cloudera®

SOLUTION BRIEF

Manufacturing Potential

- Reduced overall equipment downtime
- Reduced factory inventory costs
- Improved customer support
- More accurate yield predictions
- Global value chain insight
- Higher production quality

Characteristics of a Manufacturing Data Platform

An effective data management platform for manufacturing should:

- Scale easily and efficiently based on data growth
- Drive cost efficiency; low cost per terabyte
- Support multiple data types and structures
- Effectively handle both data at rest and data in motion
- Easily ingest and process data in real time
- Be fundamentally secure
- Support hybrid cloud environments

Data-Driven Manufacturing

Data in Manufacturing

Manufacturing as an industry has always been at the forefront of squeezing value from data. Instrumentation, highly connected systems, and automation have been part and parcel of manufacturing organisations for decades. Constrained by the state of technology more than cost, process optimisation was always achieved by making clever use of the data available and has given rise to completely new disciplines and applications.

Yet many manufacturers now feel they've bumped against a ceiling. Data volumes from both inside as well as outside the manufacturing process continue to grow. The makeup of data changes too, and is more and more unstructured. There is a proliferation of machines and sensors, industrial IoT, each spewing forth torrents of data, disrupting manufacturing industry with near certainty[1]. Getting a handle on these flows of information becomes cost prohibitive with the current approaches. In order to innovate, to remain competitive and differentiate, manufacturing organisations must break through and fundamentally change their approach to managing their data in order to gain the insight they need: an actionable view of their operations, products, customers, and supply chain.

Areas of Opportunity

What if business were no longer constrained in terms of the data at its disposal? What if the boundaries for innovation and insight were no longer set by the sliver of historical data on offer or the processing power available? There would then be no limit to the questions to be asked, simulations run, or alternatives evaluated. Manufacturing can move from working with approximations to insight based on actual data, as it happens. Overall performance is improved through better forecasts of product demand and production, through understanding of plant operations across multiple metrics, and by being able to provide service and support to customers faster. Analysts predict benefits that range from reducing overall equipment downtime up to 50 percent through IoT-enabled predictive maintenance[2] and savings between 20 percent and 50 percent on factory inventory costs through real-time inventory monitoring and optimisation measures[3]. In fact, the potential value that could be unlocked with IoT applications in a manufacturing setting could be as much as $3.7 trillion[4] by 2025!

Given the complexity and variety of manufacturing and IoT data, organisations must fundamentally rethink their data management strategy—transitioning to a platform that is optimised for the scale and complexity the data problem presents. A platform that allows organisations to take advantage of the benefits the cloud offers, yet still lets them leverage their existing infrastructure. A platform that can either complement or replace existing databases and data warehouses, that can handle data offloads and warm and cold data. More importantly, a platform where the real value of the data can be exploited by being able to combine and correlate data streams with data from other internal and external data sources. The applications are wide and varied yet there are three areas of impact and opportunity that are common to all manufacturing organisations:

- **IoT and Connected Manufacturing:** With IoT, manufacturers can gain a comprehensive view of what is going on at every point in the production process and can make real-time adjustments to maintain an uninterrupted flow of finished goods and avoid defects. A data management platform lets companies ingest, store, process, analyse, and drive insights into all of the data from manufacturing processes and IoT sensors.

[1] Ovum, Understanding the IoT Opportunity: an Industry Perspective, 2015
[2] McKinsey Global Institute, The Internet of Things: Mapping the Value Beyond the Hype, June 2015
[3] McKinsey Digital, Industry 4.0: How to Navigate Digitization of the Manufacturing Sector, 2015
[4] McKinsey Global Institute, Unlocking the Potential of the Internet of Things, June 2015

Cloudera for Manufacturing

- Effectively handle both data at rest and data in motion
- Easily ingest millions of events per second
- Industry leadership in Spark
- Real-time processing and analytics
- Hybrid cloud deployments
- Effectively combine sensor and system data with other internal and external sources
- Data security beyond compromise
- Proven success across diverse manufacturing use cases

