



## Customer Insight

# BUILDING A DATA LAKE FOR INTEGRATION OF DATA ACROSS SILOS AND UNLOCKING NEW VALUE FROM DATA THROUGH CROSS-FUNCTIONAL ANALYSIS

////////////////////////////////////

**BUSINESS CHALLENGES**

- A need for infrastructure capable of storing vehicle driving data on a long-term basis
- A need for a Hadoop platform capable of deploying a variety of data cross-functionally

**DEPLOYMENT RESULTS**

- Introduction of a data lake capable of storing all types of company data
- An increase in demand for and deployment of data

////////////////////////////////////

Nissan Motor Co., Ltd (Nissan) collects a wide range of data. In order to deploy the data (which includes vehicle driving data and product quality data) cross-functionally, Nissan implemented an Apache Hadoop® platform that serves as a data lake. The company is well on its way to discovering high value-added uses for its data, which it plans to deploy in the field of IoT and across a global data lake.

### An Urgent Need for Infrastructure Capable of Storing and Analysing a Burgeoning Amount of Vehicle Data

Imagine connected cars—networked automobiles that can deliver a range of services leveraging on driving data, vehicle status, and nearby road conditions. This vision is one that is aggressively pursued by automobile manufacturers.

Nissan is one of them, and its efforts to work towards a society of connected cars go a long way back. In 1998, it established Compass Link to provide drivers with operator support. Then, in 2002, it launched CARWINGS, an upgraded, Japan-wide telematics information service which continues to this day. Later, in 2010, to coincide with the launch of the Nissan LEAF electric car, Nissan introduced the NissanConnect EV service.

Speaking about the company’s efforts to date, Mr Satoshi Kitsuki, General Manager, Nissan Global IT Division, said, “Nissan’s connectivity services are the means by which we deliver useful information, such as driving tips to maximize battery efficiency and battery charging points. In the course of doing so, the company has accumulated large volumes of driving data. This data is extremely valuable. Take for example battery usage data, which provides information on how, and the conditions in which the batteries were used. We are able to analyze how these factors affect battery degradation, and apply the findings to the company’s development and support functions. Previously, we lacked the infrastructure to store and analyze all this data on a long-term basis.”

### The Deciding Factor in Selecting Cloudera: Its Open-Source Nature

Nissan had Hadoop in mind as a solution for its big data problem.

“It was around 2014 that we began considering Hadoop. At that time, our Nissan Research Center Silicon Valley team was investigating the latest available technologies, and Hadoop was one of them. Cloudera’s offerings received high customer ratings compared to other products in the market. After discussions about whether the product would be as effective for use in Japan, we ultimately decided to go with Cloudera,” says Kitsuki.

“The fact that Cloudera’s Data Platform is the most open-source compliant product available made a big difference. Unlike other products, it did not require installation of any proprietary (non-open source) modules,” said Mr. Akinori Baba, Manager of IT Architecture and Production Services, Global IT Division.

Kitsuki went on to explain the benefits of open-source software:

“Many of Nissan’s projects make use of open-source technology. If I were to name two major benefits of open source, firstly, there are large numbers of open source engineering talent in

"The fact that Cloudera's enterprise data platform is the most open-source compliant product available made a big difference. Unlike other products, it did not require installation of any proprietary (non-open source) modules."

Mr. Akinori Baba, Manager of IT Architecture and Production Services, Global IT Division

#### About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

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

the market. Secondly, its open-source nature makes it easier to discontinue use of the technology, should circumstances change down the road."

The Cloudera Data Platform's foundation is a Hadoop distribution platform based on Apache Hadoop, YARN and the Hadoop Distributed File System (HDFS). Nissan runs the Apache Hive™, Apache HBase® data access modules and the Apache Oozie™ scheduling module on the platform.

Though the platform was introduced based on Nissan's long-term driving data storage and usage needs, the company is also experiencing an increased demand for big data analytics.

"Internal cross-functional data analysis is a growing need. For instance, when it comes to quality, there are many aspects to it, including design quality, production quality, and final product quality. Traditionally, different departments have been responsible for each aspect, but we've found that focusing on a single aspect of quality usually does not get to the root of problems. Instead, examining cross-functional quality data helped us understand, for example, that certain problems with final product quality were caused by something in the production process. Cloudera proved very effective in managing the large volume of quality data we had. Thus, it was deployed to other functions in order to meet internal demand," says Kitsuki.

#### Implementation of a Data Lake Capable of Storing All Types of Company Data

As part of Nissan's efforts to fully implement Cloudera's enterprise data platform, it is in the process of constructing a data warehouse.

"We've created the enterprise data platform that can house driving, quality, and other data. The basic structure will soon be in place, allowing deployment of a large data lake to host all kinds of company data. To promote awareness internally, we shared one of the key benefits of the platform with our employees — a dashboard presenting visualized data analysis results. The response was great, and the company's data utilization rate has seen a huge increase since."

Explaining the benefits of implementing the data storage platform, Kitsuki said,

"Nissan has a long history, but over the years, our data silo problem had gone worse. This is not the company's first attempt at integrating its data into a single warehouse. Unfortunately, our previous efforts to do so failed. However, this time, there is a sense that with Cloudera, we will finally overcome the problem."

The project is still in its early stages, but by 2017 it is expected that the platform will host up to 500TB of data.

#### The Future: Adoption of Other Modules for IOT Functionality

Nissan is currently investigating the viability of implementing the Spark module.

"Though the framework is presently intended to be used as a data lake, in the future we expect that it will have application in the IoT field. The plan is to deploy Spark with a machine learning library, creating a very short-range mesh network for analyzing fragmented IoT data and discovering trends," says Baba.

On top of this, Nissan is also working to set up a coordinated data warehouse across its locations globally.

"In North America, Europe, and other locations outside Japan, the Hadoop platform is being put in place. Data federation across these locations will enable sharing and creation of a global data lake. As business becomes more global, we need to utilize data from all regions effectively so as to enable new business. Instead of fragmented handling of data, our goal is the transparent presentation of data on the horizontal axis," says Kitsuki.

Nissan continues to expand into new areas of business by actively seeking fresh applications for its valuable data.