Apache Phoenix in HDP
- Apache Accumulo in HDP
- Apache Storm in HDP
- Apache Ranger in HDP
- Apache Spark in HDP
- Apache Kafka in HDP

4. Optimized Performance

Customizing Hadoop architectures and tuning clusters are critical to achieving fast, consistent batch processing. For example, a company may need a specialized configuration for its Hadoop clusters to scale and meet workload demands.

The Rackspace Managed Big Data Platform is a customized, single-tenant hosting solution designed to address these needs. Rackspace engineers work with customer teams to design an optimal network, compute and storage for each environment based on customer needs, including networking up to bonded 10G.

By leveraging Rackspace's infrastructure expertise and experience with the Hadoop application, companies can reap the benefits of a purpose-built, highly tuned Hadoop environment from the very beginning. And with **Fanatical Support**, customers have services available to optimize architecture for the full lifecycle of their project.

5. Flexibility

The Rackspace Managed Big Data Platform gives companies the flexibility to tailor Hadoop to their unique use cases and then adapt as needed. Companies can take advantage of hybrid deployment options, open technologies for data portability and customer support for all of their Hadoop tools — plus a range of integrations with third-party business intelligence and analytics applications out of the box. The Rackspace Managed Big Data Platform supports Hortonworks' technology application partners.

Commitment to Open Source

The Managed Big Data Platform is built on core open-source technology, not a proprietary version. Hortonworks is similarly committed to driving innovation in the open-source community, and is the only provider that works entirely within the Apache Software Foundation process. By relying on the OpenStack cloud from Rackspace and the Hadoop platform from Hortonworks, businesses can take advantage of flexible and convenient data portability.

Breadth and Depth of Tools and Integrations

The Hortonworks Platform includes all core Hadoop tools, and Hortonworks and Rackspace will continue to keep pace with the rapidly expanding Hadoop ecosystem.

Big data applications power a wide range of applications and platforms. The Rackspace Managed Big Data Platform environment comes ready to integrate with business intelligence and analytics applications or primary data stores. For example, Rackspace offers a MongoDB file-system connector and affordable object-storage platforms like Rackspace Cloud Files. Users can also deploy on traditional bare metal or other supported storage options, such as EMC Isilon or OpenStack Swift.

Developers can deploy and leverage Spark for real-time data processing, including the ability to rapidly write applications using Java, Scala or Python. Spark can also be used interactively from the Scala and Python shells.

6. Companies With Security and Compliance Requirements

Single-tenant solutions are inherently more secure than multi-tenant solutions, and they are highly customizable to help companies implement security features and manage access requirements.

As part of the Rackspace Big Data Platform service, Rackspace can combine customers' Hadoop clusters with available appliances and services from an extensive security portfolio that customers can purchase to help them address their compliance needs. The Rackspace portfolio includes intrusion detection, log management, web application firewalls, vulnerability scanning and data access monitoring.

