Get a Farm-to-Table View of Your Data

Tracking data quality and lineage on-premises and in the cloud, on and off the cluster

Dr. Tendü Yoğurtçu, Chief Technology Officer

syncsort

Who is Syncsort?

>7,000
Customers

syncsort

84
of Fortune 100 are Customers

The global leader in Big Iron to Big Data

500+
Experienced & Talented
Data Professionals

100+
Countries We Do Business In

3x
Revenue Growth
In Last 12 Months

Customer Use Cases & Strategic Partnerships

Data Infrastructure Optimization

- Mainframe Optimization
- Cross-Platform Capacity Management
- EDW Optimization
- Application Modernization

IBM · bmc · ca technologies · Compuware · splunk> · HORTONWORKS · NTT DATA · ORACLE

Data Availability

- High Availability & Disaster Recovery
- Mission-Critical Migration
- Cross-Platform Data Sharing
- IBM i Data Security & Audit

IBM · hms · Agilysys… · SIRIUS · AVNET · fiserv.

Data Integration

- Mainframe Access & Integration for Machine Data
- Mainframe Access & Integration for App Data
- High-performance ETL

splunk> · cloudera · HORTONWORKS

Data Quality

- Data Governance
- Customer 360
- Big Data Quality & Integration
- Data Enrichment & Validation

collibra · cloudera · asg · whitepages PRO · HORTONWORKS
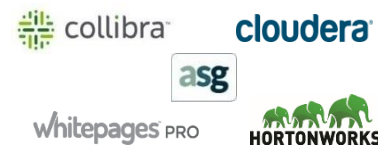
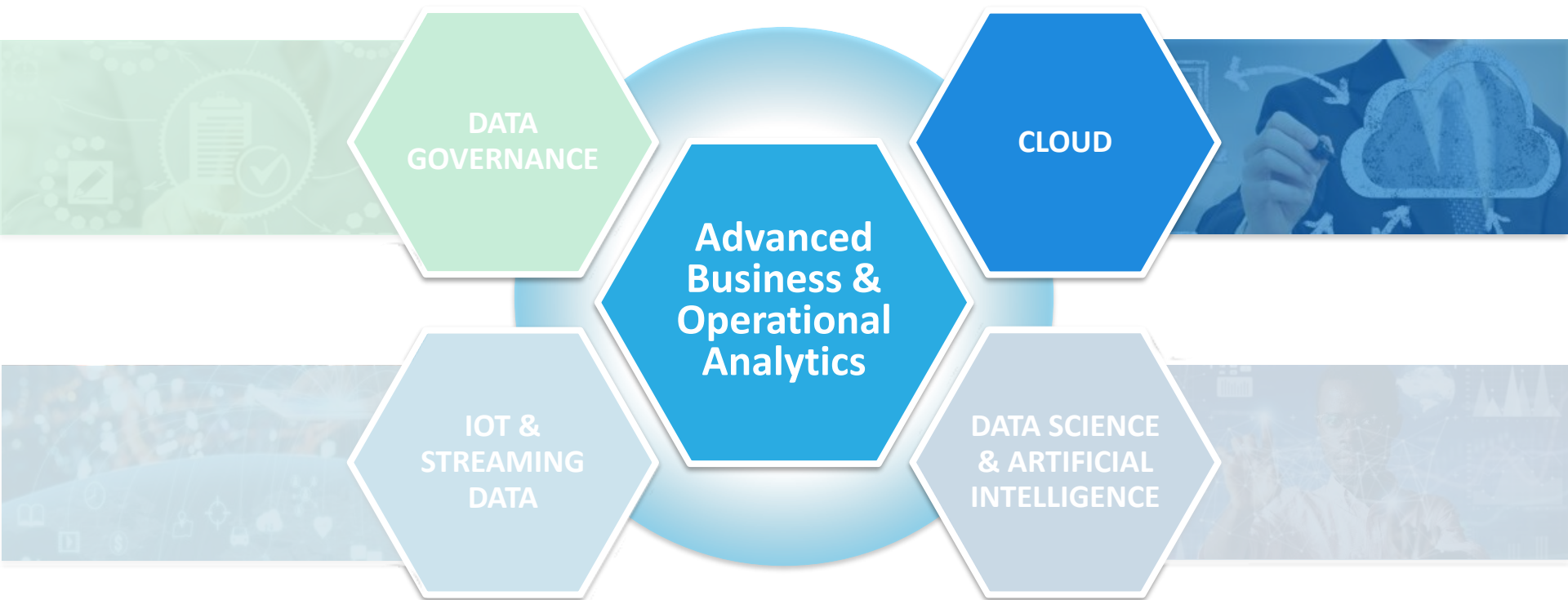## Big Iron to Big Data

