



## PRODUCT SHEET

# Syncsort DMX-h

Simplifying Big Data Integration

## Goals of the Modern Data Architecture

---

Data warehouses and mainframes are mainstays of traditional data architectures and still play a vital role today. However, they are not designed to cost-effectively scale to handle the massive increases in data volumes, as well as the many new types of data sources that organizations need to analyze. In response, organizations are evolving their environments to a Modern Data Architecture to maximize the value of all their enterprise data – from both traditional and newer sources – to deliver meaningful insights to business decision makers as quickly as possible. Modern Data Architectures adhere to the following core principles:

### Centralize all your data

**Simple, high volume data access and ingestion to big data repositories. Collect raw data from every source within the enterprise, regardless of complexity.**

The heart of the Modern Data Architecture is the Data Lake or Enterprise Data Hub, which enables organizations to access and retain more data from more sources in a highly scalable and cost-effective manner. Some organizations overlook legacy data sources, such as the mainframe, due to their complexity. These legacy systems power many mission-critical applications throughout the enterprise, and mainframes house 70% of all transactional enterprise data, including growing volumes of mobile and IoT data. Accessing and delivering usable data from all data sources, batch and streaming, to analytic environments – is essential for both operational analytics and business analytics.

### Turn your raw data into insight

**Integrate all data assets for advanced analytics and machine learning.**

Give data context and meaning so decision makers can leverage this data for bigger business insights. Decision-makers need all their organizations' data at their fingertips to be able to make the best decisions. But raw data won't tell them anything. More than ever, teams are challenged to manage, transform and utilize overwhelming volumes of data into meaningful actionable insight. Modern Data Integration enables your data scientists to cleanse, blend and transform data to uncover new information that can be added to your data pipeline. Data in an analytic format is transformed into insight that can be acted upon.

### Maintain governance and security standards

**Ensure data lineage, security and efficiency. Protect your data and comply with regulatory and internal mandates.**

As the big data repositories enable storing and processing significantly more data (structured and unstructured), in turn, more users and tools are accessing the data. The opportunity to drive greater insights is outstanding. However, more data, more users and more tools create even a bigger data governance challenge. Providing secure access to all enterprise data and metadata lineage across platforms become critical for the next generation data architecture.

### Simplify and Optimize IT Operations

**Automate and optimize your data pipeline, keep pace with the evolution of technology, and standardize platforms and infrastructures.**

Modern Data Architectures are designed to automate and optimize data pipelines, reduce risk and reduce the burden on staff. The architecture streamlines the development process and enables the adoption of newer technologies as they mature within the ecosystem, while reducing maintenance costs.

## Hadoop: The Centerpiece of the Modern Data Architecture

Apache™ Hadoop has become the central component of the modern data architecture and Apache™ Spark has the promise to be the single compute framework for multiple types of workloads, including deep learning, advanced analytics as well as batch and streaming data pipelines. These complex data processing and integration tasks are required for modern business intelligence and operational analytics.

As a result, Hadoop adoption has risen sharply as the answer to many organization's Big Data challenges. However, once inside the Hadoop ecosystem, new challenges arise regarding accessing and integrating complex data sources, specialized skillsets, long development cycles, and rapidly changing technology stack.

## Modern Data Management without the Complexity!

Syncsort DMX-h is specifically designed to help you achieve your modern data management objectives with a single interface for accessing and integrating all your enterprise data sources – batch and streaming – to turn raw data into insight.

Its game-changing architecture gives you the flexibility to design your jobs once and deploy them anywhere – across Hadoop, Spark, or single-server systems – on premise, or in the cloud. This architecture, plus native integration with Hadoop, enables DMX-h to quickly evolve with the ecosystem, so you can stay current on the latest technologies without rewriting your jobs or acquiring new skills.

Finally, you can feel confident processing your most critical data with DMX-h, as it packages the comprehensive support you need to manage, secure and govern the entire process.

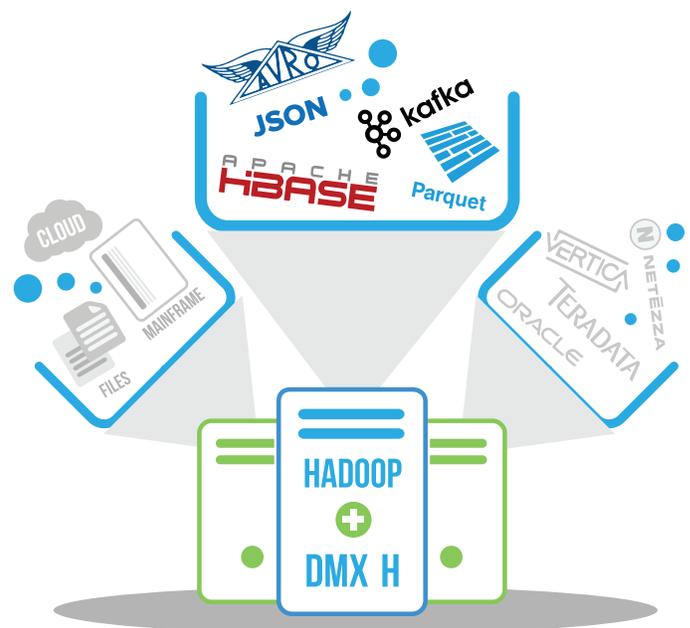
Intelligent Execution allows users to design sophisticated data transformations, focusing solely on business rules, not on the underlying platform or execution framework.

## ACCESS All Enterprise Data

Get best in class data ingestion capabilities for Hadoop – Mainframes, RDBMS, MPP, JSON, Avro/Parquet, NoSQL, and more

Don't miss out on vital insights by ignoring complex data sources such as the mainframe. Use DMX-h to connect securely to virtually any source using native drivers to extract data for guaranteed optimal speed and efficiency.

- Collect virtually any data from any source, including:
  - JSON
  - Kafka
  - Mainframe
  - MPP
  - NoSQL
  - RDBMS
  - S3
  - ... and more!
- Move hundreds of tables – including whole database schemas – into your data lake at once, with the press of a button
- Access both batch & streaming from the same interface
- Access, re-format and load data directly into Avro & Parquet. No staging required
- Load more data into Hadoop in less time. Let DMX-h dynamically split the data and load it to HDFS in parallel



## INTEGRATE and Achieve the Fastest Path from Raw Data to Insight

Design Streaming & Batch processes in a single interface



With DMX-h, you can quickly cleanse, blend and transform your data, giving it context and meaning so your organization can execute fast.

- Enrich your data on-the-fly at lightning speeds before loading into Hadoop
- High-performance connectivity to Big Data and NoSQL databases, such as Cassandra, HBase & MongoDB
- The fastest parallel loads to Amazon Redshift, Greenplum, Netezza, Oracle, Teradata and Vertica
- Create Tableau and Qlikview files with one click
- Integrate Enterprise-Wide Data with Real-Time Sources

Use DMX-h's graphical interface to subscribe, transform and enrich data from real-time Kafka queues. DMX-h can also publish these enriched datasets to Kafka, simplifying the creation of realtime analytical applications by cleansing, pre-processing and transforming data in motion.

## Goals of the Modern Data Architecture

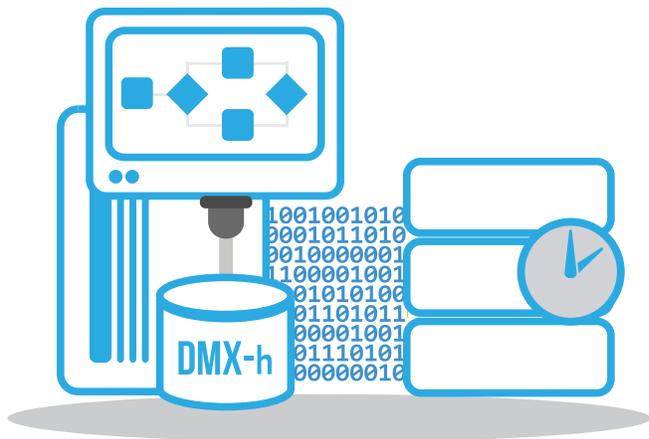
Syncsort combines cutting-edge technology and decades of experience with both mainframe and Big Data platforms to offer the best solution for accessing and integrating mainframe data with Hadoop. Syncsort provides the experience so you don't have to.

**Get mainframe data into Hadoop - in a mainframe format - and work with it like any other data source! Integrate mainframe data with emerging data sources!**

- Preserve your data exactly as it was on the mainframe to meet governance & compliance mandates
- Enable non-mainframe developers to work with native mainframe data on the cluster

**Integrate mainframe data with emerging data sources**

- Directly access and understand VSAM files, mainframe fixed & variable files, and DB2 data
- Give your data meaning with COBOL Copybooks mapped directly to the mainframe data
- Stop wasting weeks of development time just to understand the data



## COMPLY with Security, Management and Governance Protocols

Secure and govern with seamless Hadoop integration – Improve metadata management and track data lineage

Due to Syncsort's ongoing contributions to Apache™ Hadoop projects, DMX-h is natively integrated into the Hadoop data pipeline, providing interoperability and scalability. DMX-h offers:

**Management:** Full integration with Cloudera Manager and Apache™ Ambari for monitoring, maintenance and deployment across hundreds of nodes

### Security:

- Native LDAP and Kerberos support
- Apache™ Ranger and Apache™ Sentry certified
- Secure mainframe data access through FTPS and Connect:Direct

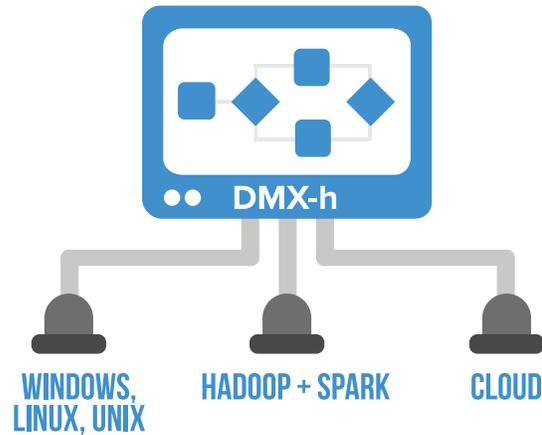


### Governance:

- Cloudera Navigator certified
- Tight integration with HCatalog for metadata management and data lineage
- Work directly with mainframe data in its native format – preserving data lineage across platforms

## SIMPLIFY – Design once.

Execute anywhere; on premise or in the Cloud  
Insulate your team from the underlying complexity, and everchanging skills requirements



Adaptability and simplification are critical in a successful modern data architecture. DMX-h is built with Intelligent Execution (IX) to transcend operating systems and execution platforms, simplifying your data management and future-proofing your applications. With DMX-h, you can:

- Visually design your jobs once, and deploy them anywhere – Hadoop, Spark, Linux, Unix, Windows – on premise or in the cloud. No changes or tuning required.
- Easily move applications from standalone server environments and from MapReduce to Spark – as easy as clicking on a dropdown menu
- Future-proof applications for emerging compute frameworks
- Avoid tuning -- Intelligent Execution dynamically plans for applications at run-time based on the chosen compute framework
- Insulate your users from the underlying complexities of Hadoop and leverage ETL skills
- Scale applications on-premise and in the Cloud
- Cut development time in half

## DMX-h Features to Support your Modern Data Architecture

These are our most popular features. If you are looking for a feature not listed, contact us at [info@syncsort.com](mailto:info@syncsort.com) to see if we support it!

### ACCESS

- Database: Amazon Redshift, DB2/UDB, Greenplum, Netezza, Oracle, SQL Server, Sybase ASE/IQ, Teradata, Vertica, IBM Websphere MQ, Salesforce.com, SAP Netweaver, SAP Hana
- Hadoop: Apache Avro, Apache Parquet, Apache Hive/HDFS, HCatalog
- Mainframe: Mainframe fixed, Mainframe variable (Hadoop distributable), Mainframe variable with block descriptor, Mainframe VSAM, DB2/z, IMS
- Kafka Topic source and target
- Data Visualization: QlikView data eXchange, Tableau TDE
- NoSQL databases (Cassandra, HBase, Mongo DB, etc.) via ODBC/JDBC drivers
- All other data stores via ODBC/JDBC if compliant
- C/Java data connector API available for additional data sources

### INTEGRATE

- Complex Filtering, Compression, Format Conversion, Data Cleansing (string manipulation, arithmetic calculation)
- Joins and lookups for data enrichment utilizing Intelligent Execution and dynamic optimizer for guaranteed performance throughout data and environment changes
- Set Level: High-performance Sort, Aggregate, Join, Merge
- Field Level: Comprehensive String, number and date functions
  - Conditional transformations (including regular expressions)
  - Built-in Hashing Algorithms (CRC32 and MD5)

### COMPLY

- Management: Cloudera Manager and Apache™ Ambari
- Security: Apache Ranger and Cloudera Sentry certified, Kerberos
- Metadata: HCatalog, Cloudera Navigator, file based open metadata

### SIMPLIFY

- One interface to design jobs to run on:
  - Single Node, Cluster
  - MapReduce, Spark, Storm, Future Platforms
  - Windows, Unix, Linux
  - On-Premises, Cloud
  - Batch, Streaming
- Intelligent Execution to insulate from rapidly changing big data technology stack
- No changes or tuning required, even if you change execution frameworks
- Future-proof job designs for emerging compute frameworks, e.g. Spark

### SUPPORT FOR ALL MAJOR DISTRIBUTIONS

- Cloudera, Hortonworks,
- MapR, Big Insights,
- Apache™ Hadoop and
- Apache™ Spark



Take DMX-h for a test drive at  
[www.syncsort.com/try](http://www.syncsort.com/try)