Brochure

# Enterprise-grade Hadoop with HPE Enterprise Services

Unlock the most value and performance out of Apache Hadoop and scale without compromise

Hewlett Packard Enterprise

As organizations strive to identify and realize the value in Big Data, many now seek more agile and capable analytic systems. Apache Hadoop is a software framework that provides measurable savings and value for storage and data processing at tremendous scale. While many have piloted Hadoop as a data repository for simple workloads, there is much more value that can be created from Hadoop by leveraging the data in the platform more, to interact with the data and uncover new business insights. But challenges exist, limiting this value and leaving opportunity on the table. Hewlett Packard Enterprise has developed a solution that solves these challenges through a robust, yet flexible, offering to deliver a future-proof data-centric foundation that scales with your evolving business needs.

## Today's business challenges

Data is the fuel that drives the idea economy. But to harness all your relevant data, you will need to leverage new engines to consume this fuel and deliver superior business outcomes. Over the last 10 years, Hadoop has become a critical engine and fuel tank, gaining a critical mass of developers, data scientists, and IT professionals as supporters. But while Hadoop has worked for some major visionaries, it has largely been limited to science experiments, simple workloads, and pilots for the vast majority of early adopters. Hewlett Packard Enterprise sees the challenges and opportunities presented by Hadoop and can empower your organization to turn Hadoop into an enterprise-grade cornerstone of your data-centric platform.

60%
HDFS performance[1]

100%
of your relevant data

Up to 8X
faster analysis[2]

| Customer challenges | Customer needs |
|---|---|
| **Inability for existing technologies to handle the data growth:** Most of the data available today was created in the last two years, often from new data sources and types, and the existing architectures, databases, and approaches to data management are reaching their limitations. A survey shows that 41 percent of business surveyed noted that their systems cannot process large volumes of data from different sources.[3] | **Consolidated, scalable, and affordable repository for Big Data:** Organizations need a centralized location via Hadoop to cost-effectively store and manage a wide range of data, from business data, free of the rigid structure of traditional databases to the wide range of structured, semi-structured, and unstructured data and leverage standard SQL as well as more advanced tools. |
| **Scaling Hadoop without compromise:** Organizations want to get more value out of their Hadoop investments and scale their data lakes. But, Hadoop doesn't include the analytics wherewithal and performance optimization for complex workloads, and the lack of robust security with increased scale increases risks. | **Enterprise-grade workload management, performance, availability, and security:** Flexible, yet robust, solutions leveraging proven infrastructure and analytics technologies that leverage ANSI SQL queries and results in the right amount of compute, storage, and analytics wherewithal based on the use cases. Security integrated with Hadoop that protects your data in motion, at rest, and in use. |
| **Determining how to get the value from Hadoop:** Organizations are struggling to identify the right projects and use cases to extract the most value out of Hadoop. The Hadoop framework enables many workloads and use cases, but organizations struggle to identify which ones to execute and assess the business impact and value. | **Partner that understands your business and is experienced with complex, heterogeneous IT environments:** Your organization needs to access and prioritize Hadoop use cases of proven value, and integrate it with existing applications and the larger business intelligence (BI) and analytics landscape. IT environments are difficult and complex, and require the partner to have the deep expertise in solving these challenges. |

## Value proposition

Hewlett Packard Enterprise understands what's required for you to not just survive but thrive in the Idea Economy. Ideas have to be converted to successful experiments then into full-scale business solutions. This two-stage conversion maps to a need to convert data into insights and then those insights into better business outcomes. Hewlett Packard Enterprise has the ability to partner with you to discover the value of your data. We can help you build a data-centric foundation to turn that data into insights, and then turn those insights into superior business outcomes at the speed of business, incorporating Hadoop as a cornerstone of your Big Data analytics solution.

[1] "**HPE Big Data Reference Architecture: A Modern Approach, 4AA5-6141ENW, October 2015, Rev. 2**."

[2] HPE internal study.

[3] "Capitalizing on the promise of Big Data," PWC paper, January 2013.

**Hadoop expertise:** Hewlett Packard Enterprise has optimized all aspects of the Hadoop environment to enable our customers to run multiple applications and consolidate multiple data stores into a single system. HPE reference architectures have been developed to support adaptation to new demands, allowing you to independently scale compute and storage, breaking the standard Hadoop local node compute and storage chain, as well as optimize traditional architectures. These advancements enable you to get the most out of YARN, Spark, Kafka, Mahout, and of course the core modules—Hive, HBase, MapReduce, and Hadoop Database File System (HDFS). Additionally, Hewlett Packard Enterprise has industry-leading software to deliver additional functionality over and above Hadoop through Vertica for SQL on Apache Hadoop, Vertica Enterprise Edition for predictive analytics, and a range of information governance, management, and data security suites to help ensure full enterprise-grade capabilities.

