



Highlights

- **Superior performance and flexibility** — Up to 1.7 times performance advantage versus competing x86-based solutions.¹ Right-size to your workload with an unmatched range of Linux[®] server options.
- **Maximum efficiency** — Reduce storage infrastructure requirements by three times compared to HDFS when integrating with IBM[®] Elastic Storage[™] Server with IBM Spectrum Scale[™] erasure coding.²
- **Ready for cognitive** — Allows seamless integration of machine and deep learning using IBM's PowerAI deep learning platform.
- **Committed to client success and open innovation** — Complete, enterprise-ready solution built on open hardware and software technology that is fully tested and offered with industry-leading support and expertise.

Hortonworks Data Platform on IBM Power Systems

Secure, enterprise-ready open source Apache Hadoop distribution for the leading open server for Big Data Analytics and artificial intelligence.

Hortonworks Data Platform (HDP) on IBM Power Systems[™] delivers a superior solution for the connected enterprise data platform. With industry-leading performance and IT efficiency combined with the best of open technology innovation to accelerate big data analytics and artificial intelligence, organizations can unlock and scale data-driven insights for the business like never before.

Hortonworks Data Platform

An industry-leading, secure and enterprise-ready open source Apache Hadoop distribution, HDP addresses a range of data-at-rest use cases, powering real-time customer applications and delivering robust analytics to accelerate decision-making and innovation.

HDP uses the Hadoop Distributed File System (HDFS) for scalable, fault-tolerant big data storage and Hadoop's centralized Yet Another Resource Negotiator (YARN) architecture for resource and workload management. YARN enables a range of data processing engines including SQL, real-time streaming and batch processing, among others, to interact simultaneously with shared datasets, avoiding unnecessary and costly data silos and unlocking an entirely new approach to analytics.

An open and flexible data platform, HDP includes a comprehensive set of capabilities including data access, governance and integration, along with security and operations management. HDP's open source community development model allows your organization to take advantage of rapid innovation and deep integration across the enterprise.



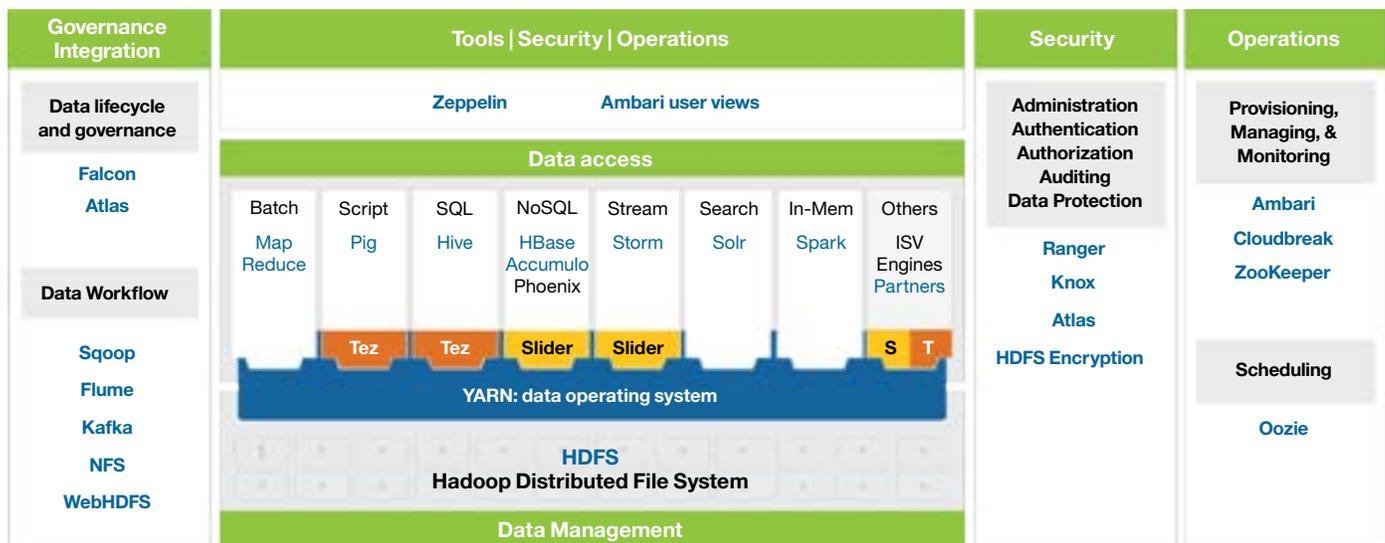


Figure 1: Hortonworks Data Platform

Hortonworks Data Platform on Power Systems — better together

Power Systems with IBM POWER8® processors and differentiated hardware acceleration technology are designed to deliver breakthrough performance for big data analytics workloads. The POWER8 processor delivers industry-leading performance for big data analytics applications running on HDP, with multi-threading designed for fast execution of analytics (eight threads per core), multi-level cache for continuous data load and fast response (including an L4 cache) and a large, high-bandwidth memory workspace to maximize throughput for data-intensive applications.

IBM Power Systems OpenPOWER LC server family

Designed for flexibility and seamless integration to existing clusters and clouds, the new IBM OpenPOWER LC server family offers the data-crushing POWER8 processor in a range of purpose-built system configurations, from compute-dense to storage-rich. The LC family's innovative design in partnership with the OpenPOWER Foundation offers hardware accelerator-offload for compute, storage and

networking workloads—for incredible speed-ups to analytics and massive efficiencies in data movement.

