



Implementing predictive maintenance using IoT can reduce equipment downtime by as much as 50 percent and reduce maintenance costs of factory equipment by 10 to 40 percent.

– Source - McKinsey Global Institute

Characteristics of an IoT Data platform

An effective data management platform for IoT should

- Scale easily and efficiently based on IoT data growth
- Handle the volume & variety of IoT data
- Drive cost efficiency - Low cost/ TB
- Effectively handle both data at rest and data in motion
- Easily ingest and process data in real-time
- Be fundamentally secure
- Support both Cloud & On -prem deployments

Driving Predictive Maintenance in a Connected World

IoT Enabled Predictive Maintenance

Today, IoT is fundamentally transforming the maintenance model paradigm from repair-and-replace to predict-and-prevent. By capturing and utilizing data streaming from sensors and devices, businesses can now gain visibility into the condition of their valuable assets and their specific components in real time.

The key to predictive maintenance is to monitor the condition of the machine/ assets in real-time, and intervene before the machine actually fails or breaks down. Utilizing IoT and sensor data from equipment, predictive maintenance enables organizations to effectively predict when and how an asset might fail by detecting variances, understanding warning signals, and tracking any patterns that may indicate a potential breakdown. Leading organizations are now using machine learning to accurately predict the chances of a machine being down — including when and how an asset will fail. This enables early and corrective measures to be planned and introduced in the most effective way, thereby avoiding unplanned downtimes and costly staff and resources.

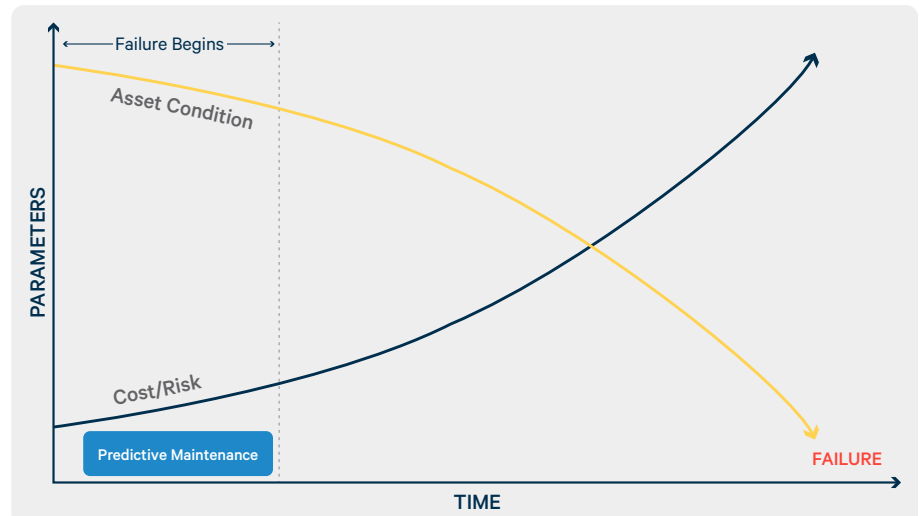


Figure 1: Condition-Based Monitoring

Predictive maintenance can have a profound impact, especially on asset-heavy industries such as manufacturing, automotive, mining, energy and utilities, etc., delivering lowered costs associated with maintenance, lowered equipment downtimes, and improved productivity and quality.

Key Data Management Challenges

Even though there is significant value to be gained from continuous monitoring of assets and driving predictive maintenance, organizations struggle to keep up with the volume, variety, and velocity of the IoT data streaming in from millions of sensors in real time. Some of the key challenges organizations face with respect to data management for predictive maintenance include:

- **Cost of Data Management:** Traditional data management mechanisms tend to be notoriously expensive, do not scale easily, and are not ideal for capturing and processing the petabytes of data streaming from connected machinery. Organizations need a data management platform that can easily ingest, store, manage, and process all of these streaming data at a **lower cost per terabyte**.

Cloudera for IoT

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

- **Ability to handle the volume and variety of IoT Data:** To enable continuous monitoring and IoT, organizations need a different kind of data platform that can handle **diverse data structures and schemas**, including everything from intermittent readings on temperature, pressure, or vibrations per second to handling fully unstructured data (e.g., images, video, text, spectral data) or other forms such as thermographic or sound/noise from machines.
- **Managing the complexity of real-time data:** In order to drive continuous monitoring and predictive maintenance, organizations need a platform that can ingest, store, and process the data streaming in from sensors in **real time** or near-real time in order to instantly drive insights.
- **Diverse Analytical Capabilities:** Existing platforms offer limited ability to provide insights and analytics into real-time asset performance and condition. In order to effectively drive predictive maintenance, organizations need a platform that can provide a **wide range of analytical options**, including SQL analytics, search, machine learning capabilities, and integration with business intelligence (BI) tools.
- **Predictive Modeling Capabilities:** Current platforms provide limited to no modeling or machine learning capabilities to predict and prevent issues before they impact the operations.