A fast-growing market segment composed of solutions that optimize traditional data systems and deliver mission-critical data from these systems to next-generation analytic environments.

syncsort

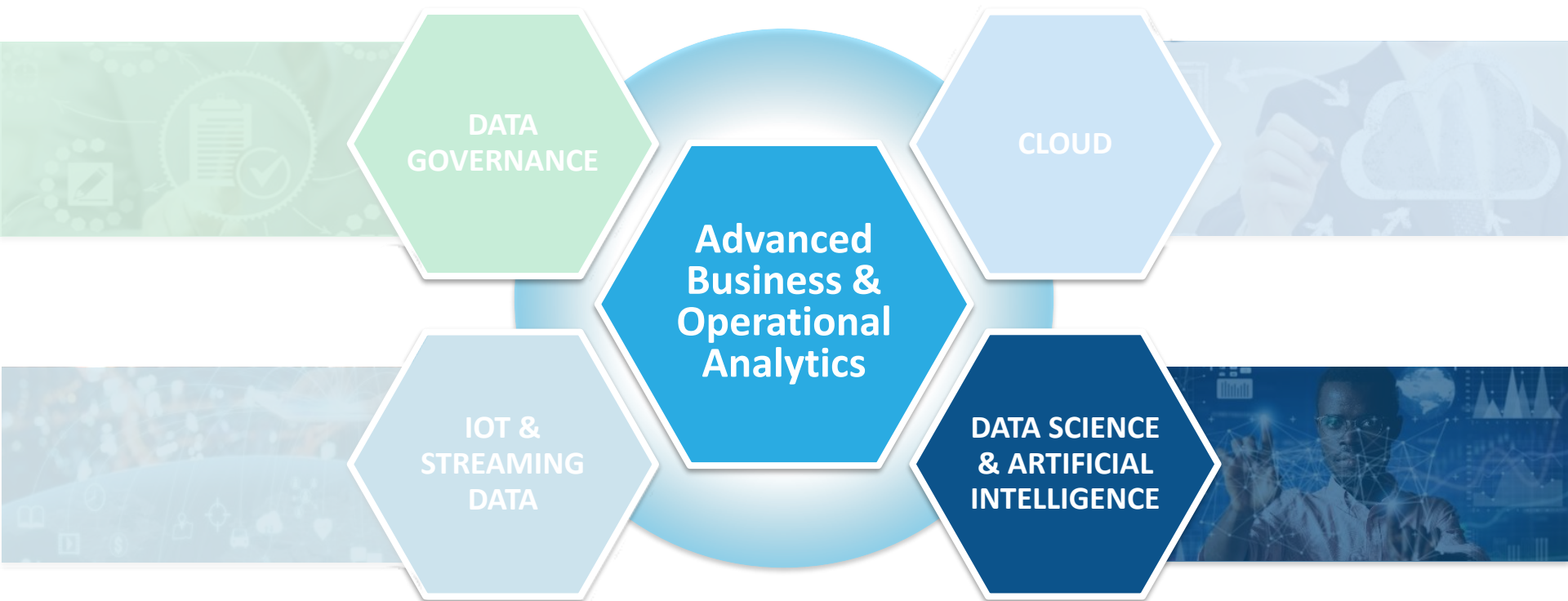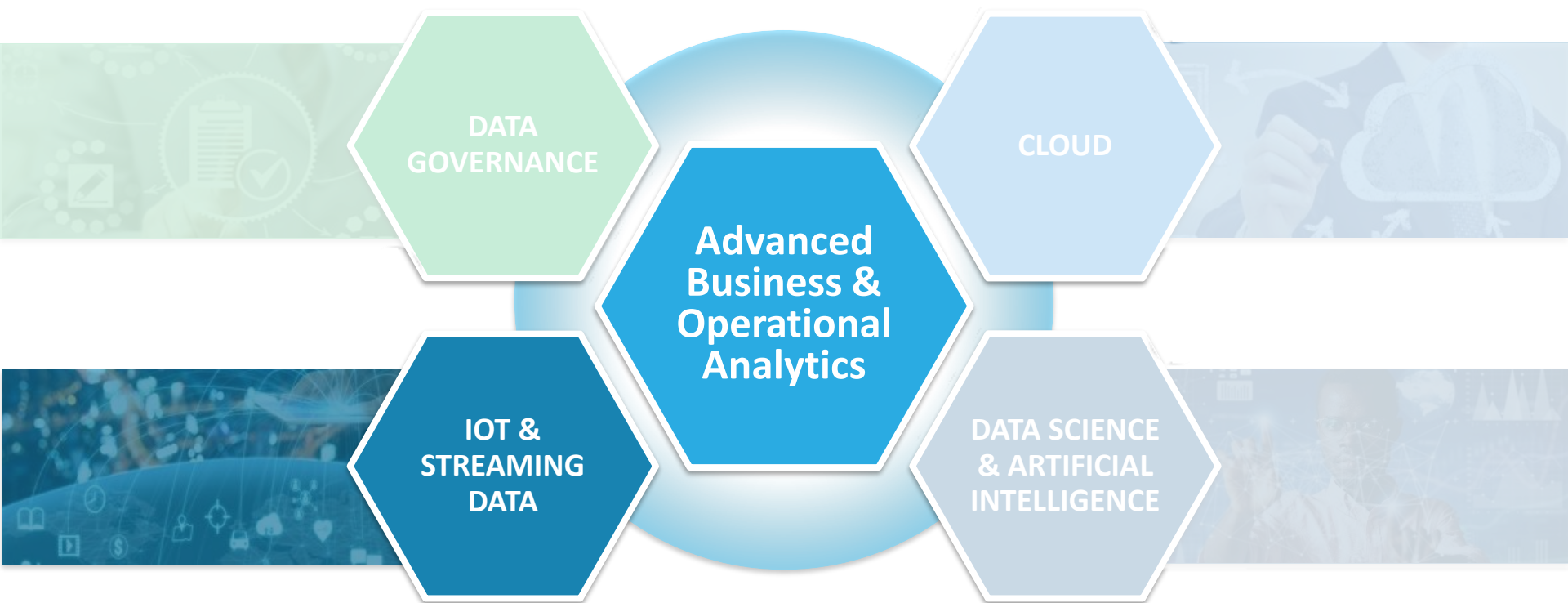# Farm to Table

Technology Trends Advancing Data

DATA GOVERNANCE

CLOUD

Advanced Business & Operational Analytics

IOT & STREAMING DATA

DATA SCIENCE & ARTIFICIAL INTELLIGENCE

Technology Trends Advancing Data

DATA GOVERNANCE

CLOUD

Advanced Business & Operational Analytics

IOT & STREAMING DATA

DATA SCIENCE & ARTIFICIAL INTELLIGENCE

Technology Trends Advancing Data

DATA GOVERNANCE

CLOUD

Advanced Business & Operational Analytics

IOT & STREAMING DATA

DATA SCIENCE & ARTIFICIAL INTELLIGENCE

Technology Trends Advancing Data

DATA GOVERNANCE

Advanced Business & Operational Analytics

CLOUD

IOT & STREAMING DATA

DATA SCIENCE & ARTIFICIAL INTELLIGENCE

Data Governance

Business imperative across platforms and deployment models, on-premise and in the cloud

GOALS

- Regulatory compliance
- Understand data context, meaning
- Accuracy, completeness, consistency, relevancy, timeliness, validity of data

CHALLENGES

- Multi-platform, data volume and complexity
- Diversity and consistency of sources
- Compliance demands: broader & deeper

Data Governance

▪ Requires a multi-faceted approach

QUALITY

- Discover sources of, relationships between, data
- Apply business rules to measure data quality continuously

SECURITY

- Protect the confidentiality, integrity and availability of data
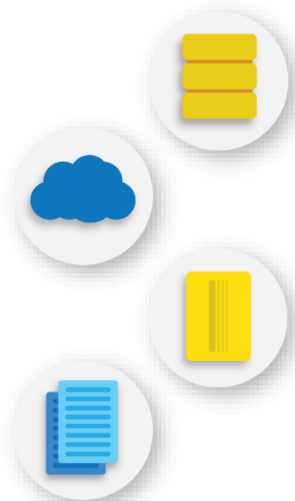
LINEAGE

- Get insights into where data came from, what changes were made and where it lands

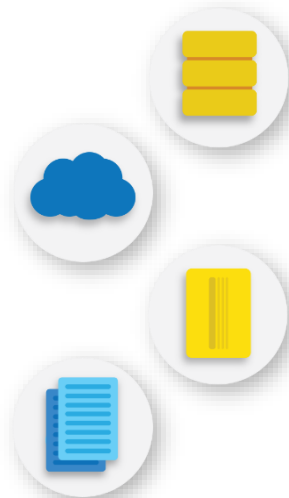# End to End Data Lineage in Cloudera Navigator