OpenPOWER brings the leading processor together with the best of our partners and end users across the ecosystem—from High Performance Computing installations, enterprise IT and hyperscale data centers and to system designers worldwide.

The HDP on IBM Power Systems reference configuration suggests options for a solution that's sized to your specific needs, built with a combination of the high-performance, storage-rich Power Systems S822LC for Big Data and the lightweight, yet compute-powerful Power Systems S821LC. See Figure 2 for further detail.

Superior performance for Apache Hadoop and Spark Workloads

HDP on Power Systems delivers more data faster, enabling valuable analytics for better and quicker decision making. In IBM performance testing for typical Apache Hadoop workloads, HDP on Power Systems versus x86-based solutions demonstrated 70 percent more queries per hour based on an average response time and 40 percent reduction on average in query response time.¹

	System Management Node			Worker Node		
	Master Node	Edge Node		Balanced	Performance	Storage Dense
Server Type	1U S821LC	1U S821LC	1U S821LC	2U S822LC	2U S822LC	2U S822LC
Count (Min / Max)	1 / 1	3 / Any	1 / Any	8 / Any	8 / Any	8 / Any
Cores	8	20	20	22	22	11
Memory	32GB	256GB	256GB	256GB	512GB	128GB
Storage – HDD	2x 4TB HDD	4x 4TB HDD	4x 4TB HDD	12x 4TB HDD	8x 6TB HDD	12x 8TB HDD
Storage – SSD					+ 4x 3.8TB SSD	
Storage Controller	Marvell (internal)	LSI MegaRAID 9361-8i (2GB cache)				
Network – 1GbE	4 ports (internal)	4 ports (internal)	4 ports (internal)	4 ports (internal)	4 ports (internal)	4 ports (internal)
Network – 10GbE	2 ports	2 ports	2 ports	2 ports	2 ports	2 ports

Figure 2: HDP on IBM Power Systems reference configuration

In addition to enabling faster Hadoop workloads, the POWER8 processor's leading thread density, large cache and memory bandwidth and superior I/O capabilities are a great match for in-memory Apache Spark workloads including SQL, streaming, graph and machine learning analytics.

IBM Spectrum Scale and Elastic Storage Server

An integrated storage system running IBM Spectrum Scale software on IBM Power Systems, IBM Elastic Storage Server can serve as the underlying storage for HDP. IBM Spectrum Scale is a software-defined storage system based on a parallel file system architecture that provides File (NFS, SMB, POSIX) and Object (S3, Swift) access and supports HDFS APIs. Support for HDFS APIs enables in-place analytics on enterprise storage instead of copying data from enterprise storage to analytics silos. In-place analytics not only eliminates

duplication of data but also avoids the problems of running analytics on stale data. In addition, Spectrum Scale provides shared storage to HDP, which allows for de-coupling of compute and storage to enable optimized configurations.

Elastic Storage Server provides a consolidated storage solution that can store a wide variety of data types and a range of applications with standard access methods, creating a dynamic, shared data ocean that scales capacity and performance with demand. A shared data lake (a storage repository that holds a vast amount of raw data in its native format until it is needed) allows for the same data to be shared across different application domains and global locations while helping reduce the need for data movement and copies. This helps save significant costs for storage, floor space and administration.

Ready for Cognitive with PowerAI

HDP on IBM Power Systems clients can seamlessly integrate with IBM's fully optimized and supported PowerAI platform for deep learning. Optimized for blazing performance on the Power Systems S822LC for HPC with NVIDIA NVLink Technology and Tesla P100 GPUs, PowerAI includes the most popular deep learning frameworks, pre-compiled and easily deployed for maximum-throughput, scalable deep learning on your connected data.

Commitment to client success

HDP on IBM Power Systems and Elastic Storage Server is fully tested, validated and supported by Hortonworks and IBM who provide deep industry expertise and dedicated commitment to client success. IBM and Hortonworks are leading supporters of the open source Apache Hadoop and Spark communities, driving innovation and advancements in the big data ecosystem. In addition, IBM and Hortonworks are founding members of an open data platform initiative, ODPi, a nonprofit organization committed to simplification and standardization of the big data ecosystem with common reference specifications and test suites.

IBM and Hortonworks partner to offer a wide range of complementary solutions to support a data lake based on HDP and Power Systems. For example, use [Attunity Replicate](#) to capture change data effectively from a wide range of enterprise data sources to populate the data lake, use [Zettaset BDEncrypt](#) to add proven enterprise-class data privacy and protection or add [Synerscope](#) for advanced high speed data visualization.

For more information

To learn more, contact your IBM representative or visit: ibm.biz/hortonworksOnPower.



© Copyright IBM Corporation 2017

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
April 2017

IBM, the IBM logo, ibm.com, IBM Elastic Storage Server, IBM Spectrum Scale, POWER8 and Power Systems are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or TM), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product or service names may be trademarks or service marks of others.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

1 Performance results are based on preliminary IBM Internal Testing.

2 Up to three time reduction in storage is due to HDFS making two additional copies for data protection and availability while IBM ESS only requires 30 percent overhead. Additional copies can be avoided with ESS by sharing the same data over NAS and Object protocols with other enterprise applications.



Please Recycle