Thus, in order to effectively drive condition-based monitoring and predictive maintenance, organizations need a scalable, elastic, and cost-effective data management platform that can handle the volume, variety and velocity of data presented by IoT. Furthermore, the platform should be able to handle the complexity of both data at rest as well as data in motion, offer enterprise-grade security and management tools, and deliver a range of analytical options including proven machine learning and predictive modelling capabilities.

Apache Hadoop—The Data Management Platform for IoT

More and more organizations in industrial, manufacturing, automotive, and utilities are utilizing Cloudera's Enterprise Data Hub (EDH), powered by Apache Hadoop, to drive predictive maintenance and other IoT use cases in order to **improve performance, reduce costs, and enable insights** in real time.

With **Cloudera Enterprise**, organizations can easily bring together real-time sensor data from machines together with data from diverse manufacturing systems and multiple sources into a single, unified platform at considerably lower cost. Since Hadoop is built on a highly scalable and flexible file system, any type of data—including structured data from manufacturing systems or historians or streaming data from sensors—can be loaded into the platform without altering its format, in order to perform machine learning and analytics to predict failures. Data generated by machines and sensors, including log files, can be collected in real time and streamed directly into an EDH instead of being staged in a temporary file system or data mart.

And because Hadoop uses industry-standard hardware, the cost per terabyte of storage is, on average, **10 to 100 times** cheaper than a traditional data management mechanisms.

Cloudera for IoT and Predictive Maintenance

Some key attributes of Hadoop and Cloudera Enterprise that make them perfect for driving IoT-enabled predictive maintenance include:

- **Real-Time Data Ingest:** Easily ingests data from multiple data sources, and supports both batch and real-time data ingest from sensors using tools such as **Apache Kafka** and **Apache Flume**.
- **Manage Data Variety:** Provides a platform that can handle all types of structured and unstructured data sources. Based on the specific use case, you can easily bring in multiple data types, structures, and schemas—including structured data from diverse manufacturing systems to intermittent sensor readings of temperature and pressure or live video and audio feeds.
- **Real-Time Data Processing and Insights:** Supports real-time data processing on streaming data using **Apache Spark** and **Spark Streaming**, supported by storage options like **Apache HBase** and **Apache Kudu**.

- **Machine Learning Capabilities:** Provides out-of-the-box machine learning libraries using Apache Spark that enable organizations to easily build predictive models and continuously iterate on them.
- **Diverse Analytical Options:** Offers diverse analytical engines, including search and SQL analytics, with tools such as [Apache Impala](#) (incubating), statistical analysis, and [machine learning libraries](#), to suit the diverse needs within your organization. Moreover, Cloudera is already integrated with industry-leading BI and visualization tools, so that you can continue to leverage your existing BI tools and assets.
- **End-to-End Data Security:** Delivers complete protection of sensitive data, including total encryption, key management, access policy enforcement, audit, lineage, and complete data governance capabilities using [Cloudera Navigator](#).

Cloudera for Predictive Maintenance

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

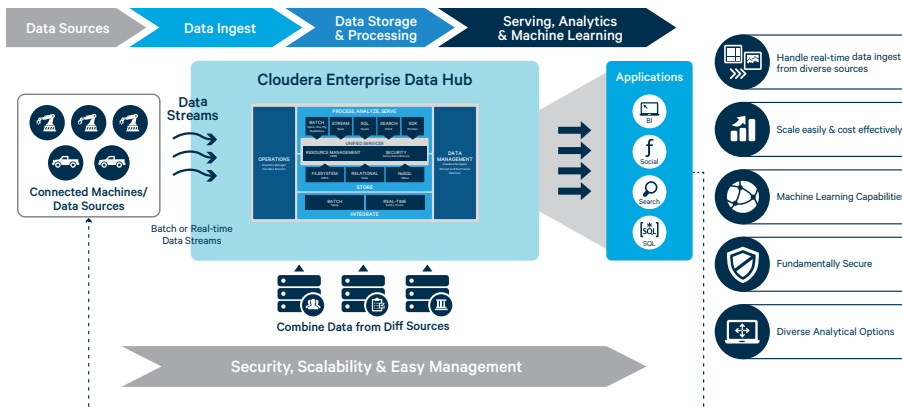


Figure 2: IoT Data Management Value Chain Using Cloudera Enterprise

Solution Delivery Models

When it comes to putting all of this together, Cloudera offers a variety of options in deploying the predictive maintenance solution. Organizations have a choice of a) building a custom solution, b) assembling a solution using certified technology partners, or c) buying a prepackaged solution from systems integrators to implement the use case.

