

---

# DATA STRATEGY DRIVES CONNECTED MANUFACTURING SUCCESS

Real Time, Scalable Data Lifecycle Solutions Enable Industry 4.0

---

Connected Manufacturing is at a turning point and it is catalyzed by a real, measurable change and shift in data types. Real-time and time series data have come into the forefront of manufacturing use cases because they enable real-time insight and more importantly autonomous decision making. Cloudera Data Platform lowers the cost of data management, handles the volume and variety of IoT data and manages the complexity of real-time data with unique capabilities that address hybrid and multi-cloud environments, solves the most demand manufacturing use cases, provides edge to AI security and governance and is founded on an open continuously innovating platform architecture.

## Connected Manufacturing's Pivot to an Enterprise Data Solution

Connected Manufacturing is at a turning point and it is catalyzed by a real, measurable change and shift in data types—real-time and time series data is growing 50% faster than latent or static data forms and streaming analytics is projected to grow at a 34% CAGR, leaving legacy data platforms that specialize in static historical data solutions, functioning on-prem or in discrete clouds, inadequate in addressing today's real-time insight needs. Streaming data's growth is powered from the fact that it enables real-time insight and more importantly, autonomous decision making.

This shift has been enabled by the proliferation of inexpensive process sensors tailored to specific use, robust edge computing devices allowing repetitive autonomous decisions, cloud computing performing both analytics and storage, and soon to come - 5G, which opens the lanes of the data superhighway, freeing manufacturing processes from the chains of hard wired connections. But the benefits of streaming data create a challenge in managing its massive data volume, its diverse data structure, and its real-time speed of light velocity across enterprise business processes.

## Limitations to Traditional Connected Manufacturing Data Solutions

Today's manufacturers face limitations addressing the stressful complexities of digitalization, much of these limitations are due to the rapid evolution of new & connected data sources and the massive volume of data spewn out. Some of the critical challenges and key considerations that organizations face with respect to data management for Connected Manufacturing are:

- **The cost of data management**—Traditional data management mechanisms tend to be notoriously expensive, do not scale easily, and were not built for capturing and processing the petabytes of IoT data streaming from connected devices. Today, organizations need a more flexible and scalable data management & analytics platform that can easily ingest, store, manage, and process streaming data from IoT sources at a lower cost.
- **Handling the volume and variety of IoT data**—To enable process monitoring & optimization, predictive maintenance and emerging IoT use cases, information architects need a platform that can handle all types of diverse data structures and schemas, including everything from intermittent readings of temperature, pressure, and vibrations per second to handling fully unstructured data (e.g., images, video, text, spectral data) or other forms such as thermographic or acoustic signals, from the edge delivered through diverse supported drivers and protocols.
- **Managing the complexity of real-time data:** In order to drive continuous process monitoring, throughput optimization or predictive maintenance, a data management platform needs to enable real-time analytics on streaming data. The platform also needs to effectively ingest, store, and process the streaming data in real time or near-real time in order to instantly deliver insights and action.

### Hot Topics to Consider with Industry 4.0

#### EDGE

48% stated Manufacturing's number one priority is to install smart sensors linked to edge device computing. Edge computing is best suited for applications requiring anatomy, high bandwidth and low latency.

#### 5G

5G fundamental enabling technology is its ability to manage bandwidth, the number of devices connected and latency, for routine autonomous decisions at the edge or in the cloud for complex enterprise wide aggregation and analysis for machine learning applications.

#### MACHINE LEARNING

To effectively leverage machine learning, vast data volumes of diverse data are demanded. The quandary though is amassing vast quantities of data needed to build and test credible machine learning tools, when it resides in multiple locations, structured in a multitude of forms and doesn't have a governance protocol needed to safely move data.

- **Freeing data from independent silos**—Specialized processes (innovation platforms, QMS, MES, etc) within the value chain reward disparate data sources and data management platforms that tailor to unique siloed solutions. These narrow point solutions limit enterprise value by considering only a fraction of the insight cross-enterprise data can offer, in addition, duplicate siloed solutions divide the business, limiting collaboration opportunities. Also, the platform must have the ability to ingest, store, manage, and process streaming data from all points in the value chain, combine it with Data Historians, ERP, MES and QMS sources and leverage it into actionable insights.
- **Diverse analytical capabilities**—Existing platforms offer limited ability to provide insights and analytics into platform usage and performance. For Connected Manufacturing IoT solutions, the platform must provide a wide range of analytical options—including everything from SQL analytics and search capabilities, to tools to support machine learning and modelling, along with tight integration with leading business intelligence (BI) solutions that offer specialized dashboard and business analytics capabilities.
- **Predictive modeling capabilities**—Predictive modelling capabilities are key to delivering insights, and current platforms provide little to no modeling or machine learning capabilities used to predict and prevent disturbances before they impact operations. When building these models, leveraging disparate data sources or data types promotes a robust and well rounded model resulting in stronger prediction capabilities.

Given the complexity and variety of manufacturing and IoT data, manufacturers are reinventing their data management and analytics strategy—transitioning to a multi-function open platform that is optimized for the massive scale and complexity of manufacturing data. They also need an integrated suite of proven and open data management tools and analytics engines driving insights and analytics—all with the robust security, governance, data protection, and management capabilities that organizations demand.

### A Successful Industry 4.0 Enterprise Data Solution

Analytics, machine learning, and artificial intelligence are reshaping the manufacturing landscape. The most innovative manufacturing companies are applying industrial IoT concepts, coupled with data and analytics, to transform their product development, supply chain, and manufacturing operations.

However in order to capitalize on this fundamental shift, manufacturers need the ability to ingest, process, store, and analyze all types of data (structured, unstructured, and semi-structured data) regardless of where it lands—at the edge, on the factory floor, from Data Historians, in any public cloud, or in a hybrid cloud.

Today's manufacturers need to be able to generate insights and use cases not only from data at rest, but also from data in motion and streaming data sources in real time. Manufacturers are utilizing machine learning, advanced analytics, and artificial intelligence to identify patterns from petabytes of data, detect anomalies, and predict potential outcomes for their businesses.

With greater visibility and insights at the point of decision, manufacturers can use the data they generate every day to gain real-time visibility into their operations and supply chain, improve operational performance, enable predictive maintenance, and help drive new revenue streams.

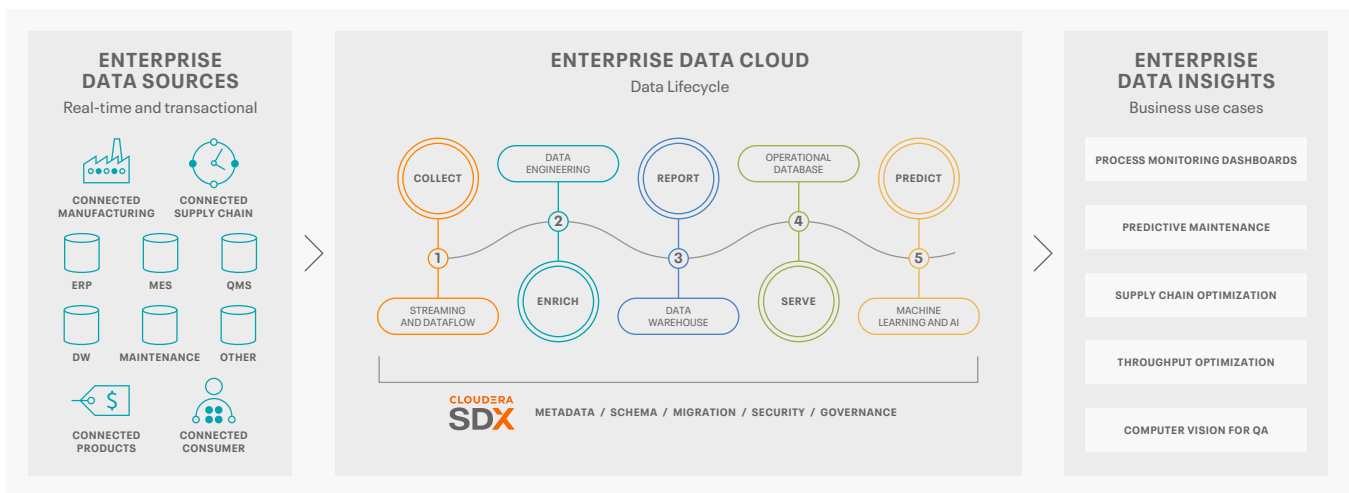
## End-to-End Manufacturing Data Flow with Cloudera

Today, leading manufacturing organizations worldwide are adopting an enterprise data cloud strategy using the [Cloudera Data Platform](#) managing the end-to-end data journey from absorbing raw data at the source, creating autonomous decision making, to driving actionable insights and value producing use cases.

With Cloudera, manufacturers can ingest data from a variety of sources including streaming and enterprise data sources, store and process it across hybrid infrastructures, run analytics or apply machine learning algorithms to all data, all while maintaining strict enterprise data security, governance, and control across all environments.

### Cloudera Data Platform (CDP)

Cloudera Data Platform (CDP), the enterprise data platform, transcends silos in real time by supporting shared analytics and collaboration across teams also enabling data unification and security across the enterprise from the time a product is conceptualized, through scale-up, manufacture to engagement of the customer in post-sales service and maintenance relationships.



Attributes unique to CDP include:

- **Hybrid and multi-cloud**—provides choices to manage, analyze and experiment with data either on-prem and/or in any public or private cloud environments for maximum choice and flexibility.
- **Multi-function**—solves the most demanding business use cases—applying real-time stream processing, data warehousing, data science and iterative machine learning across shared enterprise and real time data at scale.
- **Secure and governed**—simplifies data privacy and compliance for diverse enterprise data with a common security model and governance (powered by Cloudera SDX) to control data in an on-premise, public or private cloud - or hybrid environments.
- **Open**—facilitates the continuous innovation of the open source community, the choice of open storage and compute architectures, and the confidence and flexibility of a broad eco-system.

“From day one, we wanted to build a system that is able to handle the increasing amount of data... What makes AVIATAR unique, is it’s the only independent platform that enables our customers to have full control of their own data. Privacy, transparency and security allows our customers a piece of mind to fully analyze and gain insights from their data and use it effectively.”

Jan Stövesand, Senior Director, Data and Analytics, Lufthansa Technik

### Lufthansa Technik

#### CHALLENGE

Airline MRO operations are costly and time consuming as dispatch reliability is more important than ever due to tighter flight schedules and higher penalties for delays. Lufthansa Technik, sought to leverage the vast data generated in maintenance, operations and customer experience optimize its operations

#### SOLUTION

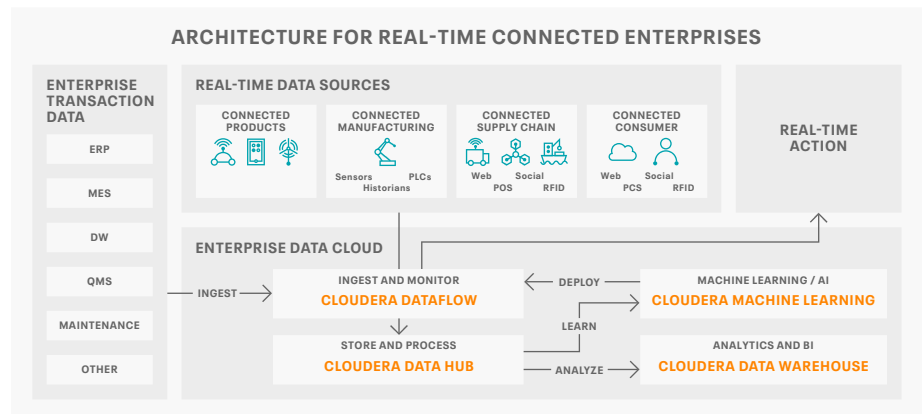
Lufthansa Technik built AVIATAR it’s platform as a service strategic answer to these challenges. AVIATAR optimizes the entire operation, from predictive maintenance to automated fulfilment solutions and combines profound airline operation expertise, data science, and engineering knowledge by aggregating data into one source of truth

#### IMPACT

A fully integrated data platform has enabled airlines serviced by Lufthansa Technik to lead to 40% less predicted components removal.

## Data-driven Manufacturing with Cloudera

The diagram below provides a snapshot of the successful enterprise data platform used for Connected Manufacturing’s predictive analytics use case, built on Cloudera.



Cloudera Data Platform leverages the full of capabilities to ingest, store, analyze and deploy streaming and historical data with the following components:

### Cloudera DataFlow (CDF)

[Cloudera DataFlow](#) provides scalable, real-time streaming analytics that ingests, curates, and analyzes data for key insights and immediate actionable intelligence. It can ingest and process real-time data from streaming data sources (such as connected machines and connected products) and also from traditional enterprise data sources such as ERP, MES, and QMS systems. It addresses the key challenges enterprises face with data in motion including the ability to:

- Ingest and process real-time data streaming at high volume and high scale
- Drive stream processing and analytics on data-in-motion
- Track data provenance and lineage of streaming data
- Manage and monitor edge applications and streaming sources

### Cloudera Data Hub

[Cloudera Data Hub](#) provides massively distributed storage and processing engines for large data sets, storing and processing of any kind of data including unstructured data, semi-structured data (i.e., sensor data) and structured data (i.e., transactional data from ERP, maintenance warehouses, SCM, CRM data, etc.) and provides the ability to execute a wide range of data processing workloads in an extremely high-performance manner.

### Cloudera Data Warehouse

[Cloudera Data Warehouse](#) is an enterprise-grade, hybrid cloud solution designed for self-service analytics enabling organizations to share petabytes of data to drive analytics and BI with the security, governance, and availability that large enterprises demand.

### Cloudera Machine Learning

[Cloudera Machine Learning](#) helps accelerate data science at scale to build, test, iterate, and deploy machine learning models in production by taking advantage of massively parallel computing and expanded data streams, delivering a self-service experience to data scientists developing and prototyping new machine learning projects and easily deploying them to production.

- Accelerates data science at scale to build, test, iterate, and deploy machine learning models in production
- Experiment faster, with on-demand compute and secure data access
- Enables data scientists to push these models out to the edge to continuously monitor digital signatures from connected data sources and drive action in real-time

## The Possibilities are Endless

The power of the enterprise can come to bear when data is integrated, flows across siloed departmental boundaries, and extends from the first days of product conception, through the edge of process monitoring to finally, the customer's in situ product use conditions in a real-time streaming manner. When data truly flows through an organization the possibilities are boundless:




- **Product Development**—with an integrated enterprise data philosophy, voice of customer data is flowed back to Product Development so that the next generation of products are embedded with valued performance, not value-killing fluff.
- **Operations**—production scale-up is shortened or eliminated as the product arrives fully baked from Product Development. Real time product availability is communicated to the Supply Chain, lowering inventory costs, enabling just-in-time delivery, supply chain logistics optimization and driving global coordination. Autonomous decisions are implemented at the edge in real-time maximizing uptime and productivity.
- **Supply Chain**—extends the value provided by leveraging insight in product quality through in-process rejections and real-time field performance - building the strength of the supply chain.
- **Marketing, Sales and Service**—connected and fully embedded into the organization feeding Product Development, Operations and Supply Chain insights, reaping the benefits of an efficient value chain promoting customer satisfaction, transitioning the business to Service-as-a-Platform.

**About Cloudera**

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at [cloudera.com](https://cloudera.com)

Successful manufacturing use cases embodying this vision:

CUSTOMER	BUSINESS PROCESS CHALLENGES	SOLUTIONS AND BUSINESS IMPACT
<p>Tenaris—A manufacturer of steel tubing for the energy sector</p> 	<p><b>Process Optimization</b>—Legacy systems lacked the flexibility to combine data from various sources limiting the ability to create forecasts and predictive models</p>	<p>Manufacturing and energy consumption efficiencies saved costs and higher product quality drove market share gains</p>
<p>Komatsu Mining—A heavy equipment manufacturer to the mining industry</p> 	<p><b>Predictive Maintenance as a Service</b>—Existing systems could not ingest, process and analyze real-time data and analytics obtained from its smart, connected devices and assets</p>	<p>Real time insight into equipment performance doubled utilization of the Joy Global longwall mining system for one coal mining customer</p>
<p>Micron—A Semiconductor Fab</p> 	<p><b>Process Optimization</b>—Legacy systems could not aggregate feeds from over 50 data sources across 14 manufacturing steps at eight fabrication site into one unified enterprise vision</p>	<p>Shortened the time it took to identify misprocessed dies from 7 days to under an hour, thus limiting defects from flowing downstream into more costly and labor intensive fab operations</p>

**Conclusion**

Effective Connected Manufacturing implementation doesn't start and end with point solutions solving tactical challenges within boundaries of business functions and processes resulting in limited enterprise value. Effective Connected Manufacturing adoption lies in the foundational realization that data is the heart of the true potential of Connected Manufacturing, one that allows data to flow across the organization in real time as the unifying ribbon tying all parts of the business together. Cloudera delivers the promise of Connected Manufacturing with its data lifecycle platform enabling data ingestion, storage and analysis from edge to AI in real time.