Powering the Internet of Things with Apache Hadoop: Four Customer Use Cases

cloudera®

Table of Contents

Introduction

With over 30 billion things connected by 2020,[1] the Internet of Things (IoT) has the potential to be one of the most disruptive technological advancements in recent ages. Organizations are starting to leverage IoT to drive key business objectives, including monitoring and improving product performance, driving internal efficiencies, improving customer experience, and enabling new revenue and usage-based business models.

However, IoT will also drive an explosion of data that will need to be processed, stored, managed, analyzed, and served (in some cases in real time) to enable business value. With millions of data events generated per minute from a multitude of sources, IoT will create far greater volume and variety of data than most information leaders are currently familiar with—requiring a modernization of information infrastructures to realize value.

Generating value from IoT entails effectively managing both data at rest as well as data in motion. In fact, the success of IoT deployments will depend on the ability of organizations to gain insights out of all this fast-moving, high-volume data. The need to drive real-time insights means that each step of the data-in-motion process—ingestion, processing, and serving—is critical.

A Data Management Platform for IoT

Today, leading organizations worldwide 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 IoT data.
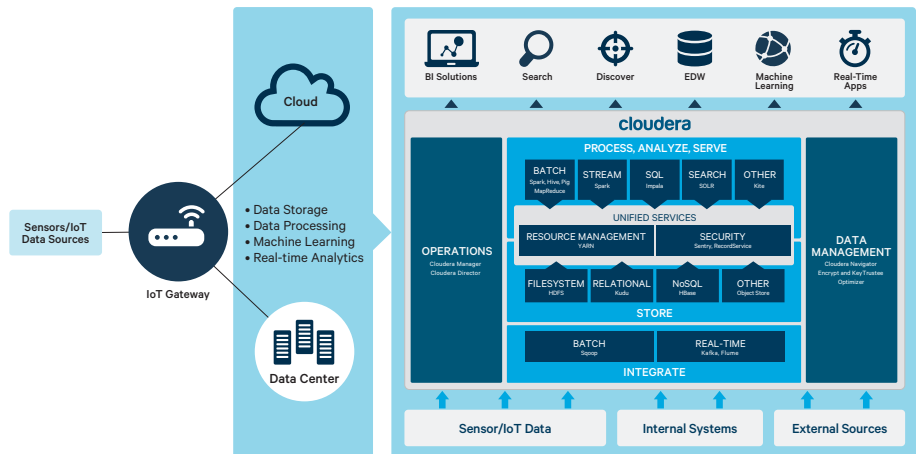

Figure 1: Cloudera Enterprise—The Data Management Platform for IoT

With Cloudera Enterprise, organizations can easily bring data from multiple sources onto a single, unified, secure platform—combining and correlating IoT sensor data streams with activity logs, transaction data, customer data, external data, and much more. Any type of data (whether structured, unstructured, or semi-structured) can be loaded into Cloudera Enterprise without altering its format—preserving data integrity and delivering complete analytic flexibility.

Cloudera Enterprise provides a scalable, elastic, and enterprise-grade data management platform that enables easy ingestion, storage, processing, and analytics of both data in motion and data at rest. Organizations can easily ingest both batch and streaming data from sensors using tools such as Apache Kafka and Apache Flume. It also supports both batch and real-time data processing data using Apache Spark and Spark Streaming, supported by storage options like Apache HBase and Apache Kudu. Also, depending on the use cases, Cloudera Enterprise provides a range of analytical options, including SQL analytics with tools like Apache Impala (incubating), machine learning libraries, and integration with leading business intelligence (BI) solutions.

Furthermore, because it runs on industry standard hardware, appliances, or in the cloud, the cost per terabyte of storage and processing is, on average, ten times less expensive than traditional data management mechanisms.

A number of leading organizations across diverse industry verticals—automotive, manufacturing, utilities, industrial automation, insurance, healthcare, telecom, and technology—have already adopted Cloudera Enterprise as their data management platform for IoT, ingesting and processing millions of events per second to generate actionable business insights.

In fact, Cloudera is powering some of the most compelling IoT use cases in the industry today, including connected vehicles and telematics, predictive maintenance to connected homes, industrial IoT, smart cities, usage-based insurance, and healthcare IoT. Take a look.

Case Study 1: Using IoT to Lower Downtime for Connected Vehicles

Use Case: Connected Vehicles/Predictive Maintenance
Customer: Leading automotive manufacturer
Geography: North America

Overview
With revenues of more than US$10 billion in 2015, this customer is one of the leading North American automotive manufacturers offering a wide range of commercial trucks, buses, defense vehicles, and engines.

Business Drivers
Real-time View of Vehicle Performance
- To truly understand truck performance and improve vehicle uptime, the automotive manufacturer wanted to capture and analyze detailed data on more than 180,000 trucks in real time—including everything from speed, braking, and engine performance to fuel economy, geolocation, idle times, potential failures, and more.
- They also wanted to gain insights into how their trucks perform under a variety of conditions, and spot engine troubles well before the vehicles break down so as to minimize failures and fleet downtime.

Solution
Real-time Remote Diagnostics Solution Built on Cloudera
- The manufacturer developed a remote diagnostics tool, leveraging Cloudera Enterprise to collect telematics and geolocation data across their trucks in real time.
- The sensors in the vehicle capture and stream data on everything from speed, engine performance, and fuel efficiency to coolant temperature and brake wear. The data is collected at 15- to 60-second intervals and ingested into Cloudera Enterprise for advanced analytics.
- Engineers and analysts then use various analytical tools and machine learning capabilities to identify engine problems early, predict maintenance issues, and address them before a truck breaks down.

Impact
A Better Customer Experience Coupled with Reduced Fleet Downtime
- **Reduced downtime:** The solution enabled fleet owners to easily monitor the performance of their fleet in real time from their smartphones or tablets, proactively schedule maintenance, and reduce unplanned repairs and vehicle downtime by as much as 30 percent.
- **Lowered maintenance costs:** With real-time monitoring of their vehicles, fleet owners were able to reduce maintenance costs by as much as 80 percent per mile (from US$0.12–US$0.15 per mile to US$.03 per mile).
- **Enabled new services:** The manufacturer now offers analytics as a service to fleet owners and plans to build an online portal that integrates telematics data with additional GPS data and parts inventory information. This offering will enable fleet owners to quickly find the nearest dealer service location that has the parts they need in stock and service locations that have available technicians and bays.

> The remote diagnostics IoT platform built on Cloudera's data hub has helped reduce fleet maintenance costs by an average of 30 to 40 percent.

cloudera

Case Study 2: Delivering Usage-based Insurance and Telematics

Use Case: Automotive Telematics/Usage-based Insurance
Customer: Large European auto insurance company
Geography: Europe

Overview

A large European auto insurance company uses Cloudera and telematics to gather, store, and analyze data in real time from black box devices installed in their clients' vehicles. The company uses this information to adjust rates and deductibles based on each individual's driving patterns, mitigating risk while increasing both profit and market share.

Business Drivers

Improve Competitiveness, Reduce Fraud, and Deliver Value-added Services

- Driven by competitive pressures as well as an increasing average cost of damages per claim, the auto insurance company wanted to effectively launch a new line of service, delivering usage-based insurance (UBI) to their customers.

- Furthermore, the company needed a telematics-based solution to help them reduce fraud and also offer a variety of new value-added services to customers, including roadside and accident assistance, stolen vehicle recovery, remote diagnostics fuel management, and more.

Solution

A Telematics-based Usage Solution Using Cloudera Enterprise and Intel Architecture

- Cloudera, together with Intel, enabled the insurer to launch a telematics-based usage insurance solution that allowed it to personalize coverage and charge customers based on their actual driving habits.

- A black box fitted to a participating customer's car constantly records information, such as GPS location, driving speed, distance and time of drive, type of acceleration and braking (e.g., rapid or smooth), and cornering habits.

- Cloudera Enterprise is used to gather, store, and analyze this large volume of telematics data, gathered in real time from the black box devices, along with data from other external sources, such as weather data, traffic reports, and collisions involving other vehicles. The data is extracted, cleansed, transformed, and aggregated into Apache Impala (incubating) and Apache HBase tables.

- The company then analyzes this data to create a personalized record of each customer's driving habits. By focusing on the individual characteristics of a customer's driving tendencies, the company will improve their ability to accurately predict the odds and cost of the customer filing a claim, and can adjust rates and deductibles accordingly.

Impact

Safer Driving and Decreased Claims

- **Decreases claims by 30 percent:** The insurer can now make more consistent and accurate predictions based on drivers' real-world habits and driving conditions, and adjust rates accordingly. The new insight has significantly helped minimize fraudulent claims and reduce the number of all claims by 30 percent.

- **Attracts and retains low-risk (safer) drivers:** With the ability to offer lower cost incentives to safer drivers, the company has been able to increase their user base and, at the same time, improve their profitability.

- **Encourages good driving habits:** With the advanced analytics Cloudera Enterprise delivers, the insurer can provide their customers with an online portal to study and improve their driving skills, and encourage high-risk drivers to change their driving habits.

> The new insight has significantly helped minimize fraudulent claims and reduced the number of all claims by 30 percent.

cloudera®

Case Study 3: Enabling Predictive Maintenance in an Industrial Setting

Use Case: Industrial IoT/Predictive Maintenance
Customer: Leading industrial automation organization
Geography: North America

Overview

One of the leading global industrial automation companies in North America complements their product offerings with factory management software, motion control sensors, and machine safety components to help manufacturers improve efficiencies and increase equipment and production uptime.

Business Drivers

Real-time Operational Monitoring to Drive Efficiencies

- The company needed the ability to monitor, store, and analyze data from thousands of diverse manufacturing systems across multiple factories, in real time, in order to do predictive analytics and drive operational efficiencies.

- With customers generating almost 50 gigabytes (GB) to 1 petabyte (PB) of data per factory per month, including sensor data that measures everything from the speed, force, temperature, pressure, and revolutions per minute of factory equipment, the company was looking for a scalable data management platform that can process enormous data volumes, including both data at rest as well as data in motion.

Solution

A Cloud-based Connected Factory Solution Running on Cloudera Enterprise

- Through their industrial IoT solution, based on Cloudera Enterprise and running on Microsoft Azure, the company was able to leverage predictive analytics and machine learning capabilities across petabytes of IoT data to help their customers reduce production downtime and increase competitiveness.

- A proprietary diagnostic and analytics application, built on top of Hadoop, enabled users to continuously monitor critical machine data signatures like torque, speed, and position errors, and determine the mechanical wearing of the movers for supporting component-level maintenance.

- By running Cloudera Enterprise in the cloud, data and analytics could easily be delivered as a service, enabling them to offer real-time apps, machine learning, and business intelligence (BI) solutions to help customers more efficiently manage their production resources. Customers using the company's IoT solution can now effortlessly monitor the health and functioning of their machinery in real time via a tablet or mobile device to ensure smooth operations.

Impact

Predictive Maintenance and Real-time View of Industrial Assets

- **Achieving real-time predictive maintenance:** By processing and analyzing time-series sensor data from production systems, the company can now effectively detect signs of mechanical wear and equipment issues well before they are visible to factory staff, so that they can take action before a failure occurs.

- **Driving toward zero downtime:** By combining sensor data streams from factory floors with structured data from internal and external systems, company engineers are now able to spot and fix emerging problems before they affect production.

- **Reducing costs:** Reduced downtime helps prevent millions of dollars in lost production. Additionally, when maintenance is needed, machine learning capabilities help the company match the right technician or engineer with a problem more quickly, enabling them to scale service worldwide without investing millions in additional resources.

> Reduced downtime helps prevent millions of dollars in lost production, and, when maintenance is needed, machine learning capabilities help the company match the right technician or engineer with a problem more quickly.

cloudera®
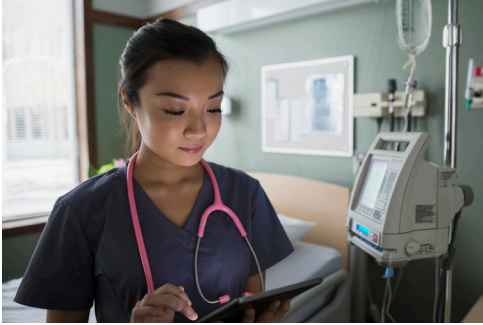
Case Study 4: Revolutionizing Healthcare and Parkinson's Research

Use Case: Healthcare IoT and Wearables
Customer: The Michael J. Fox Foundation for Parkinson's Research
Geography: North America

Overview

An estimated seven to 10 million people live with Parkinson's, a neurodegenerative disease, which is second only to Alzheimer's in its prevalence. The Michael J. Fox Foundation for Parkinson's Research (MJFF) is dedicated to finding a cure for Parkinson's disease (PD) through an aggressively funded research agenda, and to ensuring the development of improved therapies for those living with Parkinson's today. Established by actor Michael J. Fox in 2000, the foundation has since become the largest nonprofit funder of Parkinson's disease research in the world, investing more than $450 million in research to date.

Business Drivers

Improving Parkinson's Data Collection and Management

- One of the biggest challenges in Parkinson's research was the limited data available—the data came from small sample sizes, was collected infrequently, and was very subjective.

- Parkinson's disease was still subjectively measured largely the same way it had been about 200 years ago. Typically, data collection for research came from patient visits to the doctor's clinic, maybe once every three months, and also from patient self-evaluations and journals.

- During office visits, patients would report their symptoms and doctors would ask them to perform several activities, which would then be subjectively ranked by the physicians.

Solution

Intel Wearable-based Solution Built on Cloudera

- Intel has partnered with MJFF for a multiphase research study that uses wearables and a big data analytics platform (built on Cloudera) to monitor symptoms, collect data, and detect patterns in participant data.

- Using wearable devices to unobtrusively gather and transmit objective, experiential data in real time, 24 hours a day, seven days a week.

- A big data analytics platform based on Cloudera CDH was built to collect, store, manage, and drive analytics on all of the data streaming from wearables. The data platform is deployed on a cloud infrastructure optimized on Intel architecture, allowing scientists to focus on research rather than the underlying computing technologies.

- The infrastructure can manage up to 1GB of data per patient per day, which can be easily scaled as the data and the sample size grow.

Impact

Revolutionizing Data for a Cure for Parkinson's Disease

- **Better understanding Parkinson's disease:** Using wearables technology, MJFF can now easily gather and bring together experiential data (including slowness of movement, tremor intensity, sleep quality, and more) in real time to assemble a better picture of the clinical progression of Parkinson's and track its relationship to molecular changes.

- **Improving data access and quality:** With this new platform, researchers could go from looking at a very small number of data points to collecting over 300 readings per second from thousands of patients and attaining a critical mass of data to detect patterns and make new discoveries.

- **Creating a centralized data repository for researchers:** The platform currently acts as the centralized repository with the largest collection of data on Parkinson's. In the near future, the platform could store other types of data, such as patient, genomic, and clinical trial data. In addition, the platform could enable other advanced techniques, such as machine learning and graph analytics, to deliver more accurate predictive models that researchers could use to detect change in disease symptoms.

> Researchers could go from looking at a very small number of data points to collecting over 300 readings per second from thousands of patients and attaining a critical mass of data to detect patterns and make new discoveries.

cloudera®

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. Our customers can efficiently capture, store, process, and analyze 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