Data Sources

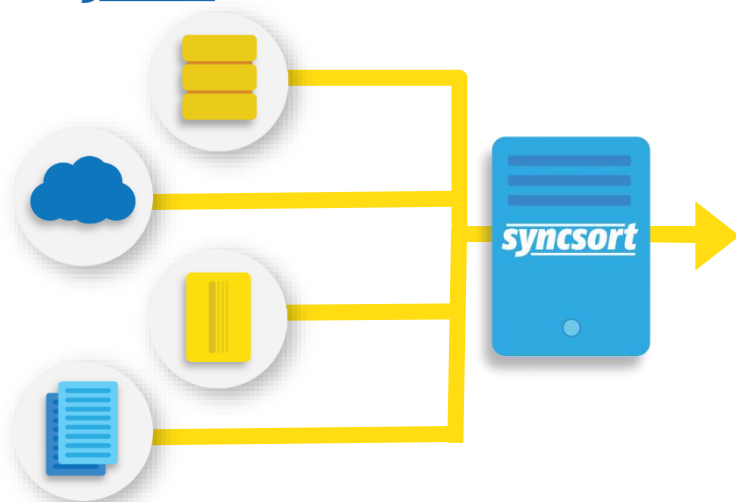# End to End Data Lineage in Cloudera Navigator

Data Sources

syncsort

Syncsort accesses data from sources outside cluster.

End to End Data Lineage in Cloudera Navigator

Data Sources

syncsort

Syncsort accesses data from sources outside cluster.

Syncsort onboards data, modifies on-the-fly to match Hadoop storage model.

End to End Data Lineage in Cloudera Navigator

Data Sources

Data Hub

Syncsort accesses data from sources outside cluster.

Syncsort onboards data, modifies on-the-fly to match Hadoop storage model.

Syncsort changes, enhances, joins data in cluster with MapReduce or Spark.

End to End Data Lineage in Cloudera Navigator

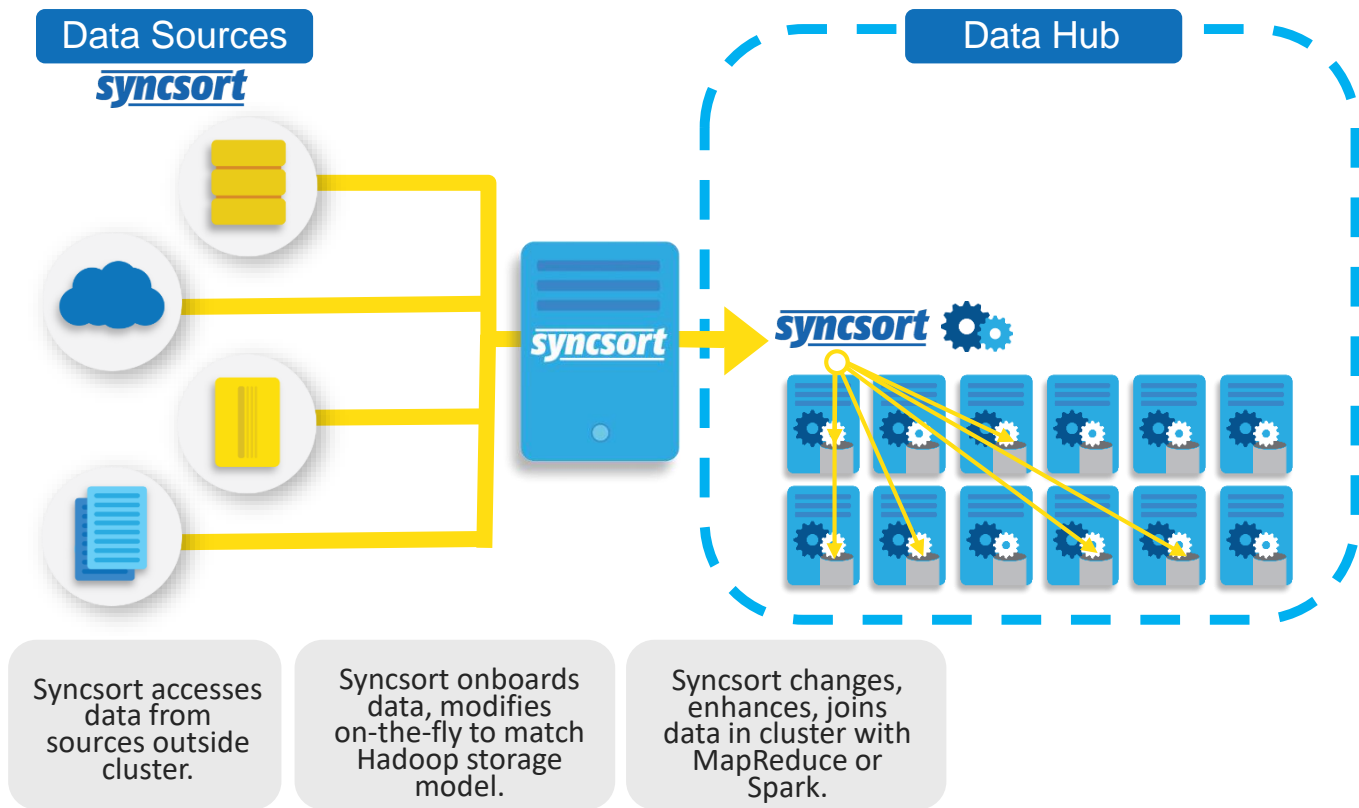Data Sources

syncsort

Data Hub

cloudera navigator

syncsort

| Syncsort accesses data from sources outside cluster. | Syncsort onboards data, modifies on-the-fly to match Hadoop storage model. | Syncsort changes, enhances, joins data in cluster with MapReduce or Spark. | Syncsort passes source-to-cluster data lineage info to Navigator. |

End to End Data Lineage in Cloudera Navigator

Data Sources

Data Hub

cloudera navigator

Data changes made by MapReduce, Spark, HiveQL.
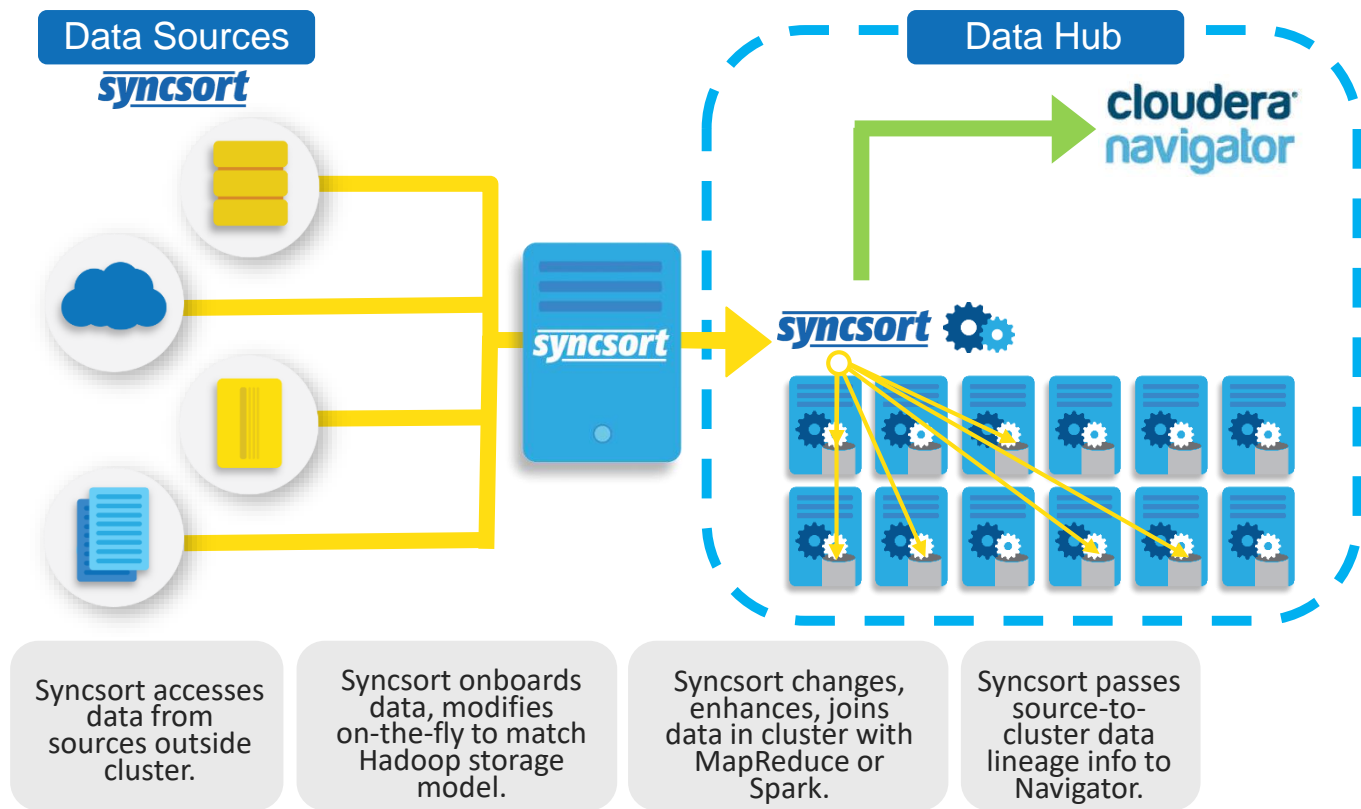
Syncsort accesses data from sources outside cluster.

Syncsort onboards data, modifies on-the-fly to match Hadoop storage model.

Syncsort changes, enhances, joins data in cluster with MapReduce or Spark.

Syncsort passes source-to-cluster data lineage info to Navigator.

Navigator gathers any other changes made to data on cluster.

End to End Data Lineage in Cloudera Navigator

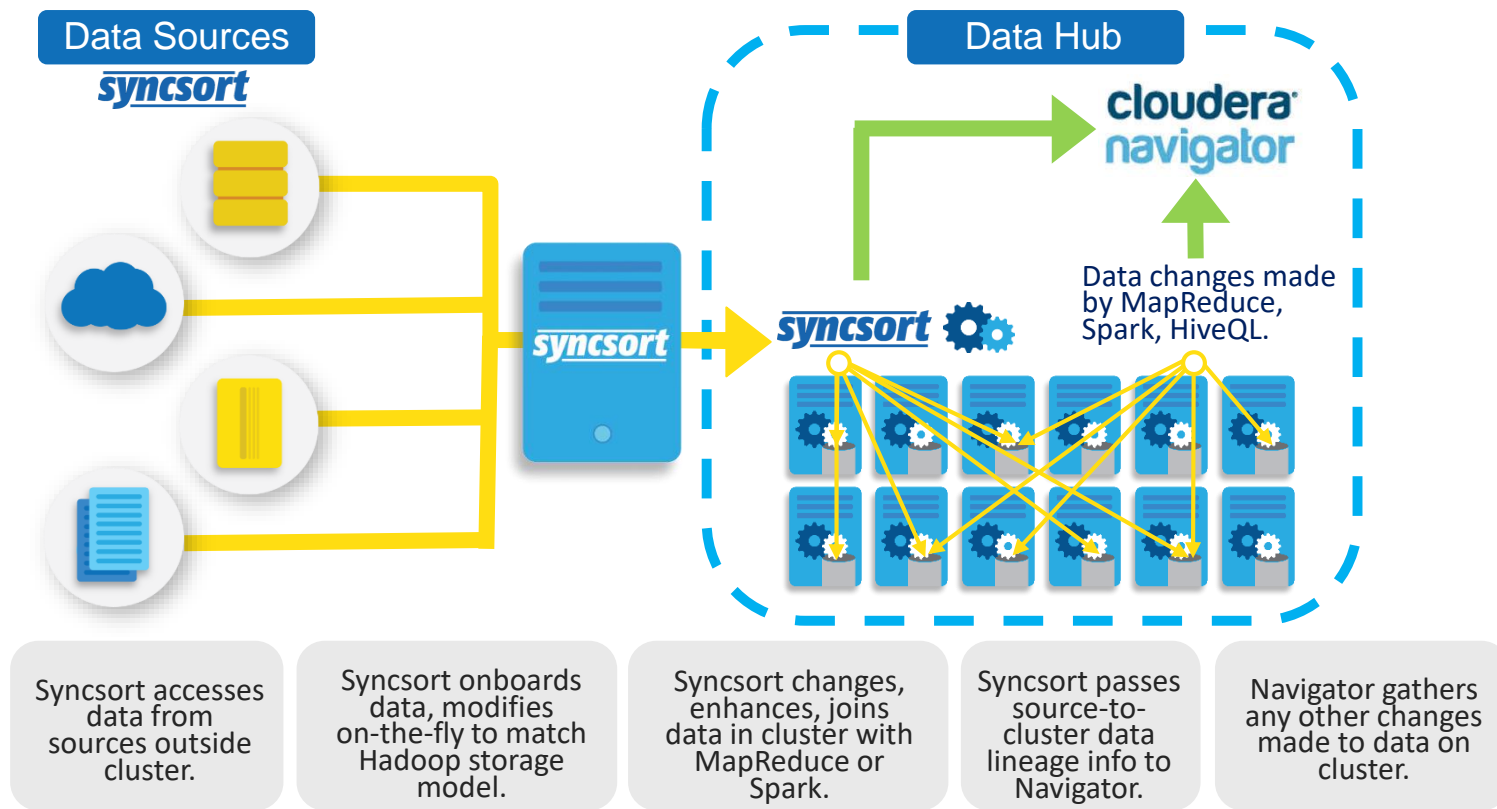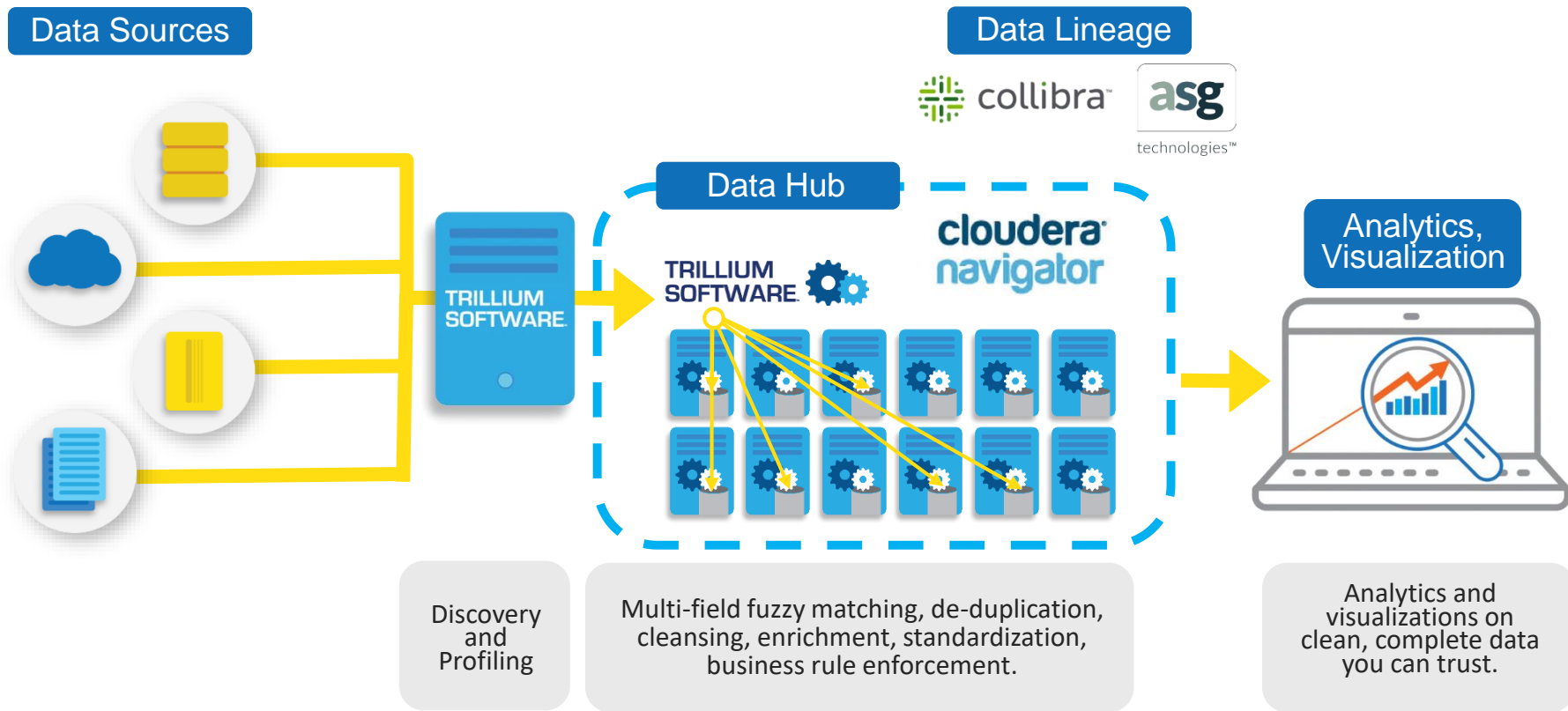| Syncsort accesses data from sources outside cluster. | Syncsort onboards data, modifies on-the-fly to match Hadoop storage model. | Syncsort changes, enhances, joins data in cluster with MapReduce or Spark. | Syncsort passes source-to-cluster data lineage info to Navigator. | Navigator gathers any other changes made to data on cluster. | Analytics and visualizations get complete data. |

Syncsort DMX-h + Cloudera Navigator for End-to-End Lineage
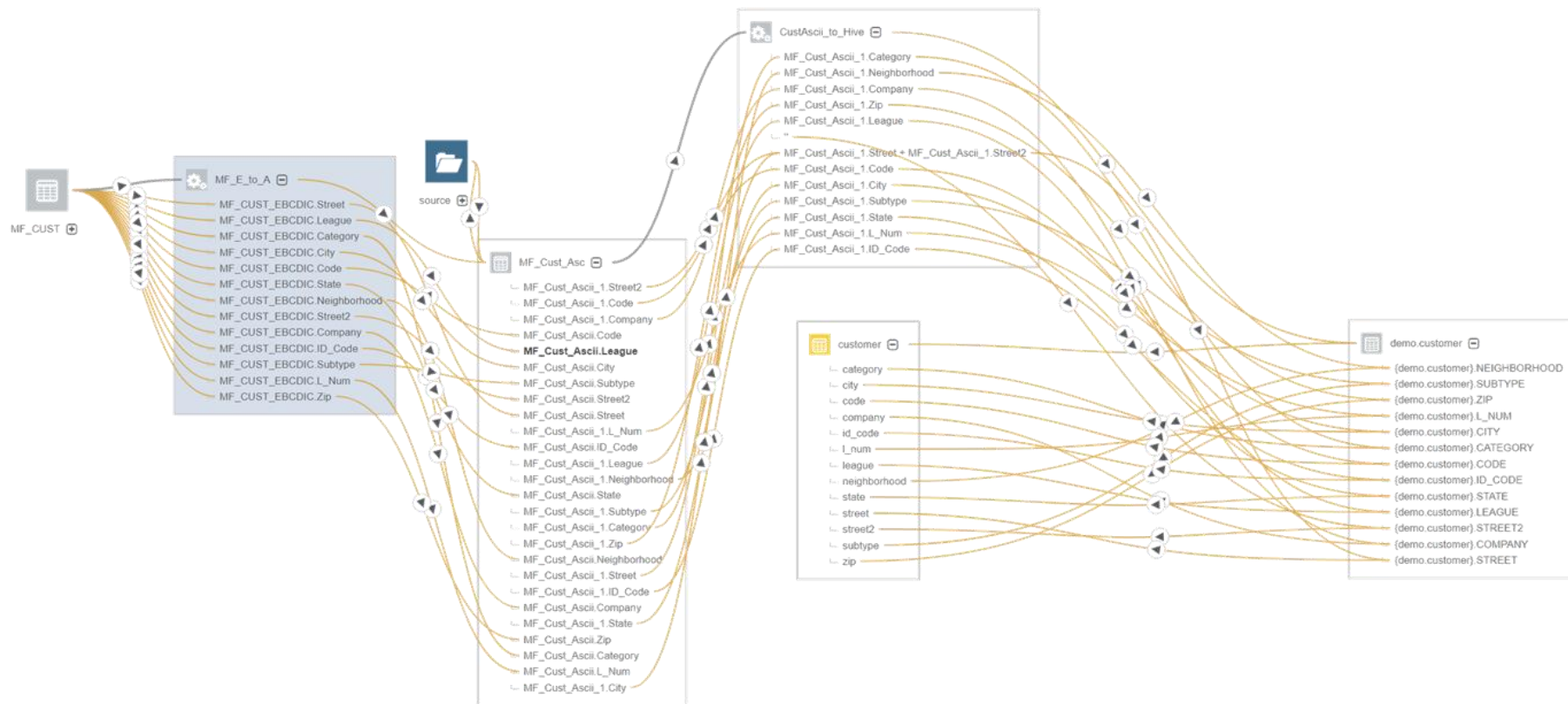
Data Lineage + Data Quality = Foundations of Data Governance