7. Rackspace Managed Big Data vs. Unmanaged Hosting

	Unmanaged Hosting	Rackspace Managed Big Data Platform
Deployment Process	<p>Customer handles deployment</p> <ul style="list-style-type: none"> • Need for significant investment in Hadoop expertise • Manual configuration of storage, networking, security, monitoring and more • Manual testing and optimization, with a tradeoff between timely deployment and future scalability • Lengthy implementation and procurement cycles 	<p>Rackspace oversees deployment</p> <ul style="list-style-type: none"> • Rackspace experts gather customer requirements and optimize based on extensive experience and up-to-date Hadoop best practices • Rackspace handles configuration and deployment
Deployment Options	<ul style="list-style-type: none"> • Limited deployment flexibility — all dedicated servers or all public cloud 	<p>Multiple deployment options architected specifically for Hadoop:</p> <ul style="list-style-type: none"> • Public cloud • Private cloud • Bare metal • Hybrid cloud
Data Portability	<ul style="list-style-type: none"> • Expensive proprietary technologies and risk of vendor lock-in 	<ul style="list-style-type: none"> • Infrastructure built on OpenStack • Hortonworks platform works 100% within the Apache Software Foundation process
Performance	<ul style="list-style-type: none"> • Lower performance on generic servers not optimized for high-I/O database workloads • Noisy-neighbor problems on public cloud create inconsistency 	<ul style="list-style-type: none"> • Consistent database performance on hardware configured and tuned to make Hadoop run faster
Maintenance and Monitoring	<ul style="list-style-type: none"> • Customer allocates significant resources to monitor performance, maintenance, patching, etc. 	<ul style="list-style-type: none"> • Rackspace can proactively monitor the network, server and Hadoop clusters for connections, lag, status of replicas, size, number of queries, etc. — 24x7x365
Ecosystem Support	<ul style="list-style-type: none"> • Limited or no support for key Hadoop tools like Spark, Hive and Pig 	<ul style="list-style-type: none"> • Support available for core tools in the Hadoop ecosystem
Security Options	<ul style="list-style-type: none"> • Bare bones out-of-the-box Hadoop security • Additional security must be provided by additional vendors and often managed by the customer 	<ul style="list-style-type: none"> • Range of managed security options available like intrusion detection, log management and network security solutions
Vendor Accountability	<ul style="list-style-type: none"> • Separate infrastructure and platform providers can point fingers when things go wrong 	<ul style="list-style-type: none"> • Hortonworks platform and Rackspace infrastructure support on one support contract, giving customer a single POC

8. Use Cases

The Rackspace Managed Big Data Platform uses the popular Hortonworks Data Platform to help businesses address a variety of common use cases, including:

- Financial Analysis
- Clickstream Analysis
- Healthcare Record Retention and Processing
- Social Media Data and Sentiment Analysis

Financial Analysis

In the financial industry, controlling risk is paramount, and the only way to control risk is to understand it. Hadoop offers diverse options for obtaining deep, data-driven insights and using them to quickly evaluate potential risks, including the risk of fraud.

Rackspace Managed Big Data offers single-tenant isolation so customers can address security needs for their financial data.

Clickstream Analysis

Keeping customers on a website requires understanding their motivations. Using Hadoop, businesses can efficiently store and process enormous volumes of clickstream data — information about how users view, scroll over and click through web pages. This can give them the opportunity to analyze, visualize and understand how their visitors are behaving on their websites. Companies can also use clickstream data to instantly respond to customer behavior, offering users special deals or strategically re-targeting them. Forward-thinking retailers have made a business practice out of this use case, and are more effectively managing their ecommerce sites and online catalogs.

The Rackspace Managed Big Data Platform simplifies the deployment and maintenance of Hadoop, allowing marketers to speed up customer intelligence initiatives without overburdening IT departments. Traditionally, marketing organizations haven't budgeted for infrastructure or IT, so it is important to choose a platform that restrains costs.

Healthcare Record Retention and Processing

Many healthcare organizations store large volumes of data for extended periods of time and they also need to meet stringent security standards to protect patient privacy.

Hadoop is an inherently distributed system that can allow healthcare IT teams to store data far more affordably than they could with relational data-warehouse systems. And they will gain access to shared compute, enabling them to run analytics and ad hoc queries on patient data. When doctors are guided by data-driven insights and have instant access to health records, they can make better-informed decisions faster.

With Rackspace Managed Big Data Platform, Rackspace and Hortonworks experts facilitate the normally difficult process of deploying Hadoop. Moreover, Rackspace offers a number of security solutions that healthcare companies can choose from to help them meet their compliance needs.

Social Media Data and Sentiment Analysis

With the Hadoop, businesses can quickly mine, archive and analyze many thousands of social media conversations for sentiment data — the ratio of positive comments to negative comments, for example. Companies can assess customers' attitudes about their products, services, brand and competitors, and then track those opinions over time.

This kind of data can be leveraged in a number of ways, allowing a business to make targeted, real-time decisions. From gathering data on social media platforms after a product launch or tracking a decline in customer engagement on social media posts, Hadoop makes it all possible.

Hadoop makes it possible to manage and exploit big data, whether that means streaming social media feeds or querying, refining and visualizing the information they contain. With Rackspace Managed Big Data, you won't need Hadoop experts, and social analytics projects like these will become much more feasible for both IT and marketing. And for teams with limited Hadoop experience, Rackspace experts are always on hand, ready to help users leverage Flume to stream Twitter feeds into HDFS, HCatalog to build a relational view of the data or Hive to query and refine it.

9. Conclusion

For businesses seeking an easier way to stand up and manage Hadoop, the Rackspace Managed Big Data Platform is an ideal choice. Using the popular Hortonworks Data Platform, it allows companies to reap the benefits of big data without getting bogged down in deployment, optimization and maintenance challenges. Each environment is custom-designed to fit customers' exact requirements, so they can take advantage of big data insights while dedicating internal resources to initiatives that drive the business forward.

And Rackspace is constantly adding new and requested open-data platforms like MongoDB, Redis, Elasticsearch and MySQL that work alongside Hadoop. Choosing a vendor who can help you implement a multi-database strategy is difficult, since few vendors offer a meaningful breadth of open-source data platforms and support.

To learn more about the Rackspace Managed Big Data Platform, contact a Rackspace sales representative at **1-888-480-7667**. Or visit www.rackspace.com/managed-hosting/data/big-data.

About Rackspace

Rackspace® (NYSE: RAX) is the #1 managed cloud company. Its technical expertise and **Fanatical Support**® allow companies to tap the power of the cloud without the pain of hiring experts in dozens of complex technologies. Rackspace is also the leader in hybrid cloud, giving each customer the best fit for its unique needs — whether on single- or multi-tenant servers, or a combination of those platforms. Rackspace is the founder of OpenStack®, the open-source operating system for the cloud. Based in San Antonio, Rackspace serves more than 300,000 business customers from data centers on four continents.

GLOBAL OFFICES

Headquarters Rackspace, Inc.

1 Fanatical Place | Windcrest, Texas 78218 | 1-800-961-2888 | Intl: +1 210 312 4700

www.rackspace.com

UK Office

Rackspace Ltd.
5 Millington Road
Hyde Park Hayes
Middlesex, UB3 4AZ
Phone: 0800-988-0100
Intl: +44 (0)20 8734 2600
www.rackspace.co.uk

Benelux Office

Rackspace Benelux B.V.
Teleportboulevard 110
1043 EJ Amsterdam
Phone: 00800 8899 00 33
Intl: +31 (0)20 753 32 01
www.rackspace.nl

Hong Kong Office

9/F, Cambridge House, Taikoo Place
979 King's Road,
Quarry Bay, Hong Kong
Sales: +852 3752 6488
Support +852 3752 6464
www.rackspace.com.hk

Australia Office

Rackspace Hosting Australia PTY LTD
Level 1
37 Pitt Street
Sydney, NSW 2000
Australia

© 2015 Rackspace US, Inc. All rights reserved.

This whitepaper is for informational purposes only and is provided "AS IS." RACKSPACE MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, AS TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS DOCUMENT AND RESERVES THE RIGHT TO MAKE CHANGES TO SPECIFICATIONS AND PRODUCT/SERVICES DESCRIPTION AT ANY TIME WITHOUT NOTICE. RACKSPACE RESERVES THE RIGHT TO DISCONTINUE OR MAKE CHANGES TO ITS SERVICES OFFERINGS AT ANY TIME WITHOUT NOTICE. USERS MUST TAKE FULL RESPONSIBILITY FOR APPLICATION OF ANY SERVICES AND/OR PROCESSES MENTIONED HEREIN. EXCEPT AS SET FORTH IN RACKSPACE GENERAL TERMS AND CONDITIONS, CLOUD TERMS OF SERVICE AND/OR OTHER AGREEMENT YOU SIGN WITH RACKSPACE, RACKSPACE ASSUMES NO LIABILITY WHATSOEVER, AND DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO ITS SERVICES INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT.

Except as expressly provided in any written license agreement from Rackspace, the furnishing of this document does not give you any license to patents, trademarks, copyrights, or other intellectual property.

Rackspace and Fanatical Support are either registered service marks or service marks of Rackspace US, Inc. in the United States and other countries.