

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
- Easily ingest data in real-time
- Be efficient in real-time data processing and analytics
- Be fundamentally secure
- Support hybrid cloud environments

Driving Value from the Internet of Things (IoT) in the Public Sector

IoT—A revolution in the making

With billions of devices managed by the public sector—embedded in cars, roads, parking meters, wearables, buildings, utilities, transit vehicles—connected to the Internet, the Internet of Things (IoT) has the potential to be the most disruptive technological advancement in recent ages. According to research from IDC, about 32 billion things will be connected by 2020,¹ helping government agencies 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 differentiate services, drive internal efficiencies through predictive maintenance, better manage the public sector supply chain, and enhance the overall constituent experience.

IoT data—A new paradigm for enterprises

With 30-plus billion things connected, the Internet of Things (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 and crowd-sourcing data
- Can include diverse data structures and schemas based on the sources
- May become perishable (value of data decreases over time)

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
- 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

Given this complexity and variety of IoT data, public sector agencies must fundamentally rethink 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 be exploited only 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 outside of the public sector.

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 in government 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—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.

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, citizen 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 and application and weblog files, can be collected in real time and streamed directly into Cloudera Enterprise—instead of being staged in temporary file systems or data marts. And because Hadoop uses industry standard hardware, the cost per terabyte of storage and processing is, on average, 10 times cheaper than a traditional relational data warehouse system.

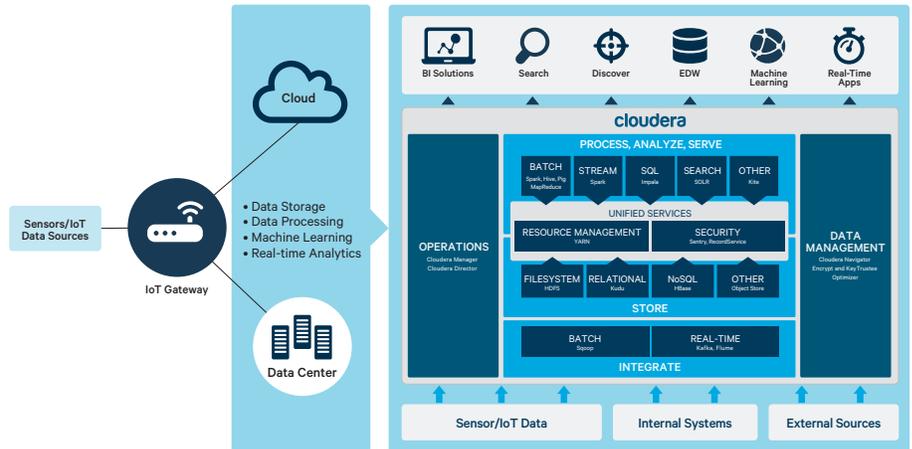


Figure 1: Hadoop as the data management platform for IoT

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

- **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](#)
- **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
- **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](#) (incubating)
- **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 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 organizational insights
- **Deployment Flexibility:** Deploy the platform on-premises, in the cloud, or in a hybrid environment based on the needs of your agency—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](#) (incubating), [machine learning libraries](#), and integrations with leading BI partner tools

A number of leading organizations across the public sector have already adopted Cloudera Enterprise as their data management platform for IoT, ingesting and processing as much as one million events per second to generate actionable insights. From connected military vehicles and predictive maintenance to road sensors, smart cities, and healthcare wearables, Cloudera is powering some of the most compelling IoT use cases in the industry today.

Conclusion

Cloudera Enterprise, powered by Apache Hadoop, has shifted the paradigm in IoT data management and analytics. Utilizing the power of 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 Hadoop while leveraging Cloudera's industry-leading data security and management tools that are required for production deployments.

IoT Use Cases Supported:

- Connected Vehicles
- Predictive Maintenance
- Smart Roads
- Smart Parking
- Transit Planning and Optimization
- Smart Buildings
- Smart Ports
- Smart Healthcare
- Military Supply Chain Tracking



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

© 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.