Key Manufacturing Use Cases Supported

- Connected manufacturing
- Predictive maintenance
- Industrial IoT
- Proactive QA
- Supply chain optimisation
- Inventory optimisation
- Warehouse optimisation
- Supplier traceability
- Production optimisation

- **Supply Chain and Inventory Optimization:** The enterprise data hub, through a sophisticated system of forecasting and logistics models, can be leveraged as a real-time supply network, balancing lead times with demand signals, ERP and MES data, retail stock, inventory capacities, and supplier throughput for totally seamless operations.
- **Proactive QA:** Quality compliance costs 100 times more to correct a problem during maintenance than in pre production. Through a single data platform, development timelines can be reduced up to 50 percent while responding to performance concerns in real time and eliminating defects prior to manufacture.

Cloudera Enterprise — The Data Management Platform for Manufacturing

Given the characteristics of manufacturing data streams, leading organisations around the globe are adopting Cloudera Enterprise—based on Apache Hadoop—as the data management and analytics platform for storing, managing, processing, and, more importantly, driving analytics from all of their manufacturing data.

With Cloudera Enterprise, organisations can easily bring information from multiple sources onto a single, unified platform at considerably lower cost per terabyte. This includes sensor readings, line productivity, transaction data, supplier data, and more. And because Hadoop is built on a highly scalable and flexible file system, any type of data (both structured and unstructured) can be loaded into Cloudera Enterprise without altering its format—preserving data integrity and delivering complete analytic flexibility. Data generated by machines and sensors, including time-series data as well as application and web log files, can be collected in real time and streamed directly into Cloudera Enterprise—instead of being staged in a temporary file system or data marts. And because Hadoop runs on industry-standard hardware, appliances, or on the cloud, the cost per terabyte of storage and processing is, on average, 10 times cheaper than a traditional relational data warehouse system.

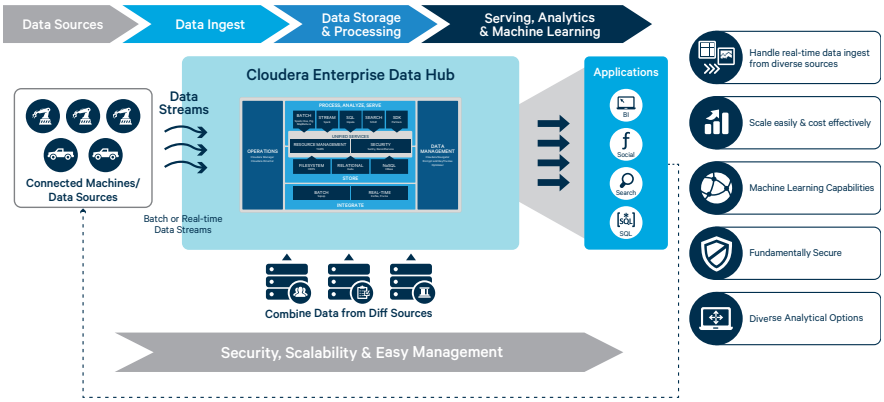

Figure 1: Manufacturing Data Management Value Chain Using Cloudera Enterprise

Using Cloudera to Drive Business Value for Manufacturing

Some key attributes of Hadoop and Cloudera Enterprise that make it perfect for manufacturing and IoT data management and analytics include:

- **Flexible Data Ingest:** Easily ingests data from multiple data sources, and supports both batch as well as real-time data ingest from sensors using tools such as Apache Kafka and Apache Flume.
- **Reliable, Scalable, Always-on Data Ingest:** Supports continuous streaming ingest, data drift (schema and semantic changes), and IoT data pipeline visualisation with our partner StreamSets. Using StreamSets Data Collector, organisations can easily develop and operate data flows and manage complex ingest pipelines from a variety of IoT data sources.
- **Data Variety:** Effectively handles multiple data types, structures, and schemas—from intermittent sensor readings of temperature and pressure to real-time location data or streaming live video feeds.
- **Real-Time Serving and Insights (Data in Motion):** Supports real-time processing and applications on streaming data using Spark Streaming, with additional support from storage options like Apache HBase and Apache Kudu.
- **Batch Processing (Data at Rest):** Apache Spark provides the open standard for flexible, in-memory data processing across a variety of workloads—including batch processing, advanced modeling, and analytics. As an integrated part of Cloudera's platform, Spark benefits from unified resource management (through YARN), simple administration (through Cloudera Manager), and compliance-ready security and governance (through Apache Sentry and Cloudera Navigator)—all critical for running in production.
- **Scalable Manufacturing and IoT Data Platform:** Scales easily and efficiently based on data growth—enabling an enterprise to store unlimited amounts of data. More importantly, the platform enables you to effortlessly combine IoT/sensor data with other internal and external data sources to ensure interoperability and drive deeper business insights.
- **Deployment Flexibility:** Deploy the platform on-premises, in the cloud, or in a hybrid environment based on the needs of your business—while still benefitting from centralised management.
- **Fundamental Security:** Security is paramount, especially when it comes to IoT. With Cloudera, organisations can take advantage of the only compliance-ready Hadoop platform with multiple layers of security and industry-leading security tools.
- **Fast Analytics:** Open up this data to self-service business intelligence and analytics with tools like Apache Impala (incubating), machine learning libraries, and integrations with leading BI partner tools.

Specifically for IoT, Cloudera also works closely with Intel and our ecosystem of partners in order to provide customers with end-to-end solutions—including everything from sensors, gateways, device management, data analytics, security capabilities, etc.—to accelerate the complete IoT journey.

Driving Value for Manufacturing

- Flexible data ingest with Kafka and Flume
- In-memory data processing using Spark
- Real-time processing with Spark Streaming
- Fastest time to insights using Impala
- Fast analytics on fast data with Kudu
- Out-of-the-box machine learning libraries
- Easy cloud deployment using Cloudera Director

cloudera®

Partner ecosystem

Speed to production is key to unlocking the potential of unlimited data. That's why Cloudera has a comprehensive ecosystem of partners to help organisations accelerate toward becoming truly data-driven. Covering cloud, platform, and software, partners' expertise covers areas such as data integration (like Datavard for all things SAP and SAP HANA), BI and analytics (for example, SAS supply chain analytics), security (like Centrify's unified identity management), and everything in between.

Customer Value

- AMD improves yield predictions, lowers infrastructure and software TCO, and now stores 90 percent of all data elements for 1.5+ years
- Lineage logistics increases cold storage capacity by 30+ percent through data-driven warehouse layout redesign
- Siemens helps organisations with global value chains save $15–25M
- NetApp improves customer support through queries that run 64 times faster

Conclusion

Cloudera Enterprise, powered by Apache Hadoop, has shifted the paradigm in manufacturing data management and analytics. Utilising the power of Hadoop, organisations can now easily ingest and store unlimited volumes and varieties of sensor, process, and IoT data, use powerful processing and analytics tools across data in motion as well as data at rest, and provide immediate search, query, and visualisation across petabytes of data to drive actionable insights. With Cloudera Enterprise, organisations are able to benefit from the power of Hadoop while leveraging Cloudera's industry-leading data security and management tools that are critical to manufacturing production deployments.

About Cloudera

Cloudera delivers the modern platform for data management and analytics. The world's leading organisations trust Cloudera to help solve their most challenging business problems with Cloudera Enterprise, the fastest, easiest, and most secure data platform built on Apache Hadoop. Our customers can efficiently capture, store, process, and analyse vast amounts of data, empowering them to use advanced analytics to drive business decisions quickly, flexibly, and at lower cost than has been possible before. To ensure our customers are successful, we offer comprehensive support, training, and professional services. Learn more at cloudera.com.

cloudera.com

1-888-789-1488 or 1-650-362-0488
Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA