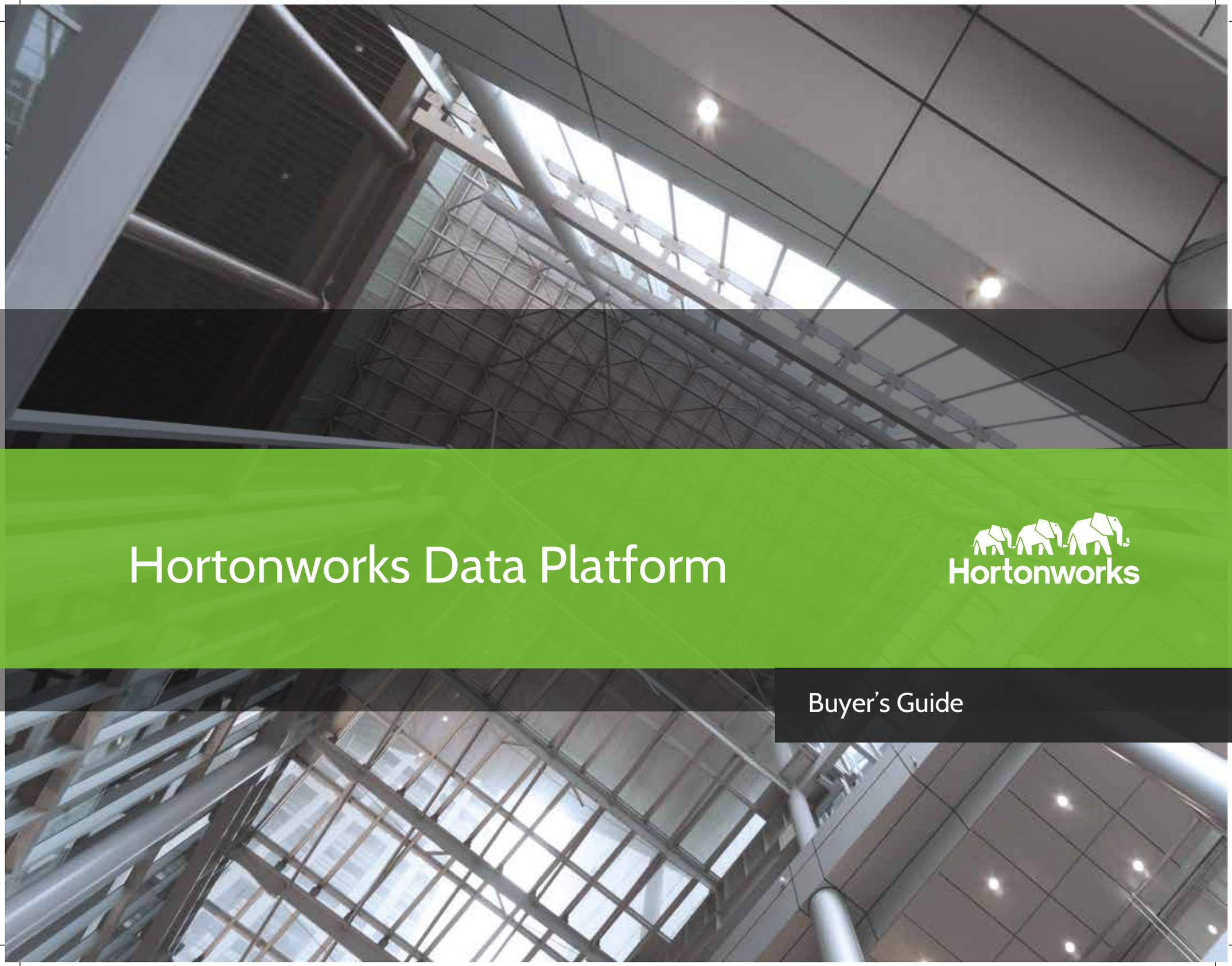
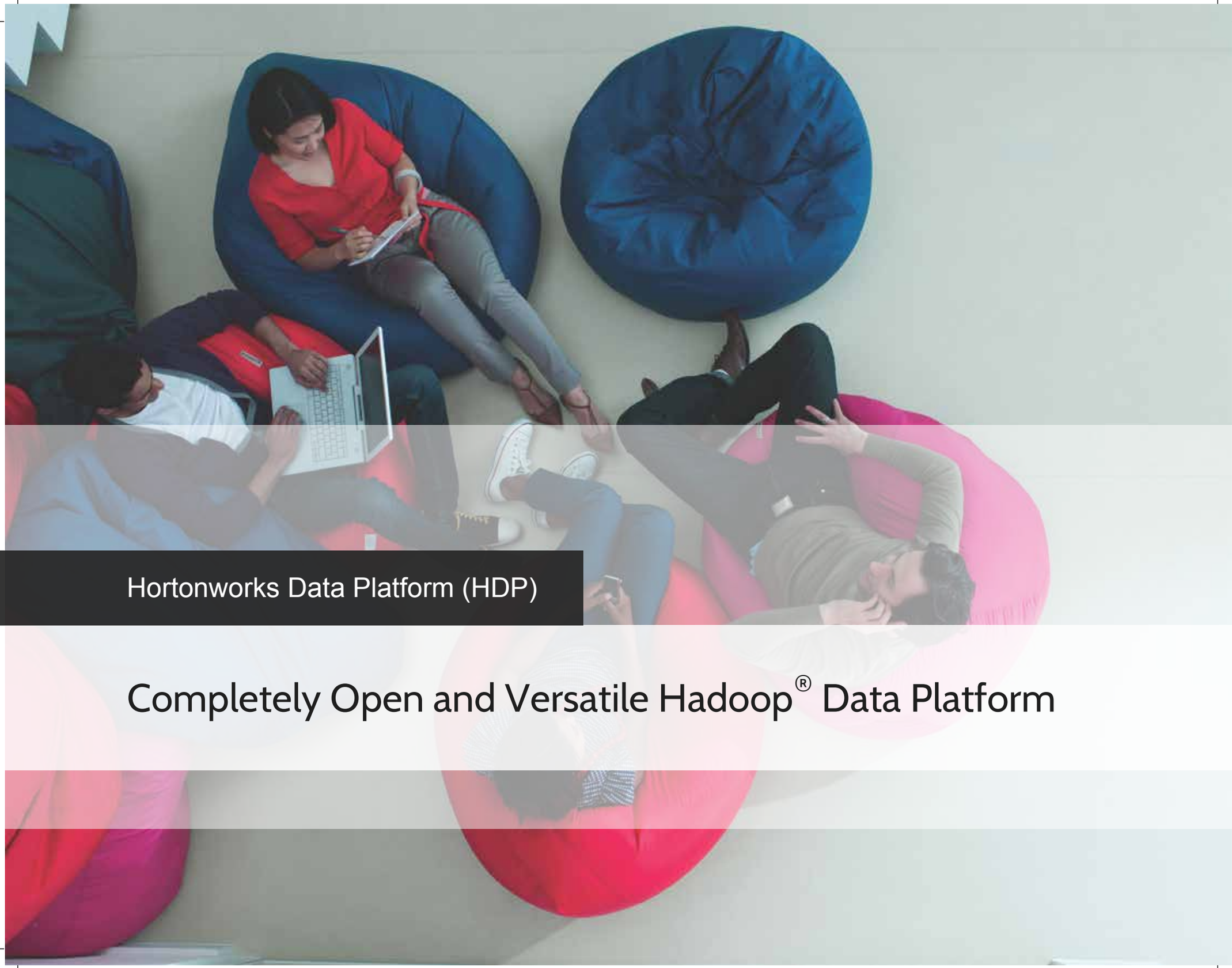


Hortonworks Data Platform

Hortonworks

Buyer's Guide



Hortonworks Data Platform (HDP)

Completely Open and Versatile Hadoop® Data Platform

Table of Contents

Hortonworks Data Platform 4

Enterprise Hadoop within a Modern Data Architecture 5

Hortonworks Data Platform: Enterprise Hadoop 6

- HDP: Data Management 8
- HDP: Data Access 9
- HDP: Data Governance and Integration 11
- HDP: Security 12
- HDP: Cluster Operations 13
- Deployment Options for Hadoop 14

HDP-enabled Processing 15

- Multiple Workloads on a Single Big Dataset 16
- Batch Processing 17
- Interactive SQL Query 17
- Application Enrichment 18
- Search 18
- Stream Analysis 19
- Future Data Access Methodologies 19

HDP for Teams 20

- The Data Architect 21
- The Data Scientist 22
- The Business Analyst 23
- The Developer 24
- The System Administrator 25

Why Hortonworks for Hadoop? 26

Hortonworks Data Platform

Architected, developed and built completely in the open, Hortonworks Data Platform (HDP) is designed to meet the changing needs of enterprise data processing.

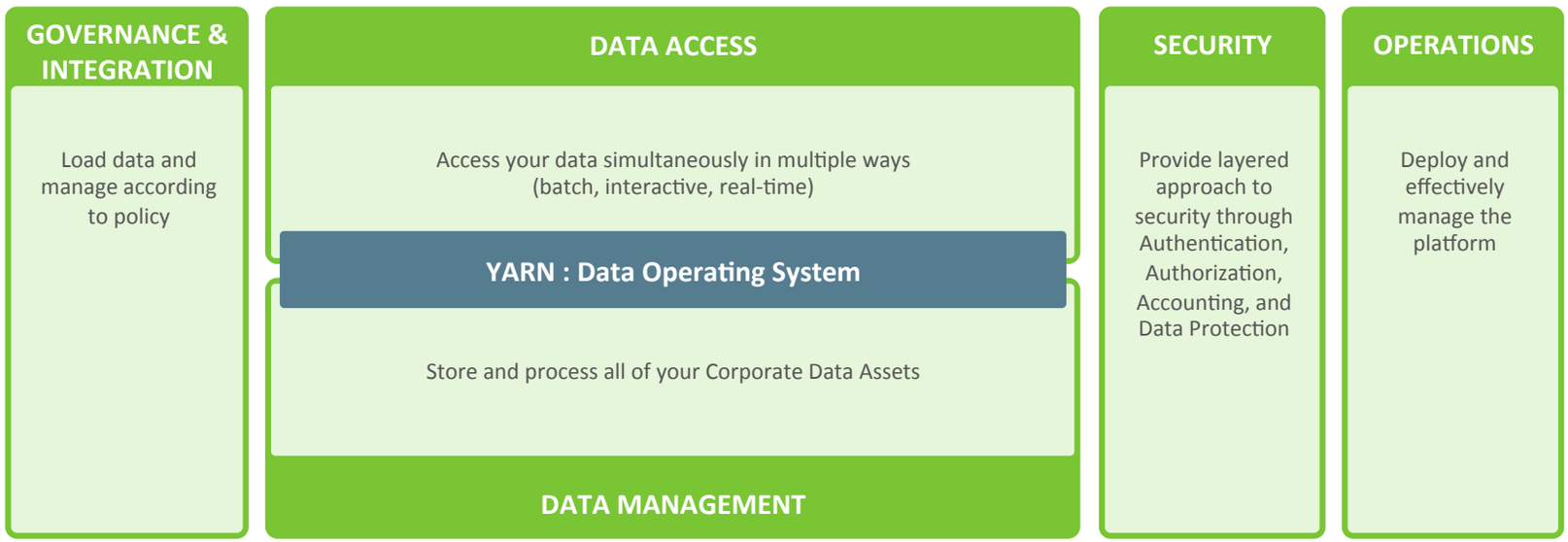
HDP is fundamentally versatile, providing linear, scalable storage and compute across a wide range of access methods, from batch and interactive to real time, search and streaming. It includes a comprehensive set of the essential data capabilities required by the modern enterprise across governance, integration, security and operations.

As the only completely open Apache™ Hadoop® platform available, HDP integrates with and augments your existing best-of-breed applications and systems so you can gain value from your enterprise Big Data, with minimal changes to your data architectures. Finally, HDP allows you to deploy Hadoop wherever you want it–from cloud or on-premises as an appliance, and across both Linux and Windows.

Enterprise Hadoop within a Modern Data Architecture

The emergence and explosion of new types of data in recent years have put tremendous pressure on the data architectures of many organizations. In response, many have turned to Apache Hadoop as a key component of their overall data architectures to manage this explosion of data.

Apache Hadoop originated as a simple project at the Apache Software Foundation (ASF) to manage and access data, and included just two components: the Hadoop Distributed File System (HDFS) and MapReduce, a processing framework for data stored in HDFS. Over time, the Hadoop platform expanded to incorporate a range of ASF projects that are required of a complete enterprise data platform. These logically fall into five distinct categories: data access, data management, security, operations and governance.



Hortonworks Data Platform: Enterprise Hadoop

The Hortonworks Data Platform (HDP) delivers all of these essential capabilities in a completely open, integrated and tested platform that is ready for enterprise usage–and most important, with no proprietary holdbacks aimed at locking customers into HDP.

HDP is deeply integrated with the incumbent technologies within your data center to deliver a modern data architecture that allows users to leverage their existing skills and investments.

The key tenets of HDP are:

Completely Open

HDP is designed for the enterprise, built in the open and tested at scale. It is comprised of 100% open source components developed by the open community within the governance of the Apache Software Foundation.

Fundamentally Versatile

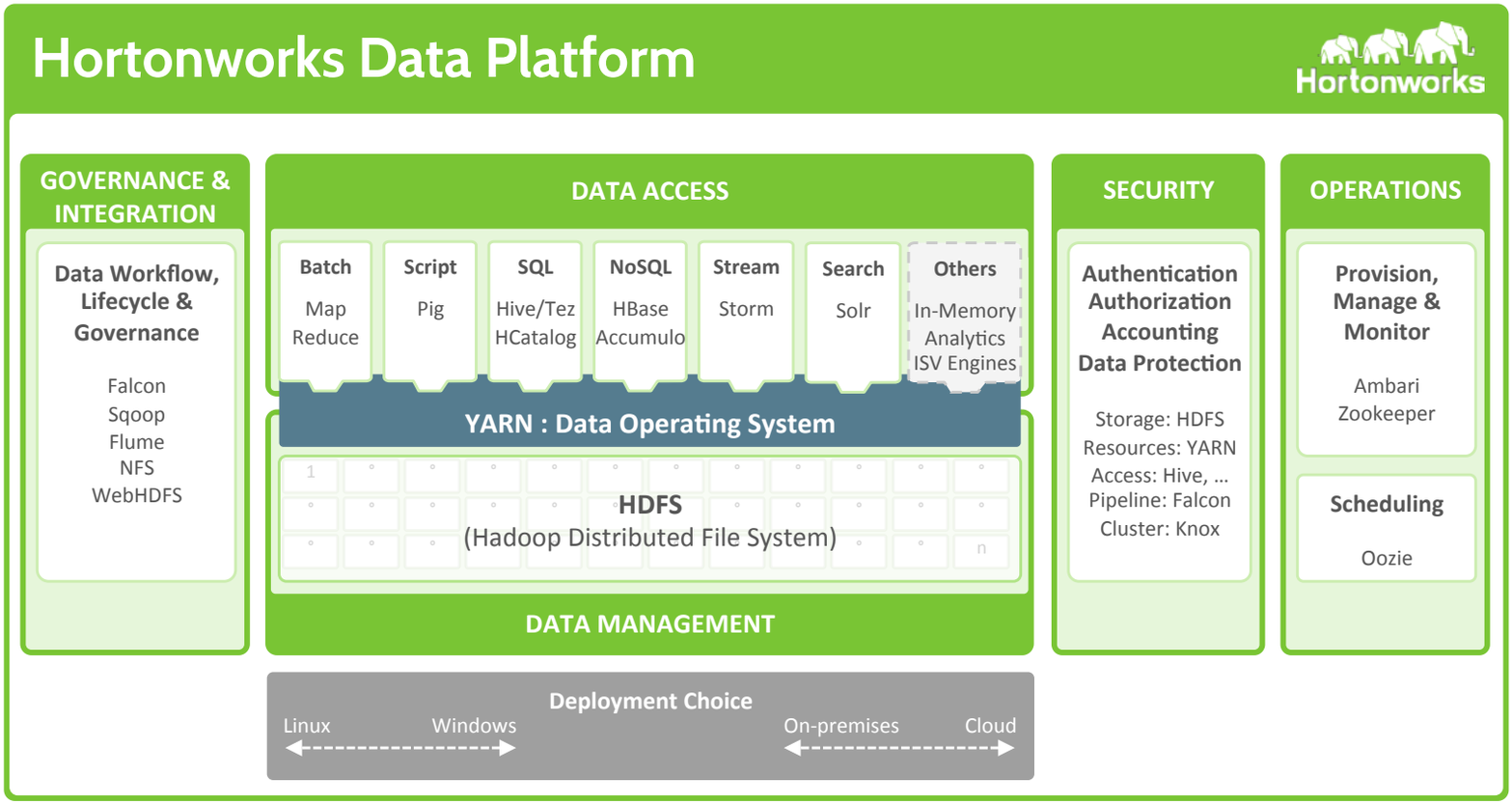
With Hadoop YARN at its core, HDP provides flexible enterprise data processing across a range of data processing engines, paired with comprehensive enterprise capabilities for governance, security and operations.

Wholly Integrated

HDP is designed to integrate deeply and augment existing data center technologies and is deployable across the widest possible range of deployment options.

Comprehensive Enterprise Capabilities

HDP presents a comprehensive set of enterprise capabilities aligned to the following functional areas: data management, data access, data governance and integration, security, and operations.



HDP: Data Management

Store and process all of your corporate data assets.

Hadoop Distributed File System (HDFS)

HDFS is the file system of Hadoop that provides linear scale and reliable data storage, designed for distributed computing across large clusters of low-cost commodity servers. With linear scale storage on commodity hardware, you more efficiently store your large corporate data assets.

Apache Hadoop YARN

YARN is the data operating system of Hadoop that enables you to process data simultaneously in multiple ways. YARN is the prerequisite for Enterprise Hadoop, providing the resource management and pluggable architecture for enabling a wide variety of data access methods to operate on data stored in Hadoop with predictable performance and service levels.

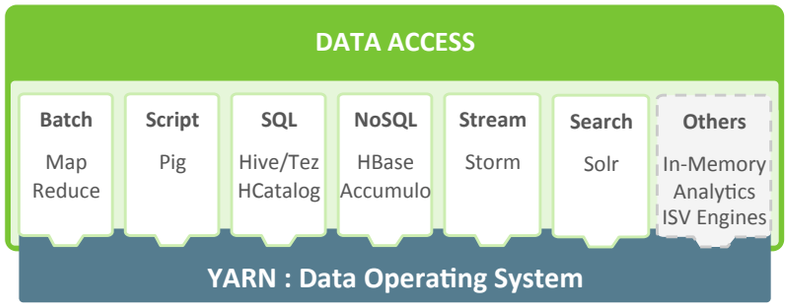
Accessing Data Simultaneously Across Multiple Engines

With Hadoop 2 and YARN, multiple processing engines can share this single data platform, enabling you to run, for instance, a Pig script to analyze a stream of data at the same time (on the same data) as you are analyzing a stream of data or feeding online applications with real-time access to HBase data. A singular system not only optimizes use of data, but also presents new opportunities.

HDP: Data Access

Access and interact with your data across a variety of engines–spanning batch, interactive, streaming and real-time use cases.

With YARN at its foundation, HDP provides a range of processing engines that allow users to interact with data in multiple and parallel ways, without the need to stand up individual clusters for each data set/application. Some applications require batch while others require interactive SQL or low-latency access with NoSQL. Other applications require search, streaming or in-memory analytics: Apache Solr, Storm and Spark fulfill those needs respectively. To function as a true data platform, the YARN-based architecture of HDP enables the widest possible range of access methods to coexist within the same cluster–avoiding unnecessary and costly data silos.



HDP natively provides for the following data access types:

Batch

Apache MapReduce has served as the default Hadoop processing engine for years. It is tested and relied upon by many existing applications.

Interactive SQL Query

Apache Hive™ is the de facto standard for SQL interactions at petabyte scale within Hadoop. Hive delivers interactive and batch SQL query across the broadest set of SQL semantics.

Search

HDP integrates Apache Solr to provide high-speed indexing and sub-second search times across all your HDFS data.

Scripting

Apache Pig is a scripting language for Hadoop that can run on MapReduce or Apache Tez, allowing you to aggregate, join and sort data.

Low-latency access via NoSQL

Apache HBase provides extremely fast access to data as a columnar format, NoSQL database. Apache Accumulo also provides high-performance storage and retrieval, but with fine-grained access control to the data.

Streaming

Apache Storm processes streams of data in real time and can analyze and take action on data as it flows into HDFS.

Apache HCatalog

Apache HCatalog provides a critical layer of metadata management for Hadoop so that users and processing engines can share a single set of schema for Hadoop data.

Apache Tez

Apache Tez is a modern, efficient processing engine for Hadoop that serves as a new foundation for higher performance within Hive and Pig. It is general purpose and implements the same API as MapReduce, so both existing and new applications can take advantage of its more powerful framework.

HDP: Data Governance and Integration

Enables you to quickly and easily load data, and manage according to policy.

HDP extends data access and management with powerful tools for data governance and integration. These tools provide a reliable, repeatable and simple framework for managing the flow of data in and out of Hadoop. This control structure, along with a set of schema or metadata on sources, is critical for the successful integration of Hadoop into your modern data architecture.

The following data governance and integration components ship with HDP:

Apache Falcon is a framework for simplifying data management and pipeline processing. Falcon simplifies the configuration of data motion and allows you to set policy for: retention and replication of datasets; lineage and traceability; and any tagging, filtering and processing (applying schema, for instance) of data on ingest.

Apache Sqoop efficiently transfers bulk data between Hadoop and structured data stores such as Teradata, Netezza, Oracle, MySQL, Postgres and HSQLDB.

Apache Flume is used to stream data from multiple sources into Hadoop for analysis. It has a simple, flexible architecture with reliable failover and recovery.

Third-Party Data Integration. Hortonworks and all its data management providers work to integrate their tools with HDP. Many have even extended their capabilities into Apache YARN.

HDP: Security

Administer consistent policy across requirements for authentication, authorization, audit and data protection.

HDP provides a centralized approach to security management whcih allows you to define and deploy consistent security policy throughout the data platform. HDP allows you to easily create and administer a central security policy while coordinating consistent enforcement across all Hadoop applications. It provides a comprehensive set of critical features for authentication, authorization, audit and data protection so that you can apply enterprise grade security to your Hadoop deployment. Consistent with all enterprise Hadoop capabilities, HDP also makes certain that you can integrate and extend your current security solutions to provide a single, consistent, secure umbrella over your modern data architecture.

Central Administration of a consistent security policy across all Hadoop data and access methods can be defined with a security administration console that is unique to HDP.

Authentication verifies the identity of a system or user accessing the system, a file or application. Apache Knox provides a single point of authentication/access for your cluster and integrates with your existing LDAP or Active Directory implementations. HDP also provides Kerberos for simple authentication.

Authorization specifies access privileges for a user or system. Hadoop provides fine-grained authorization via file permissions in HDFS, resource-level access control for YARN and MapReduce, and coarser-grained access control at a service level. HBase provides authorization with ACL on tables and column families, while Accumulo extends this further to cell-level control. Apache Hive provides Grant/Revoke access control on tables.

Audit provides the ability to track resource use within a system. HDFS and MapReduce provide audit support. Additionally, Apache Hive metastore records audit (who/when) information for Hive interactions. Apache Oozie provides an audit trail for services.

Data Protection ensures privacy and confidentiality of information. HDP encrypts data in motion across channels such as Remote Procedure Call (RPC), HTTP, JDBC/ODBC and Data Transfer Protocol (DTP). HDFS and Hadoop support encryption at the operating system level. Additionally, Apache Falcon provides the capability to extend Hadoop with powerful third-party encryption tools to protect data at rest.

HDP: Cluster Operations

Provision, manage, monitor and operate Hadoop clusters at scale.

Operations teams deploy, monitor and manage a Hadoop cluster as part of their broader enterprise data ecosystem. HDP delivers a comprehensive set of completely open operational capabilities that provide both visibility into cluster health as well as the ability to manage and configure resources. Powerful tools are provided to optimize and tune performance across all data access methods and workloads running in the Hadoop cluster.

Apache Ambari is a completely open framework to provision, manage and monitor Apache Hadoop clusters. It provides a simple, elegant UI that allows you to image a Hadoop cluster. Architected to integrate with existing operations tools such as Microsoft System Center and Teradata Viewpoint, Ambari provides a single pane of glass across your entire modern data architecture.

Apache ZooKeeper™ provides a distributed configuration service, a synchronization service and a naming registry for distributed systems. Distributed applications use ZooKeeper to store and mediate updates to important configuration information.

Apache Oozie provides a critical scheduling capability to organize and schedule jobs within Enterprise Hadoop across all data access points.

Deployment Options for Hadoop

HDP offers the broadest range of deployment options for Hadoop, from Windows Server or Linux to virtualized cloud deployments. It is the most portable Hadoop distribution available, allowing you to easily and reliably migrate from one deployment type to another.

Windows and Linux

Only HDP provides both a Windows and Linux Hadoop distribution, which means you get the best fit for your data center, whatever your choice of platform.

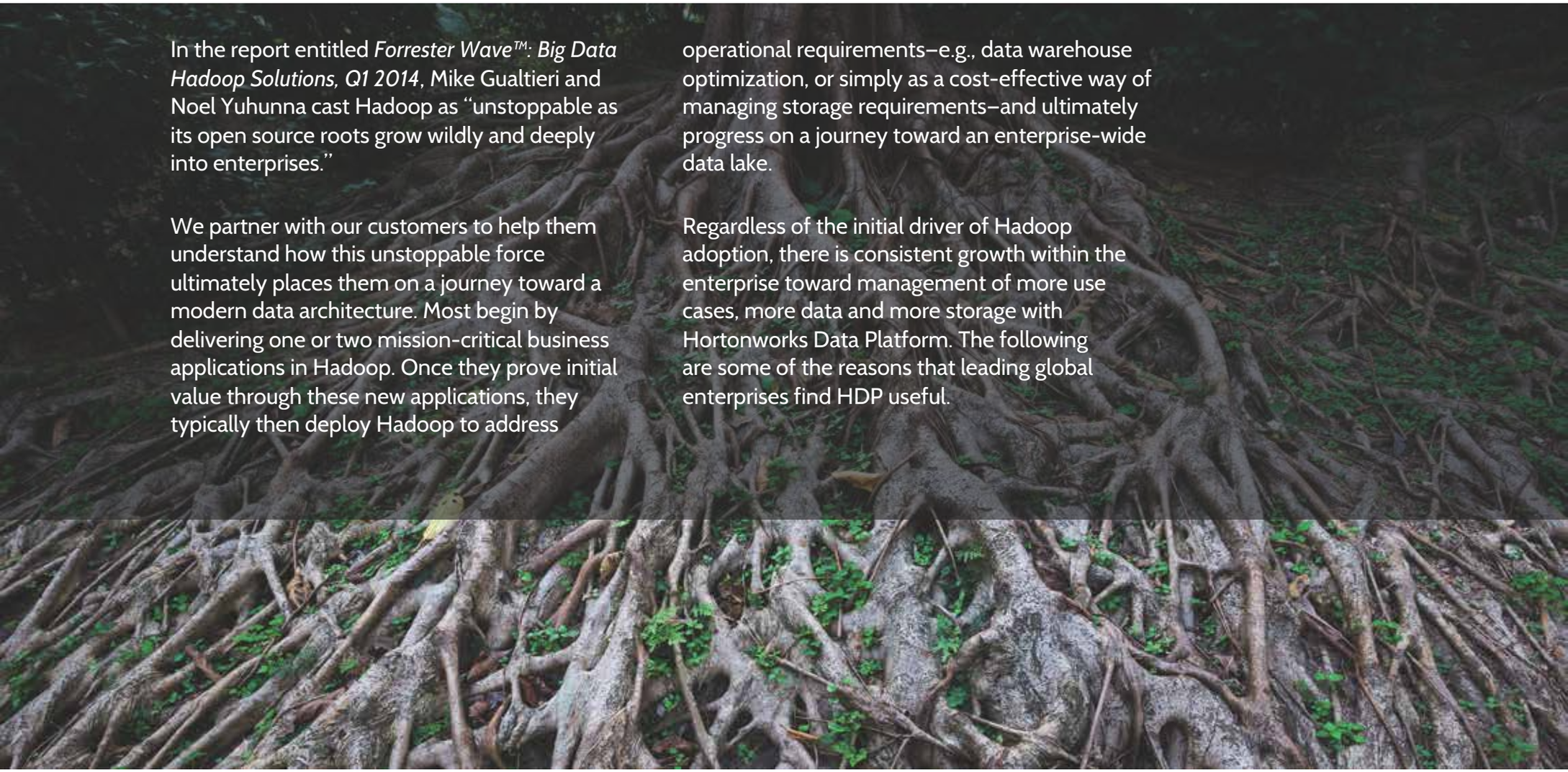
Cloud

HDP is available as Microsoft® HDInsight on the Azure cloud, and is the default distribution available for a private or public cloud at Rackspace®. Additionally, our work with OpenStack allows you to easily install and provision a Hadoop cluster in that environment.

Flexibility

Because the final package of HDP is the same across cloud deployments and on-premises, you can seamlessly migrate or port from one deployment method to another.

HDP-enabled Processing



In the report entitled *Forrester Wave™: Big Data Hadoop Solutions, Q1 2014*, Mike Gualtieri and Noel Yuhunna cast Hadoop as "unstoppable as its open source roots grow wildly and deeply into enterprises."

We partner with our customers to help them understand how this unstoppable force ultimately places them on a journey toward a modern data architecture. Most begin by delivering one or two mission-critical business applications in Hadoop. Once they prove initial value through these new applications, they typically then deploy Hadoop to address operational requirements–e.g., data warehouse optimization, or simply as a cost-effective way of managing storage requirements–and ultimately progress on a journey toward an enterprise-wide data lake.

Regardless of the initial driver of Hadoop adoption, there is consistent growth within the enterprise toward management of more use cases, more data and more storage with Hortonworks Data Platform. The following are some of the reasons that leading global enterprises find HDP useful.

Multiple Workloads on a Single Big Dataset

With a YARN-based architecture serving as the data operating system for Hadoop 2, HDP takes Hadoop beyond single-use, batch processing to a fully functional, multi-use platform that enables batch, interactive and real-time data processing. Leading organizations can now use YARN and HDP to process data and derive value from multiple business cases and realize their vision of a data lake.

The value in delivering multiple access methods on a single set of data extends beyond data science. It allows a business to set an architecture where it can deliver multiple value points all across a single set of data to create an enterprise capability previously only imagined. For instance, an organization can analyze real-time clickstream data using Apache Storm to pick off events that need attention, run an Apache Pig script to update product catalog recommendations, and then deliver this information via low-latency access through Apache HBase to millions of web visitors–all in real time, and all on a single set of data.

Many best-practice architectures have emerged to surface the value of a multi-use Big Data platform:

Advertisers target ads to their best customer segment and also analyze point-of-sale data to determine the effectiveness of campaigns.

Banks detect fraud and money laundering while also improving customer service.

Hospitals respond to patients in real time and then analyze historical data to reduce readmission rates.

Manufacturers control quality on the production line and then diagnose product defects in the aggregate.

Oil companies predict and repair equipment proactively and also analyze equipment durability under varied circumstances.

Telecoms allocate bandwidth in real time, and later discover unforeseen patterns after analyzing billions of historical call records.

Retailers make sure the shelves are stocked today and also plan their product mix for next year.

The ability to inspect and analyze data across multiple methods is liberating. Data scientists can perform a search, execute SQL and then perform in-memory analytics on results–all on a single set of data. This level of analysis of massive data sets has not been available until the advent of YARN, a critical component of HDP and the cornerstone of a modern data architecture.

Hortonworks and YARN

In 2009, Arun Murthy, a co-founder of Hortonworks, presented a new architecture for Hadoop, in which resource management was recast as a new component called YARN. Thereafter, the team at Hortonworks has led the design, development and delivery of this core technology and we continue to innovate this crucial component and work with our ecosystem partners to integrate their solutions with YARN.

Hadoop Workload:

Batch Processing

Engineers at Yahoo! developed Apache Hadoop to solve the large, web-scale batch processing problem of indexing the Internet. Today, the HDFS and MapReduce frameworks still provide a powerful engine for refining massive amounts of data in batch, but recently, Apache Tez joined MapReduce in HDP. Apache Tez now manages many of MapReduce's tasks more efficiently, with faster results.

Whether with MapReduce or Tez, Hortonworks' customers use Hadoop as a batch data refinery to distill large quantities of data into small manageable units. That refined data can then be loaded into existing data systems and accessed by familiar analytical tools–with a richer context and smaller dataset.

Hadoop Workload:

Interactive SQL Query

It can be argued that SQL serves as the universal language for data. SQL queries are, by far, the most widely used method for accessing and analyzing data with enterprise analytic tools. Apache Hive has served as the de facto SQL interface for Hadoop data since its introduction in 2008. Originally constructed to meet batch workloads, it has advanced into a powerful interactive query interface, and along with HPD, delivers the broadest set of SQL semantics to perform interactive query at petabyte scale.

Hadoop Workload:

Application Enrichment

Website customization is the canonical use case for application enrichment. For instance, many retailers store all of their web session data and use it to customize the experience for different customers who visit their websites. As a more detailed example, consider an investment services firm that enriches the information it provides to its applications, traders and analysts using Hadoop. The data is queried more than 30,000 times per second, yet this firm still meets its SLA to return the data in a fraction of a second.

Apache HBase is a NoSQL database that runs on top of the Hadoop Distributed File System (HDFS). It is columnar and provides fault-tolerant storage and quick access to billions of rows and millions of columns of data. It also adds transactional capabilities to Hadoop, allowing users to conduct updates, inserts and deletes. With HBase, data stored in Hadoop can be used to enrich the behavior of other applications.

Hadoop Workload:

Search

Freeform search of data has become a standard method for many to interact with data. It allows for exploration and helps business resources find relevant data across large, disparate data stores with mixed formats and structures found in Hadoop.

As part of HDP, Apache Solr brings full-text, interactive search, and flexible indexing to HDP. It is highly reliable, scalable and fault tolerant, providing distributed indexes, load-balanced queries, automated failover and recovery, centralized configuration and replication. Solr already powers the search and navigation features of many of the world's largest Internet sites.

Hadoop Workload:

Stream Analysis

There are many use cases for stream processing using Hadoop. HDP allows you to process a stream to pick off an event that needs to be addressed. In healthcare, a provider can actively monitor patient vital signs streaming from digital monitoring devices to respond quickly to emergencies. Utilities can analyze data streaming from smart meters to better manage electricity generation. And logistics companies can monitor streaming driving data to notify its drivers about unsafe patterns, protecting both the driver and the company's reputation. Hortonworks Data Platform includes Apache Storm for real-time analysis on streaming data.

Hadoop Workload:

Future Data Access Methodologies

YARN has enabled a diverse set of technologies to use linear scale compute and storage to execute and deliver value. Emerging applications are in development today that will allow us to look at data in new ways tomorrow. Hortonworks and HDP are committed to extending these critical access methods.

HDP for Teams

The successful rollout and ongoing use of Hadoop are predicated on enabling your existing resources to efficiently adopt Hadoop into their existing routines. But how do you extend the skillsets of your valuable resources while they remain incredibly busy supporting current infrastructure? Over the past few years, we have worked with hundreds of organizations to enable their teams to use Hadoop.

Success requires an incremental extension on existing skills across these five key resources:

1

The Data Architect

2

The Data Scientist

3

The Business Analyst

4

The Developer

5

The System Administrator

The Data Architect

Responsible for outlining the plan and architecture of a Hadoop deployment, data architects are critical to the success of integrating Hadoop into a modern data architecture. They establish standards and are ultimately accountable for ensuring that Hadoop, as part of a larger data architecture, aids the organization in achieving its strategic goals. To this end, the data architect creates a description of assets and outlines the interplay between systems. Additionally, he or she must address security concerns as well as to ensure that all the tools are in place for the daily operation of a Hadoop cluster.

HDP provides several key tools useful to the data architect, including:

Apache Falcon to set governance policy around retention/replication policies of datasets, or to tag and process data on ingest to Hadoop.

Apache Sqoop™ and **Apache Flume** to reliably and quickly integrate data with existing systems.

HCatalog to establish critical metadata management framework.

Apache Knox to set perimeter security.

The ecosystem is more important to the data architect than to any other individual involved with Hadoop and HDP. The data architect must ensure integration with existing assets like Teradata, SAP, Microsoft and many of the other important players in the data center. Applications must work together to provide value and to operate the environment. HDP is the only distribution focused on enabling these key data center partners to integrate with Hadoop and, consequently, help ease the job of the data architect.

The Data Scientist

Data scientists explore and analyze data to uncover and expose information found within Hadoop, using a wide range of tools to accomplish their work. They will look at the same data in multiple ways, using scripts and SQL to search. As well as working on advanced algorithms, they will synthesize new value from large and often nebulous data sets. They are scientists who discover new formulas that may eventually become operationalized as part of a business process.

HDP enables this unique interplay by providing a wide range of access methods to a single set of data. Some of the essential tools found within HDP to aid with this exploratory function include:

Apache Hive, the de facto standard for SQL in Hadoop, with the widest range of analytic semantics.

Apache Mahout™, a data science framework for advanced algorithms.

Apache Solr, an open and extensible framework for natural language search of data within Hadoop.

Their job cannot be accomplished wholly in Hadoop. Python, R, JavaScript and third-party visualization tools are among the many tools that integrate with and extend the value of Hadoop for these critical, highly talented resources. Hortonworks works closely with the ecosystem of vendors and the open community to ensure that these tools all work together to provide a fabric of tooling that empowers the data scientist to do great things.

The Business Analyst

Like data scientists, business analysts are tasked with identifying value within data; however, they use a more mature and established set of tooling such as business intelligence, visualization, the enterprise data warehouse and data marts. They also use spreadsheets, create charts and set forth KPIs for an organization to measure success. The challenge for business analysts in the world of Big Data is to know where to look for what. With Hadoop, their creative energy can now focus on identifying new value in data sources that were once too big and unwieldy.

HDP integrates with existing technologies to ensure a seamless continuum from business apps to Hadoop. HDP offers two key functions that enable this:

HCatalog for metadata management, which ensures that applications can talk to each other via schema.

Apache Hive, the de facto standard for SQL in Hadoop, which implements the broadest set of data types and semantics so it can execute the SQL queries you require.

Within a modern data architecture, HDP abstracts the complexity of Hadoop for analysts through tight integration with existing tools. Hortonworks works within the ecosystem to ensure that the tools and applications used by the business analyst are Hadoop-enabled. This is achieved through close engineering relationships with key partners like Microsoft, Teradata and SAP, and is instantiated in tools like Excel, SQL-H and HANA.

The Developer

Developers are responsible for building Big Data apps in, on and around Hadoop, and expect a vibrant and powerful set of tools, frameworks and interfaces to simplify this task. They are focused on delivering on the value of an application and do not want to be mired in the mechanical details of integration with Hadoop.

Java and .NET represent the large majority of development technologies used to build enterprise applications today. These technologies already have extensions for Hadoop to enable developers to build applications.

The Microsoft .NET SDK for Hadoop provides API access to HDP and Microsoft HDInsight including HDFS, HCatalag, Oozie and Ambari, as well as some Powershell scripts for cluster management. There are also libraries for MapReduce and LINQ to Hive, with the latter building on the established technology for .NET developers to access most data sources to deliver the capabilities of the de facto standard for Hadoop data query.

The Cascading framework provides Java developers with the high-level Java constructs and classes to implement their data flow easily as a data-driven journey: flowing from a source and tap, traversing data preparation and traps, undergoing transformation, and finally, ending up in a sink for user consumption.

The Spring XD framework provides Java developers with libraries to integrate Hadoop with their applications. Some of the functions available are ingest/export of HDFS data, execute real-time analytics at ingestion time and workflow management of Hadoop jobs across MapReduce, HDFS, Pig, Hive or Cascading.

HDP also features a set of native APIs that ease development. WebHDFS provides a REST interface to write, manipulate and delete HDFS files, while WebHCAT provides a critical point of integration to access metadata and schema for Hadoop data. Standard APIs are available across all the data access methods mentioned above.

The System Administrator

An Apache Hadoop cluster presents a new challenge to IT operators. It is a collection of a handful (or thousands) of machines, all working collectively on solving a problem at scale. Initial provisioning can sometimes be difficult, even with only a small number of nodes. Ongoing management and monitoring of the environment require complex networking of resources and software.

HDP packages hundreds of years of combined experience to deliver the key services required to solve the operational challenges presented by Hadoop. Apache Ambari represents the most complete experience for provisioning, managing and monitoring Hadoop clusters, and is freely available for everyone–not merely as an optional extra, but as a core component of HDP. It allows you to:

Provision a Hadoop cluster: No matter the size of your Hadoop cluster, the deployment and maintenance of hosts are simplified using Ambari. Ambari includes an intuitive web interface that allows you to easily provision, configure and test all the Hadoop services and core components.

Manage a Hadoop cluster: Ambari provides tools to simplify cluster management. The web interface allows you to start/stop/test Hadoop services, change configurations and manage the ongoing growth of your cluster.

Monitor a Hadoop cluster: Gain instant insight into the health of your cluster. Ambari preconfigures alerts for watching Hadoop services and visualizes cluster operational data in a simple web interface. Ambari also includes job diagnostic tools to visualize job interdependencies and view task timelines as a way to troubleshoot historic job performance execution.

Integrate Hadoop with other applications: Ambari provides a RESTful API that enables integration with existing tools, such as Microsoft System Center and Teradata Viewpoint. Ambari also leverages standard technologies and protocols with Nagios and Ganglia for deeper customization.

These key services are not all that is required by the system administrator. They also need to schedule jobs (using Apache Oozie) and implement a security framework.

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. Our team members are active participants and leaders in Hadoop development, with proven expertise in designing, building and testing the core of the Hadoop platform. We have years of experience in Hadoop operations and are best suited to support your mission-critical Hadoop project.

Open Leadership

Hortonworks has a singular focus and commitment to drive innovation in the open, exclusively via the Apache Software Foundation process.

Hortonworks is responsible for the majority of core code base advances to deliver Apache Hadoop as an enterprise data platform.

Ecosystem Endorsement

Hortonworks is focused on the deep integration of Hadoop with existing data center technologies and team capabilities.

Hortonworks has secured strategic relationships with trusted data center partners, including Microsoft, SAP, Teradata and Rackspace, and many more.

Enterprise Rigor

Hortonworks has world-class enterprise support and services with vast experience in the largest Hadoop deployments.

Hortonworks builds Apache Hadoop with the enterprise in mind, all certified and tested with real-world rigor in the world's largest Hadoop clusters.

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

Hortonworks develops, distributes and supports a completely open 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 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. For more information, visit www.hortonworks.com.

Hortonworks

www.hortonworks.com