Data Sources

Data Lineage

collibra

asg technologies™

Data Hub

TRILLIUM SOFTWARE

cloudera navigator

TRILLIUM SOFTWARE

Analytics, Visualization

Discovery and Profiling

Multi-field fuzzy matching, de-duplication, cleansing, enrichment, standardization, business rule enforcement.

Analytics and visualizations on clean, complete data you can trust.

Anti-Money Laundering Solution on CDH at Large Global Bank

▸ **Challenge:** Meet AML transaction monitoring and FCA compliance demands

– Data too large, diversely scattered to analyze

– Disparate data sources -- Mainframe, RDBMS, Cloud, etc

▸ **Requirements:**

– Consolidated, clean, verified data for all analytics and reporting.

– MUST have complete, detailed data lineage from origin to end point

– MUST be secure: Kerberos and LDAP integration required

– Need unmodified copy of mainframe data stored on Hadoop for backup, archive

Anti-Money Laundering Solution on CDH at Large Global Bank

▸ **Solution:**

- Syncsort DMX-h to create "Golden Record" on CDH for compliance archiving

- Trillium Quality for Big Data for cluster-native data verification, enrichment, and demanding multi-field entity resolution on Spark framework

- Full end-to-end lineage to Cloudera Navigator, from all sources, through transformations, to data landing, including HiveQL changes

▸ **Benefits:**

- New financial crimes data hub produces high performance results at massive scale

- Bank meets stringent Anti-Money Laundering compliance requirements