- **Build:** Organizations can work with Cloudera’s professional services team to build a customized predictive maintenance solution that fits their specific needs. Our service professionals will work hand in hand with client teams to engineer a solution that can be implemented on-premise, in the cloud, or in a hybrid environment, leveraging the appropriate components of data ingestion, storage, and processing mentioned earlier. A typical implementation time for a build model ranges from eight to nine weeks.
- **Assemble:** Cloudera can also work with clients to assemble a solution using the Cloudera Enterprise platform along with some of the leading technology partners from our partner ecosystem. These partners have created various components and accelerators for simplifying and providing faster time-to-value for data ingestion, data processing, or analytics. Cloudera certifies that these technologies and solution providers go through a rigorous certification process. Depending on the partner solutions deployed, a typical engagement will take around nine weeks for implementation with flexible deployment options.
- **Buy:** A third option is to buy a pre-built and proven solution from our partners that is deployed on top of Cloudera’s enterprise Hadoop platform. Some of the leading Systems Integrators (SI) have created end-to-end solutions for managing and implementing remote monitoring and predictive maintenance, using Cloudera Enterprise, that offer pre-built integrations, configurations, and compatibility to get started quickly. An example of one such solution is the [TCS Sensor Data Analytics Framework](#), which is a highly customizable big data framework, powered by Cloudera, to collect, store and analyze a variety of sensor and log data from IoT deployments. Examples of its applications include predictive maintenance, machine diagnostics, telematics processing, remote monitoring, and early warning systems.

Key IoT & Predictive Maintenance Use Cases Supported:

- Connected Vehicles
- Connected Factories
- Industrial IoT
- Heavy Machinery
- Smart Buildings
- Smart Ports
- Smart Airports

Proven IoT Leadership and Experience

Leading organizations across the globe are increasingly adopting Cloudera's Enterprise Data Hub, powered by Apache Hadoop, as the data management and analytics platform for driving predictive maintenance and IoT. Below is a summary of a diverse set of use cases that highlights how some of our customers are utilizing Cloudera and the power of Apache Hadoop to drive predictive maintenance.

Predictive Maintenance—Customer Use Cases

Setting	Use Cases	Customer Case Study—Description
Automotive	Predictive Maintenance – Connected Vehicles	One of the leading auto manufacturers in North America is using Cloudera as the data management platform to monitor the health of 180,000+ trucks in real time in order to improve uptime and reduce fleet maintenance costs by 30 to 40 percent.
Manufacturing	Predictive Maintenance – Industrial IoT	A leading industrial automation company is utilizing Cloudera in an IoT setting to ingest, store, and analyze petabytes of sensor data from thousands of diverse manufacturing systems, in real time, in order to eliminate machine downtime.
Heavy Machinery	Predictive Maintenance – Heavy Machinery	One of the biggest heavy equipment fleet manufacturers in North America is using Cloudera to parse large-volume and high-velocity data from sensors to continuously monitor performance of their fleet and to do predictive maintenance as well as advanced defect detection.
Buildings/Airports	Predictive Maintenance – Smart Buildings	One of the busiest airports in Europe is running Cloudera on Azure to capture, secure, and correlate sensor data collected from equipment within the airport (e.g., escalators, elevators, and baggage carousels) to prevent breakdowns and improve airport efficiency and passenger safety.
Ports	Predictive Maintenance – Smart Ports	A leading provider of cargo-handling solutions is utilizing Cloudera to ingest and process IoT data that is streaming from sensors in port terminal machinery, including cranes and cargo-handling equipment, to improve operational efficiencies and increase uptime.

Conclusion

Cloudera Enterprise, powered by Apache Hadoop, has shifted the paradigm in IoT-driven predictive maintenance. Utilizing the power of Hadoop, organizations can now easily ingest, store, process, and analyze unlimited volumes and varieties of sensor data; use powerful processing and analytics tools across data in motion as well as data at rest; and leverage machine learning capabilities across petabytes of data to drive real-time predictive maintenance. With Cloudera Enterprise, organizations are able to benefit from the power of Hadoop while leveraging Cloudera's industry-leading management and data security tools that are critical to IoT production deployments. Cloudera makes Hadoop fast, easy, and secure, and, together with our industry-leading partners, we can accelerate the time-to-value from your IoT investments.

About Cloudera

Cloudera delivers the modern platform for data management and analytics. The world's leading organizations 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.

cloudera.com

1-888-789-1488 or 1-650-362-0488

Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA

© 2016 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice.