



THE ENTERPRISE DATA WAREHOUSE OPTIMIZATION SOLUTION

Joint offering from Hortonworks, Syncsort and AtScale

The Enterprise Data Warehouse Optimization Solution

UPGRADE YOUR ENTERPRISE DATA WAREHOUSE FOR THE DEMANDS OF BIG DATA

The era of Big Data places unprecedented demands on data architecture. Powered by Hortonworks Data Platform (HDP®), and our partners Syncsort and AtScale, the Enterprise Data Warehouse optimization solution lets you overcome challenges of cost, complexity and expansion while extending the value of your existing enterprise data warehouse (EDW).

- **Cost** – As organizations seek to capture and use more data, data volumes continue to skyrocket and efficiency of storage capacity becomes a key factor. The EDW Optimization solution makes it possible to reduce hardware and software costs while ensuring that all data can be accessed and analyzed reliably.
- **Complexity** – The integration of new data sources allows you to deliver more powerful analytics and insights for the business—but it also increases data movement and transformation steps. The EDW Optimization solution lets you shift data extraction, loading and transformation from your EDW to Apache Hadoop with schema-on-demand for both known and unknown structures.
- **Expansion** – The structure and scale of the new Big Data domain can make it difficult to apply existing tools, and risky to enable agile functionality for traditional systems. The EDW Optimization solution provides a mechanism to ingest, store and process any type of structured or unstructured data within an integrated ecosystem.

In this way, Hortonworks along with its partners Syncsort and AtScale allows you to cut the costs associated with your data architecture while extending new analytic value to the enterprise – with the scalability to meet the needs of new data sources, advanced workloads and emerging data science methods.

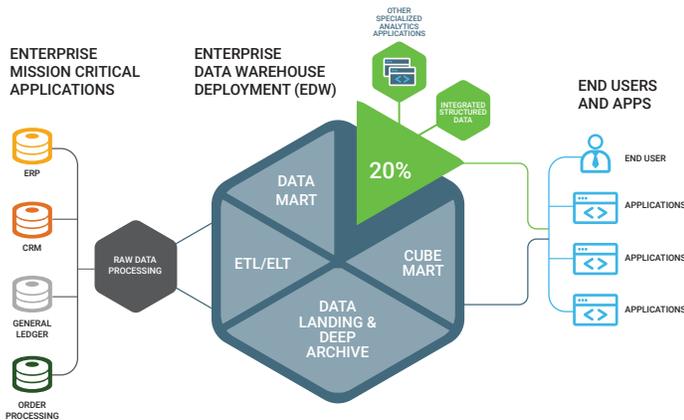


Figure 1: Traditional EDW Deployment

ADDRESS KEY BIG DATA USE CASES

- **Fast BI on Hadoop** – Store, process and feed new types of data, from your existing EDW to unlock new analytic value.
- **Onboard** – Shift ETL functions from your EDW to Hadoop to leverage compute and operational efficiencies—and free EDW capacity for high-value analytics.
- **Archive** – Move cold or less frequently used data into Hadoop storage and access it on-demand so you can store more data, longer, while controlling costs.

SOLUTION ELEMENTS

- **Hortonworks Data Platform** – Powerful open Hadoop capabilities for data governance and integration, data management, data access, security and operations—architected for deep integration with your existing data center technology.
- **Syncsort** – High-performance ETL software to access and easily onboard traditional enterprise data to HDP.
- **AtScale** – The business interface for fast business intelligence (BI) on Hadoop to bridge the gap between business users and their data.
- **Professional Services** – Expert guidance and support to quickly prove the value of your new architecture and maximize the value of the full tested and validated Hortonworks data architecture optimization solution.

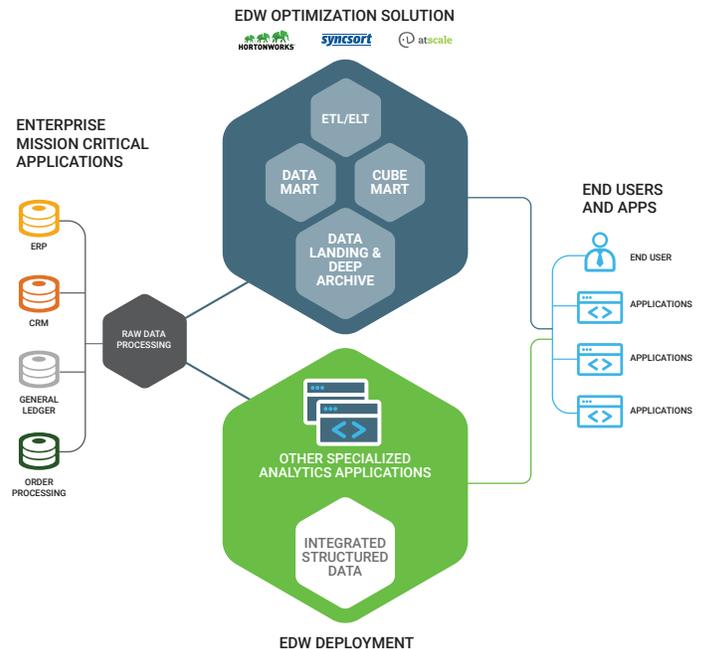


Figure 2: The EDW Optimization solution by Hortonworks, Syncsort and AtScale

Three Primary Use Cases For The EDW Optimization Solution

The EDW Optimization solution make it possible for your organization to store and analyze data of all types, from any source, with the performance, scalability and cost efficiency your business demands. Here are the three major use cases:

FAST BI ON HADOOP

Proprietary EDW systems were adopted for fast BI and deep slice-and-dice analytics, but EDW prices are unsustainably high and these systems have not adapted to modern big data challenges like unstructured data and large-scale analytics.

Hortonworks makes fast BI on Hadoop a reality, with the combination of a fast in-memory SQL engine to create data marts with an OLAP cubing engine that lets you query huge datasets in seconds. This gives you the choice of querying pre-aggregated data for maximum performance or in full-fidelity form when the finest grains of detail are needed, allowing access from any major BI tool that supports ODBC, JDBC or MDX.

ONBOARD ETL PROCESSES TO HADOOP

A typical EDW spends between 50–90 percent of its CPU cycles on ETL processing. These lower-value ETL jobs compete for resources with more business-critical workloads and can cause SLA misses. Hadoop can offload these ETL jobs with minimal porting effort and at substantially lower cost, saving money and freeing up capacity on your EDW for higher-value analytical workloads. Hortonworks makes it easy by providing high-performance ETL tools, a powerful SQL engine and integration with all major BI vendors.

ARCHIVE DATA IN HADOOP

Increasing data volumes and cost pressures force many companies to archive old data to tape where it can't be analyzed or accessed.

Hadoop offers cost per terabyte on par with tape backup solutions. Because of the appealing cost, you can store years of data rather than months. All of your enterprise data remains available for retrieval, query and deep analytics with the same tools you use on existing EDW systems.

Complementary Partner Technologies

The Enterprise Data Warehouse Optimization solution incorporates integrated partner technologies to extend its value for customers. Tested and validated under the Hortonworks Certified Technology Program, these capabilities help form a complete, production-ready solution for Big Data insight.

SYNCSORT DMX-H HIGH-PERFORMANCE ETL SOFTWARE

Designed from the ground up for Hadoop, Syncsort DMX-h simplifies big data integration with everything you need to access and integrate all your enterprise data on Hortonworks Data Platform. Syncsort integrates with Hadoop and HDP directly through YARN, making it easier for users to write and maintain MapReduce and Spark jobs graphically. An intuitive design interface and a Hadoop native runtime engine optimize the performance and efficiency of data integration processes while simplifying their creation and maintenance of these processes.

- **Access** — Ingest data from sources including mainframes, RDBMS, MPP, JSON, Avro/Parquet, NoSQL and more. Import hundreds of tables to HDFS in a single job.
- **Integrate** — Design streaming and batch processes in a single interface with fast multi-data source join and sort capabilities.
- **Comply** — Improve security and governance through seamless integration with Kerberos and Apache Ranger, and by automatically updating HCatalog when loading to Hive, Avro and Parquet.
- **Simplify** — Mask underlying complexity and reduce the need for Big Data integration expertise with design-once, execute-anywhere.

ATSCALE BUSINESS INTELLIGENCE TOOLS

AtScale makes business intelligence work on Hadoop with speed, scale and simplicity. Business users can perform interactive and multi-dimensional analysis directly on their data in HDP using the BI tools they already own, while IT maintains data security and governance they need.

- **Query data where it lays** — Eliminate the cost, latency and complexity of repeated data movement by letting users access data in its original Hadoop files.
- **Use your BI tool of choice** — Leverage existing BI tool investment and user familiarity within the context of your next-generation Hadoop environment.
- **Simplify semantic layer** — Optimize queries with a consistent, secure view of data definitions across BI tools, languages and Hadoop engines.

BUILD A FOUNDATION FOR BIG DATA SUCCESS

Realizing the transformative potential of Big Data depends on your ability to manage cost and complexity while leveraging data sources of all types. The Hortonworks solution for data architecture optimization lets you complement your EDW with the power of Hadoop to enable a highly cost-efficient environment with optimal performance, scalability, and flexibility. By putting more of your data to work, more effectively, at a lower cost, you can provide insights on-demand throughout your organization for unprecedented business impact. To learn more about the Enterprise Data Warehouse Optimization solution, please visit <http://hortonworks.com/solutions/edw-optimization/>.

About Hortonworks

Hortonworks is an industry leading innovator that creates, distributes and supports enterprise-ready open and Connected Data Platforms and Modern Data Applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Hortonworks is focused on driving innovation in open source communities such as Apache Hadoop, Apache NiFi and Apache Spark. Along with its 1,800+ partners, Hortonworks provides the expertise, training and services that allow customers to unlock transformational value for their organizations across any line of business.

Contact

For further information visit
www.hortonworks.com

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

