Hortonworks Data Platform on IBM Power Systems for Financial Services

Enabling financial services innovation in a complex, data-driven world

Highlights

- Cut across customer data silos to help create a complete 360-degree view of the customer

- Turn data from multiple sources into complete and accurate views for regulatory reporting

- Reduce costs by archiving unused data, offloading low-value workloads, and enhancing data quality

- Deploy your data platform on open hardware and software, backed up by industry-leading support and expertise

- Pair Hortonworks Data Platform with IBM® Power Systems™ servers and IBM Spectrum Scale™ file system to take advantage of performance, speed, scalability and efficiency benefits

The business world is defined by data, and financial services is no exception. Across the industry, leaders are realizing how important data is to understanding customers, ensuring compliance, and keeping up with disruptive competitors.

However, many organizations have difficulty extracting value from data, due to systems that are fragmented along geographic and business lines. These organizations often have data platforms cobbled together over time from disparate pieces. With no end-to-end coherence, these platforms can't meet the needs of big data sources.

Financial services organizations need a more coherent data platform to become data-driven businesses. Hortonworks Data Platform on IBM Power Systems with IBM Spectrum Scale can be the solution they need.

Hortonworks Data Platform on IBM Power Systems unlocks value from data

Hortonworks Data Platform (HDP) is a secure, enterprise-ready Apache Hadoop distribution that supports a variety of data use cases. With a flexible design, HDP empowers businesses to deploy across different data centers, utilizing a 100 percent open-source approach that helps drive innovation and prevent vendor lock-in.

When supported by Power Systems and Spectrum Scale, HDP provides the clarity and visibility that today's disjointed data systems lack. This allows financial services organizations to:

- Do away with complex webs of enterprise and transaction systems connected by point-to-point feeds
- Build a single enterprise-wide data lake
- Enjoy end-to-end visibility and control across all enterprise data
- Start taking advantage of the latest big data capabilities, including data discovery, artificial intelligence, machine learning, and predictive analytics

This paper will show how getting the best results from HDP requires running the right platform. HDP on Power Systems with Spectrum Scale can help financial services organizations achieve speed, performance, and cost efficiency for any big data use case.

Embarking on a digital, data-driven journey

Many financial services organizations adopt HDP as the first step on their digital transformation journey. By gradually replacing disjointed legacy data systems, you can deploy new use cases that draw value from data to decrease costs, address complexity, and better understand digital and customer data.

The use cases outlined here represent steps on your way to becoming a data-driven business. HDP on Power Systems can support each of these uses cases equally across all lines of business, while drawing from the same centralized data lake. Refer to Figure 1 for a sampling of the financial services use cases enabled by HDP.

360-degree view of the customer

Success in financial services depends on understanding who your customers are and what they want; you need data to build this understanding. The amount of data customers generate across channels means that the insights you need are out there, if you know where to look.

When it comes to understanding customers, the limits of traditional data platforms are clear: they store interaction data in silos, making it difficult to see how the pieces fit together. They are also over-reliant on structured relational databases. Companies using these systems are likely missing out on the insights available in unstructured data.

In contrast, HDP on Power Systems and Spectrum Scale lets you ingest and enrich data from different sources. You can then use machine learning, drawing from the high performance of Power Systems, to process that data quickly. In fact, when compared with x86 systems, Power Systems reduces average query response times by 40 percent.[1]
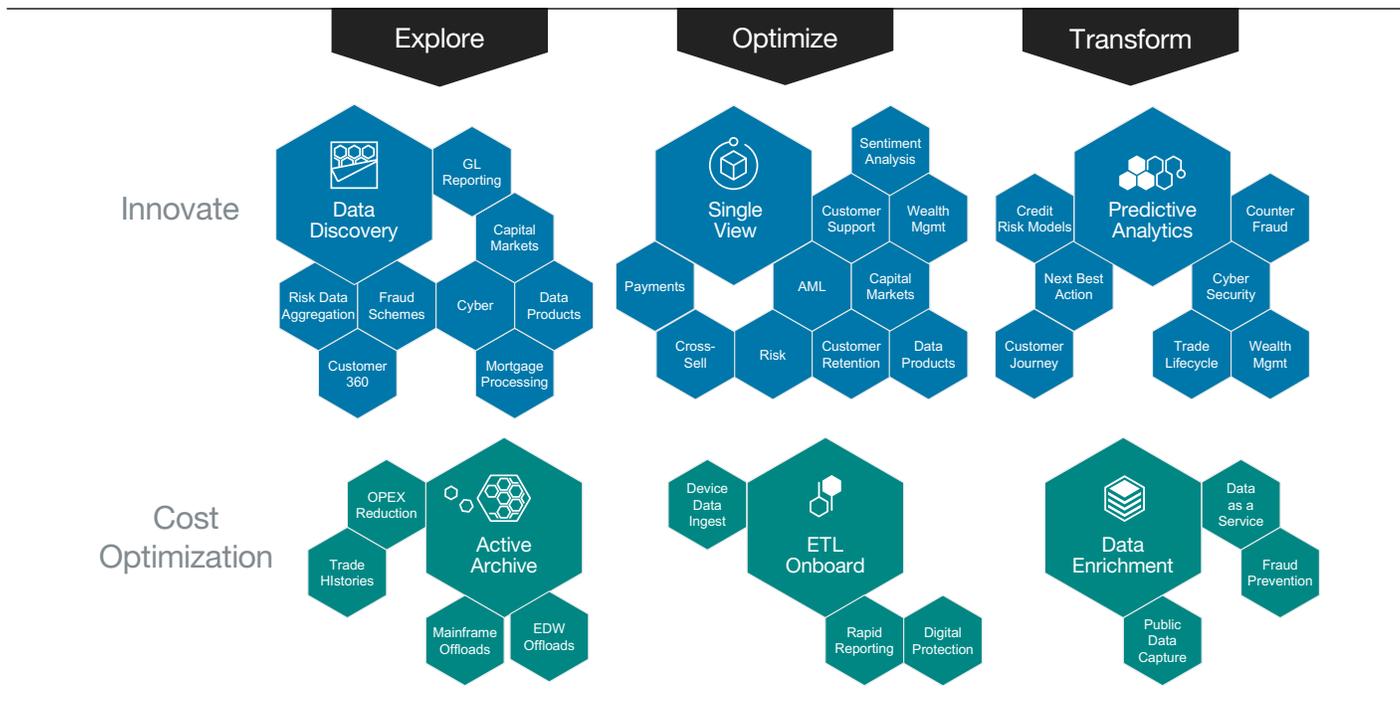

Figure 1: HDP use cases for the financial services industry

The speed of Power Systems helps make customer insights available to agents in near real-time. With these insights, agents can make informed decisions about the risks and opportunities a customer presents. For instance, they can present cross-sell and up-sell offers to some customers, or undertake retention measures with others.

Building a centralized data lake with HDP on Power Systems with Spectrum Scale can give you a 360-degree view of the customer, allowing you to understand them better, reach them across channels, develop micro-segmentation for a more targeted approach, and enable the creation of value-based metrics. In short, it's the first step toward uncovering new insights, sharing trends across the organization and building the most positive customer experience possible.

Risk and regulatory reporting

Today's financial services organizations operate in a very stringent regulatory environment. With Dodd-Frank, Basel III, BCBS 239, GDPR, and others, the compliance requirements organizations face are intense, and the risks of not complying are considerable.

To address these requirements and risks, you need a strong, centralized platform for ingesting data, processing it, adding it to a canonical data set, performing what-if scenarios and analyses on it, and finally, creating reports with the results of the analyses.

HDP is the only data platform that can perform these steps while also maintaining original data sets. This provides end-to-end traceability throughout the reporting process: if a regulation requires you to refer to your original data sets, you'll be able to do so quickly, with full data auditability and lineage.

Whether it's creating reports or accessing historical data records, organizations need to move quickly in order to meet compliance requirements and limit risk. Once again, Power Systems delivers, supporting 70 percent more queries per hour than x86 systems.[2]

Cost optimization

Employing HDP on Power Systems with Spectrum Scale to optimize your data architecture can provide many cost benefits. This is largely due to the fact that the typical enterprise data warehouse dedicates 70 percent of its storage volume to unused data, and 55 percent of its processing power to low-value extract, transform, load (ETL) workloads.[3]

Financial services organizations have the opportunity to increase efficiency and lower costs by implementing Hadoop. HDP users can modernize their data architecture in three ways:

- **Active archive:** Moving data from legacy EDWs into an HDP-based data lake allows you to store data longer, while decreasing operational costs. This can be particularly beneficial for data that isn't currently in use.
- **ETL onboarding:** By moving ETL workloads to HDP, organizations can free up processing power to dedicate to high-value analytics work. In turn, this can enable rapid reporting.
- **Data enrichment:** HDP can rapidly take in new data in a variety of structures, enrich that data, and make it available to the EDW. As a result, the EDW can benefit from a source of business value that might have gone untapped otherwise.

Data lakes can often grow to petabyte scale, resulting in the inefficient use of servers to hold all that data. By deploying IBM Elastic Storage™ Server (ESS) and Spectrum Scale as their storage solution, financial services companies can reduce storage infrastructure requirements by three times, when compared with Hadoop Distributed File System (HDFS).

About IBM Power Systems

With Power Systems, financial services organizations can build the ideal foundation for the use cases outlined in this paper.

These use cases depend on technologies like cognitive computing, artificial intelligence, machine learning, and predictive analytics to draw high-value insights from data. The opportunities these technologies present are clear, but there are also challenges: the level of performance these technologies require places added strain on IT infrastructure.

Power Systems servers deliver the outstanding performance needed to take advantage of big data applications on HDP. With simultaneous multithreading with eight threads per core (SMT-8), Power Systems offers rapid execution of analytics workloads. In addition, its multi-level cache (including an L4 cache) helps support continuous data load and fast response, while a large, high-bandwidth memory workspace drives higher data throughput.

About IBM Spectrum Scale

Spectrum Scale is flexible and scalable software-defined file storage for analytics workloads. It can be purchased as software or as a part of the pre-integrated ESS system. It scales up to more than a billion petabytes of data and hundreds of GB/s throughput.

Spectrum Scale has the most comprehensive support for data access protocols for in-place analytics (NFS, SMB, Object, POSIX and the HDFS API). This eliminates the need to maintain separate copies of the same data for traditional applications and analytics software. Using Spectrum Scale's global namespace, financial services organizations can create active, remote data copies to form data lakes across the globe.

Another area of concern in financial services is ensuring data security. Spectrum Scale meets this need with filesystem encryption for secure data at rest, a Hadoop transparency connector for secure data in transit, and secure backup and deletion.

About the IBM/Hortonworks partnership

IBM and Hortonworks are committed to bringing deep expertise and commitment to client success to financial services. Both companies are major supporters of open source projects like Hadoop and Spark, and both help drive innovation in big data technology.

Our partnership provides what financial services organizations need to start becoming data-driven businesses. By doing so, they can serve customers better, keep up with complex regulatory requirements, and innovate to stay ahead of the competition.

For more information

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

1 Performance results are based on IBM Internal Testing.

2 Ibid.

3 Hortonworks, "Optimize your data architecture with Hadoop." https://hortonworks.com/blog/optimize-your-data-architecture-with-hadoop.

Please Recycle