While Hadoop has quickly gained traction as a viable open source technology in the Big Data and analytics marketplace, as seen with the broader data revolution, a number of significant challenges have emerged. At scale and with proper security and reliability, Hadoop implementations present very complex planning, deployment, and long-term management challenges. There is currently a general lack of Hadoop skills in the marketplace. The challenges presented by this scarcity of skills are further compounded by the Hadoop technology stack's continued evolution and maturation, imposing higher degrees of difficulty and uncertainty.

Hewlett Packard Enterprise has a proven, four-step approach to the implementation of sophisticated data or analytic systems—discovery, development, integration, and implementation that has supported a global set of customers, large and small, in optimizing the value of Hadoop to achieve superior business outcomes.

**Analytic talent with global reach:** Hewlett Packard Enterprise implements mission-critical solutions for clients through our global footprint of more than 3,500 analytics experts, averaging over 18 years of analytics expertise, across nine analytics centers of excellence in four continents.

**Integrated solutions:** Our powerful portfolio of open standard products, open architectures, and services along with deep expertise helps you integrate and embed analytics into standard processes and workflows to help you turn insights into actions without locking you into inflexible proprietary solutions with diminishing returns over time.

1200+
global analytics professionals

18+
years of analytics experience

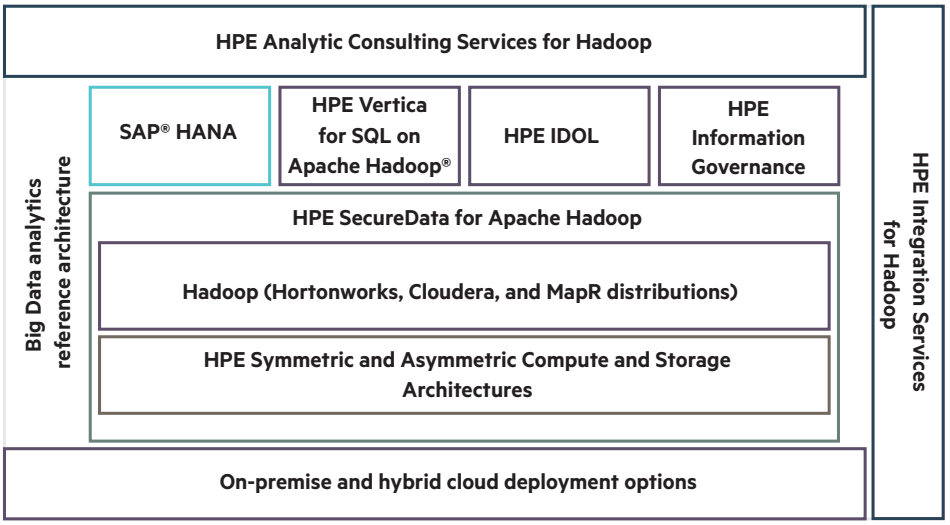


**Figure 1:** Solution overview

# Enterprise-grade Hadoop

**There is growing consensus that a data-centric foundation to handle TB–PB range Big Data across all Big Data types (machine, human, and traditional business data) will invariably have Hadoop as a core component. HDFS will be a data lake destination for data extracted from existing schema-based databases, but will also be the data lake destination for a range of new less structured and more varied data sources including social media, machine logs, and other new Big Data sources. Hewlett Packard Enterprise builds on the open source to enable an enterprise-grade Hadoop.**

| Solution components | Description |
|---|---|
| **HPE as a Service Solution for Hadoop** | HPE As a Service Solution for Hadoop is an enterprise-class, open source-based Big Data platform. It's an as-a-service, cloud-based deployment, and consumption model—in addition to our hosted and on-premise solutions. Our solution provides clients with a configured Hadoop platform in the HPE Helion cloud, without high upfront costs, high implementation risks, burden of IT resources, and "shelfware"—enabling you to focus on reaching your business goals. Available in tiered workload options, it offers the flexibility to adapt to change while reducing technology obsolescence risks.<br>The HPE As a Service Solution for Hadoop can be extended to add the analytics database capability of HPE Vertica to handle large data sets in near real time, and includes IDOL to process all forms of unstructured data. |
| **HPE Analytic Consulting Services** | HPE Analytic Consulting Services can help you determine the value of your data in hours or days instead of weeks or months. We will then help you to identify which are the appropriate components of a data-centric foundation, how to operationalize the analytics on top of that foundation at the point-of-action in your business process, and help you implement it. We will help your organization determine the optimal combination of existing BI/Enterprise Data Warehouse (EDW) investments, balance investment in real time vs. batch analytics, and Hadoop as a data lake and analytics platform. |
| **Hadoop** | Hadoop is an Apache open-source, modular software framework for Big Data management and analytics capable of scaling out to petabytes of data on open standard hardware. |
| **Data-centric foundation Hadoop integration services** | HPE Discovery and Transformation Services for Hadoop will help you to develop: a formal and structured approach to data and analytic discovery including a structured methodology, use of specialized data visualization and sharing tools as part of an integrated platform, improve the skilled of your resources and collaboration methods, and best practices to support experimentation that is both ambitious and cost-effective.<br>HPE Integration Services for Hadoop helps you optimize and modernize your infrastructure, and handles your additional system requirements to bring enterprise-grade capabilities to your Hadoop platform. |
| **HPE IDOL 10** | With HPE Intelligent Data Operating Layer (IDOL), you can access, analyze, understand, and act on petabyte ranges of human information stored in your HDFS data lake—contextually and in multiple languages—from virtually any source, including both cloud and on-premise, including 500 analytical functions to apply on text, image, audio, video, social media, and structured data in databases. IDOL contains out-of-the-box connectors to more than 1,000 standard file formats and 400 application and database interfaces. IDOL can be used with Hadoop to categorize, index, and make sense of your Hadoop data lake. |
| **HPE Vertica for SQL on Apache Hadoop** | HPE Vertica for SQL on Apache Hadoop offers you the most enterprise-ready way to perform SQL queries on your Hadoop data by providing you with full ANSI SQL syntax on MapR, Hortonworks, and Cloudera. With HPE Vertica for SQL on Apache Hadoop, your business analysts, data scientists, and developers get out-of-the-box integration and access to your traditional BI tools and the latest in industry-standard analytics and visualization tools. HPE Vertica offers enterprise-ready, advanced analytics that support full ANSI SQL, Atomicity, Consistency, Isolation, Durability (ACID) compliance, complex data types, and other capabilities only available from our SQL on Hadoop implementation. It offers optimizations like compression, columnar storage, and projections, delivering performance levels far surpassing standard SQL on Hadoop. |
| **HPE Reference Architectures for Hadoop** | To reduce risk and optimize human, financial, and data center resources, it is imperative that you employ the right architecture to support your particular Hadoop use cases: Hive, SQL, Spark's in-memory processing, Kafka streaming data, Mahout machine learning, etc. Hewlett Packard Enterprise has a multitude of reference architectures supporting workload optimization for Hadoop based on combinations of Moonshot, HPE Apollo 2000 and 4000 Gen9 series, and ProLiant XL series servers that balance compute and storage requirements for Hadoop clusters with all three Hadoop distributions: MapR, Hortonworks, and Cloudera. The Minotaur reference architecture extends the Hadoop-only reference architectures to incorporate SAP HANA, HPE Vertica, and HPE IDOL. |
| **HPE Financial Services** | Balancing your new Hadoop data lake and analytics additions with your existing infrastructure and environment requires financial flexibility. Financing helps you respond and adapt to changes throughout the implementation process, lets you fuel innovation, and align costs with use. In addition, HPE Financial Services may help you reduce reputational and regulatory risk associated with the disposal of legacy equipment, through HPE Asset Recovery Services. |
| **HPE SecureData** | HPE SecureData is a data-centric approach to security for Hadoop, encrypting and tokenizing the data, de-identifying it as close to its source as possible, transforming the sensitive data elements with usable, yet de-identified, equivalents that retain their format, behavior and meaning. This protected form of the data can then be used in subsequent applications, analytic engines, data transfers, and data stores, while being readily and securely re-identified for those specific applications and users that require it. HPE SecureData protects virtually unlimited number of data types at the data level. |

Optimize your IT investment strategy with new ways to acquire, pay for and use technology, in lock-step with your business and transformation goals.
**hpe.com/solutions/hpefinancialservices**

## What's your next step?

**Schedule a Transformation Workshop.** Talk to your HPE representative about signing up for a Transformation Workshop.

Learn more at
**hpe.com/us/en/solutions/empower-data-driven.html**

**Sign up for updates**

Hewlett Packard Enterprise



4AA6-3745ENW, August 2016, Rev. 3