DATASHEET

Hortonworks Data Platform (HDP®) 3.0 — Faster, Smarter, Hybrid Data

Companies looking at new, innovative technologies to retool business processes and better serve their customers often encounter serious hurdles along the way. These include building and deploying new applications quickly, accessing the data timely, and scaling to support the massive volumes of data generated.

In addition, the rise of machine learning and deep learning apps have brought processing challenges — with new use cases in manufacturing, insurance, and other industries — along with increasing pressure at an affordable price. Nevertheless, it is now crucial for businesses to seize these opportunities for growth and efficiency, and to take full advantage of the availability of cheaper data storage, distributed processing, and more powerful computers.

Below highlight the key challenges faced by enterprises today in their digital transformation journey:

• Flexibility and agility to build and deploy data-intensive apps quickly: developers want faster time to deployment, in minutes rather than hours or weeks

• Infinite scalability to support billions of files and thousands of nodes

• Support for deep learning apps and workloads including TensorFlow and Caffe

• Enterprise security and governance requirements for data lakes

• Real-time database (EDW) to optimize queries for both historical and real-time data, on-premises and in the cloud

• ISV strategy to integrate third-party apps including IBM Data Science Experience (DSX)
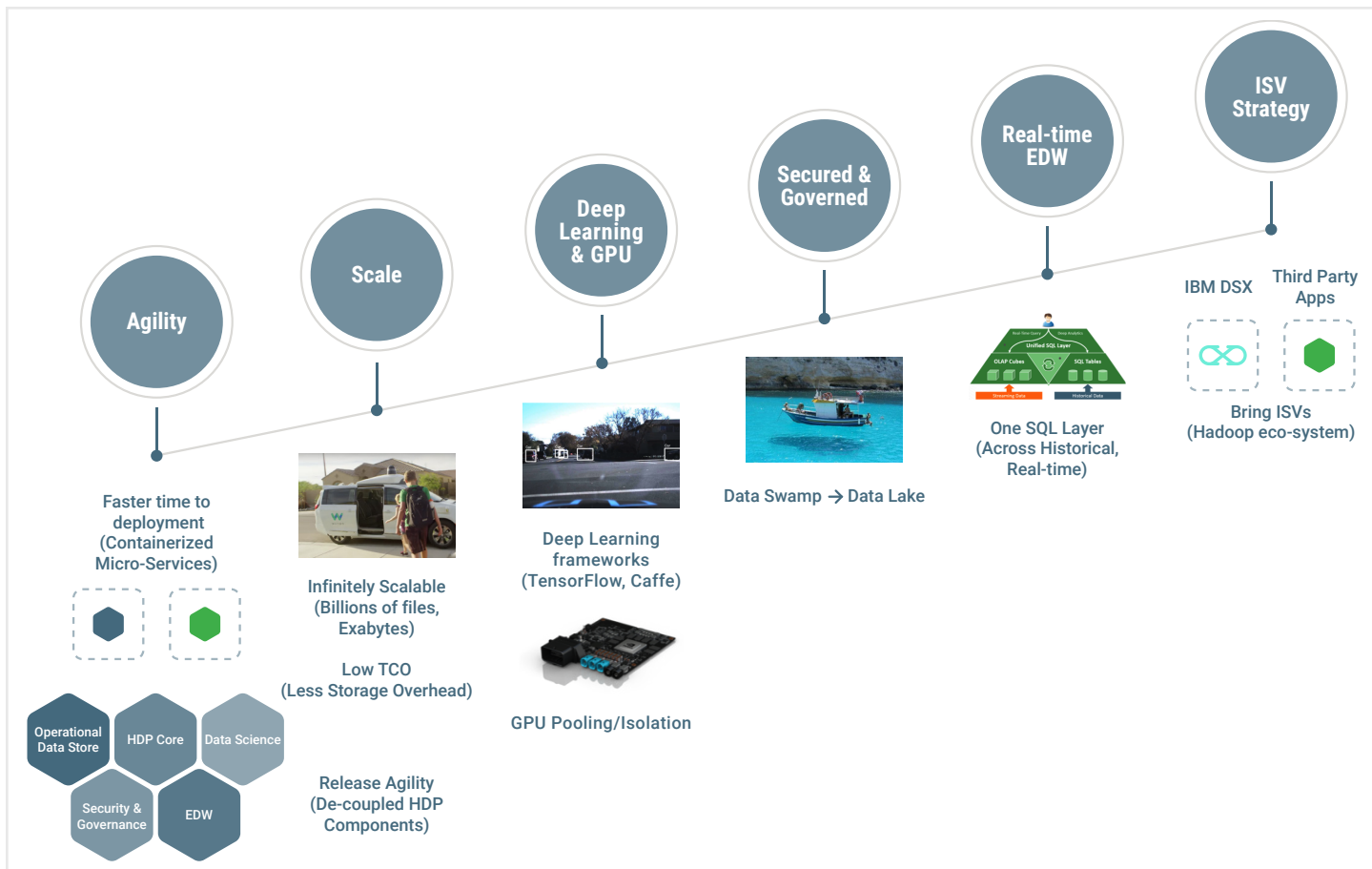
Figure 1: Market drivers

SOLUTION

Hortonworks Data Platform (HDP) 3.0 delivers significant new features, including the ability to launch apps in a matter of minutes and address new use cases for high-performance deep learning and machine learning apps. In addition, this new version of HDP enables enterprises to gain value from their data faster, smarter, in a hybrid environment.

KEY BENEFITS AND VALUE

FASTER: Agile app deployment from containerization allows customers to build data-intensive apps quickly, rolling them out in minutes and decreasing time to market. HDP 3.0 is your "cleanest" path to containerization.

SMARTER: Deep learning app support via graphics processing unit (GPUs) to make intelligent decisions empowering customers to run GPU-loving workloads, such as machine learning and deep learning apps. GPU support can accelerate the time to insight from months to days.

HYBRID: Modern hybrid data architecture includes cloud storage support to store endless amounts of data in its native format including ADLS, WASB, S3, GCP (tech preview). Data-at-rest and data-in-motion support on-premises and in the cloud.

BIGGER: Scalable with NameNode federation, allowing customers to scale to thousands of nodes and a billion files, while providing higher availability with multiple name nodes, at significantly lower TCO with Erasure Coding

REAL-TIME DATABASE: Integration of Apache Hive offers the only unified solution to provide interactive query at scale — whether the data sits on-premises or in the cloud.

DATA SCIENCE: Performance improvements around Apache Spark and Apache Hive integration, Spark connectors to other features including cloud and containerized TensorFlow tech preview combined with GPU pooling

SECURE: Comprehensive security and governance from Apache Ranger and Apache Atlas provide the ability to track the lineage of data from its origin to the data lake, allowing auditors to follow data through the entire enterprise.

KEY CAPABILITIES

Containerization provides YARN support for Docker containers, allowing 3rd party applications to run on Apache Hadoop (e.g., containerized applications), enabling:

- Faster time to deployment by enabling third-party apps

- The ability to run multiple versions of an application, enabling users to rapidly create features by developing and testing new versions of services without disrupting old ones

- Improved resource utilization and increased task throughput for containers, yielding faster time to market for services

GPU pooling and isolation ensures a first-class resource type in Hadoop, making it easier for customers to run machine learning and deep learning workloads.

- Compute-intensive analytics require not only a large compute pool, but also a fast and expensive processing pool with GPUs in tandem

- Customers can share cluster-wide GPU resources without having to dedicate a GPU node to a single tenant or workload

- GPU isolation dedicates a GPU to an application so that no other application has access to that GPU

Erasure coding offers a data protection method that until now has mostly been found in object stores.

- Hadoop 3.0 will no longer default to storing three full copies of each piece of data across its clusters (only 1.5 copies needed for data recovery)

- Boosts storage efficiency by 50%, allowing more efficient data replication

- **NameNode federation** protects data in case of failures or disaster recovery.

- Allows scaling up to thousands of nodes and billions of files

- Supports multiple standby NameNodes (if one goes down, cluster can continue to operate)

WHAT MAKES HDP 3.0 UNIQUE?

FASTER

- Fastest time to deployment for data intensive containers apps

SMARTER

- High-performance compute for deep learning and machine learning apps via GPU support

HYBRID

- Cloud storage support to store endless amounts of data in its native format including Amazon S3, ADLS, WASB

SCALABLE

- Scalability with NameNode federation, allowing customers to scale to a billion files and thousands of nodes

REAL-TIME DATABASE

- SQL data access enhancements and improved query performance to eliminate the gap between low latency and high throughput queries

SECURE

- Greater regulatory compliance (including GDPR) through full chain of data custody as well as fine-grained auditing of events

About Hortonworks

Hortonworks is a leading provider of enterprise-grade, global data management platforms, services and solutions that deliver actionable intelligence from any type of data for over half of the Fortune 100. Hortonworks is committed to driving innovation in open source communities, providing unique value to enterprise customers. Along with its partners, Hortonworks provides technology, expertise and support so that enterprise customers can adopt a modern data architecture. For more information, visit hortonworks.com.

Contact

For further information,
visit hortonworks.com

+1 408 675-0983
+1 855 8-HORTON
INTL: +44 (0) 20 3826 1405

HORTONWORKS®