

IoT

Driving Value from the Internet of Things (IoT)

IoT—A revolution in the making

IoT—A Connected Revolution**IoT MARKET POTENTIAL - 2020**

- _ 30 billion connected things
- _ \$1.7 trillion in value
- _ 15 - 20% growth YoY
- _ Over 250 million connected cars

Characteristics of an IoT Data platform**AN EFFECTIVE DATA MANAGEMENT PLATFORM FOR IoT SHOULD:**

- _ Scale easily and efficiently based on IoT data growth
- _ Drive cost efficiency - Low cost/TB
- _ Support multiple data-types and structures
- _ Effectively handle both data at rest and data in motion
- _ Support analytics at the edge while enabling machine learning
- _ Be fundamentally secure
- _ Support hybrid Cloud environments

With billions of devices—including cars, homes, airplanes, apparels, parking meters, wearables, factories, oil rigs, and heavy machinery—connected to the internet, the Internet of Things (IoT) has the potential to be the most disruptive technological advance in recent ages. According to research from IDC, about 30 billion things will be connected by 2020¹, helping enterprises drive efficiencies and launch new products and services.

As adoption continues to increase, and as new use cases emerge, organizations are starting to leverage IoT as an engine to improve and innovate product performance, drive internal efficiencies through predictive maintenance, and enhance customer experience.

IoT data—A new paradigm for enterprises

With 30+ billion things connected, IoT will drive an explosion of data that will need to be processed, stored, managed, analyzed, and served (in some cases in real time) to drive business value. IoT will generate far greater volume and variety of data than most information leaders are currently familiar with—requiring a modernization of information infrastructure to realize value.



Some of the key characteristics of IoT data include:

- _ Massive volumes of intermittent data streams—amounting to millions of events per minute
- _ Predominantly time-series data
- _ May come in streams (real-time) or batches
- _ Generated from a variety of data sources—from sensor readings of temperature and pressure to live video streams
- _ Can include diverse data structures and schemas based on the sources
- _ May become perishable (value of data decreases over time)

Given this complexity and variety of IoT data, organizations must fundamentally re-think their data management strategy—transitioning to a platform that is optimized for the scale and complexity that IoT presents. More importantly, the real value from IoT can only be exploited if organizations have the ability to combine and correlate these sensor data streams with data from other internal and external data sources; and, in some cases, even combining this with data from other IoT ecosystems.

¹ Source: <http://blogs.wsj.com/cio/2015/06/02/internet-of-things-market-to-reach-1-7-trillion-by-2020-idc/>

Cloudera for IoT

- _ Effectively handle both data at rest and data in motion
- _ Easily ingest millions of events/sec
- _ Edge analytics & machine learning
- _ 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

Key IoT Use Cases Supported:

- _ Connected Vehicles
- _ Predictive Maintenance
- _ Industrial IoT
- _ Usage Based Insurance
- _ Telematics
- _ Connected Homes
- _ Smart Buildings
- _ Smart Ports
- _ Smart Healthcare

To realize the full potential from IoT, and extract actionable intelligence, organizations and underlying architectures will need to evolve. 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.

Cloudera Enterprise—The data management platform for IoT

Given the characteristics of IoT data streams, leading organizations around the globe are adopting Cloudera Enterprise as the data management and analytics platform for storing, managing, processing and, more importantly, driving analytics and intelligence from all of their IoT data.

With Cloudera Enterprise, organizations can easily bring information from multiple sources onto a single, unified platform at considerably lower cost per terabyte. This includes sensor data, activity logs, transaction data, customer data, and more. And because the platform 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 and 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 systems 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, 10x cheaper than a traditional relational data warehouse system.

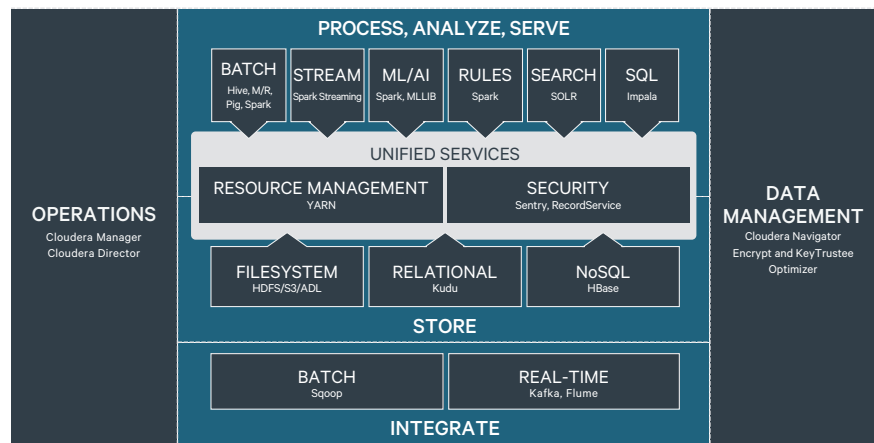


Figure 1: Cloudera Enterprise - Making Hadoop fast, easy and secure

Using Cloudera to Drive Business Value from IoT

Some key attributes that make Cloudera Enterprise perfect for 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 visualization with our partner [StreamSets](#). Using StreamSets Data Collector and StreamSets Data Collector Edge, organizations can easily develop and operate data flows and manage complex ingest pipelines from a variety of IoT data sources
- _ **Handles Data Variety:** Effectively handles multiple IoT data-types, structures, and schemas—from intermittent sensor readings of temperature and pressure to real-time location data or streaming live video feeds

- End-to-End Analytics & Machine Learning:** Enables real-time analytics at the edge (closer to the data source) or at the centralized hub, while enabling machine learning to drive intelligence and action
- 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.

Driving Value from IoT

- Flexible data ingest with Kafka & 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

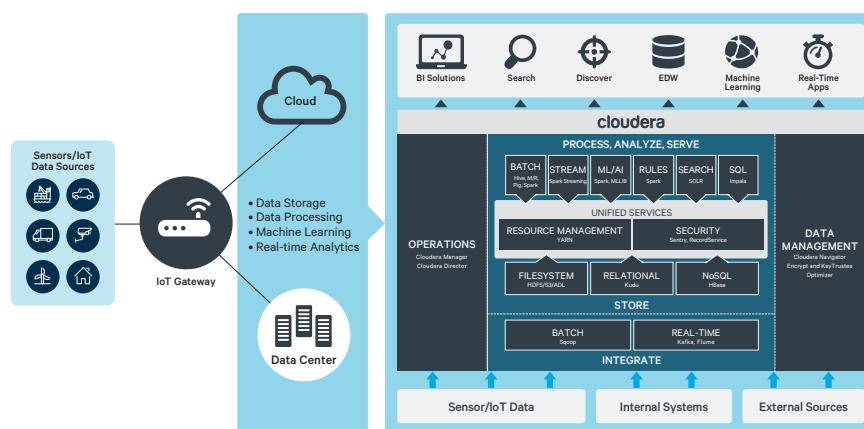


Figure 2: Cloudera Enterprise - The Data management platform for IoT

- Scalable IoT Data Platform:** Scales easily and efficiently based on the 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-premise, in the cloud, or in a hybrid environment based on the needs of your business—while still benefitting from centralized management
- Fundamentally Secure:** Security is paramount when it comes to IoT. With Cloudera, organizations 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](#), [machine learning libraries](#), and integrations with leading BI partner tools

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

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/second to generate actionable business insights. From connected vehicles and telematics, predictive maintenance to connected homes, industrial IoT, smart cities, usage based insurance, and healthcare IoT, Cloudera is powering some of the most compelling IoT use cases in the industry today.

Getting started is easy. [Download](#) Cloudera to get started today.

Conclusion

Cloudera Enterprise, the modern platform for machine learning and analytics, optimized for the cloud, has shifted the paradigm in IoT data management and analytics. Utilizing the power of Apache Hadoop, organizations can now easily ingest and store unlimited volumes and varieties of sensor 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 visualization across petabytes of data to drive actionable insights from IoT. With Cloudera Enterprise, organizations are able to benefit from the power of enterprise-grade open source technologies, while leveraging Cloudera's industry leading data security and management tools that are critical to IoT production deployments.

About Cloudera

Cloudera delivers the modern platform for machine learning and advanced analytics built on the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems by efficiently capturing, storing, processing and analyzing vast amounts of data.

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