

NOKIA

Nokia: Using Big Data to Bridge the Virtual & Physical Worlds

KEY HIGHLIGHTS

INDUSTRY

Telecommunications

LOCATION

Global

BUSINESS APPLICATIONS SUPPORTED

- > Geospatial application development
- > Content/engagement optimization
- > Network sessionization

IMPACT

- > Enables unprecedented scale and flexibility to build 3D digital maps of the globe

Company Overview

Nokia has been in business for more than 150 years, starting with the production of paper in the 1800s and evolving into a leader in mobile and location services that connects more than 1.3 billion people today. Nokia has always transformed resources into useful products — from rubber and paper, to electronics and mobile devices — and today's resource is data.

Nokia's goal is to bring the world to the third phase of mobility: leveraging digital data to make it easier to navigate the physical world. To achieve this goal, Nokia needed to find a technology solution that would support the collection, storage and analysis of virtually unlimited data types and volumes.

Use Case

Effective collection and use of data has become central to Nokia's ability to understand and improve users' experiences with their phones and other location products. "Nokia differentiates itself based on the data we have," stated Amy O'Connor, Senior Director of Analytics at Nokia. The company leverages data processing and complex analyses in order to build maps with predictive traffic and layered elevation models, to source information about points of interest around the world, to understand the quality of phones, and more.

To grow and support its extensive use of Big Data, Nokia relies on a technology ecosystem that includes a Teradata enterprise data warehouse (EDW), numerous Oracle and MySQL data marts, visualization technologies, and at its core: Hadoop. Nokia has over 100 terabytes (TB) of structured data on Teradata and petabytes (PB) of multi-structured data on the Hadoop Distributed File System (HDFS), running on Dell PowerEdge servers. The centralized Hadoop cluster which lies at the heart of Nokia's infrastructure contains .5 PB of data. Nokia's data warehouses and marts continuously stream multi-structured data into a multi-tenant Hadoop environment, allowing the company's 60,000+ employees to access the data. Nokia runs hundreds of thousands of Scribe processes each day to efficiently move data from, for example, servers in Singapore to a Hadoop cluster in the UK data center. The company uses Sqoop to move data from HDFS to Oracle and/or Teradata. And Nokia serves data out of Hadoop through HBase.

KEY HIGHLIGHTS**TECHNOLOGIES IN USE**

- > Hadoop Platform: Cloudera Enterprise
- > Hadoop Components: HBase, HDFS, Scribe, Sqoop
- > Data Warehouse: Teradata, Oracle, MySQL
- > Servers: Dell PowerEdge R510, R710

BIG DATA SCALE

- > 100+ TB structured data
- > Multiple PB multi-structured data
- > Thousands of users in multi-tenant environment

Business Challenges before Hadoop

Prior to deploying Hadoop, numerous groups within Nokia were building application silos to accommodate their individual needs. It didn't take long before the company realized it could derive greater value from its collective data sets if these application silos could be integrated, enabling all globally captured data to be cross-referenced for a single, comprehensive version of truth.

"We were inventorying all of our applications and data sets," O'Connor noted. "Our goal was to end up with a single data asset."

Nokia wanted to understand at a holistic level how people interact with different applications around the world, which required them to implement an infrastructure that could support daily, terabyte-scale streams of unstructured data from phones in use, services, log files, and other sources. Leveraging this data also requires complex processing and computation to be consumable and useful for a variety of uses, like gleaning market insights, or understanding collective behaviors of groups; some aggregations of that data also need to be easily migrated to more structured environments in order to leverage specific analytic tools.

However, capturing petabyte-scale data using a relational database was cost prohibitive and would limit the types of data that could be ingested.

"We knew we'd break the bank trying to capture all this unstructured data in a structured environment," O'Connor said.

Because Hadoop uses industry standard hardware, the cost per terabyte of storage is, on average, 10x cheaper than a traditional relational data warehouse system. Additionally, unstructured data must be reformatted to fit into a relational schema before it can be loaded into the system. This requires an extra data processing step that slows ingestion, creates latency and eliminates elements of the data that could become important down the road.

Various groups of engineers at Nokia had already begun experimenting with Apache Hadoop, and a few were using Cloudera's Distribution Including Apache Hadoop (CDH). The benefits of Hadoop were clear — it offers reliable, cost-effective data storage and high performance parallel processing of multistructured data at petabyte scale — however, the rapidly evolving platform and tools designed to support and enable it are complex and can be difficult to deploy in production. CDH simplifies this process, bundling the most popular open source projects in the Apache Hadoop stack into a single, integrated package with steady and reliable releases.

After experimenting with CDH for several months, the company decided to standardize use of the Hadoop platform to be the cornerstone of its technology ecosystem. With limited Hadoop expertise in-house, Nokia turned to Cloudera to augment their internal engineering team with strategic technical support and global training services, giving them the confidence with expertise necessary to deploy a very large production Hadoop environment in a short timeframe.

“Hadoop is absolutely mission critical for Nokia ... We can now understand how people interact with the apps on their phones to view usage patterns across applications. We can ask things like, ‘Which feature did they go to after this one?’ and ‘Where did they seem to get lost?’ ... We wouldn’t have gotten our Big Data platform to where it is today without Cloudera’s platform, expertise and support. ”

AMY O’CONNOR
SENIOR DIRECTOR
OF ANALYTICS AT NOKIA

Impact

In 2011, Nokia put its central CDH cluster into production to serve as the company’s enterprise-wide information core. Cloudera supported the deployment from start to finish, ensuring the cluster was successfully integrated with other Hadoop clusters and relational technologies for maximum reliability and performance.

Nokia is now using Hadoop to push the analytics envelope, creating 3D digital maps that incorporate traffic models that understand speed categories, recent speeds on roads, historical traffic models, elevation, ongoing events, video streams of the world, and more.

“Hadoop is absolutely mission critical for Nokia. It would have been extremely difficult for us to build traffic models or any of our mapping content without the scalability and flexibility that Hadoop offers,” O’Connor explained. “We can now understand how people interact with the apps on their phones to view usage patterns across applications. We can ask things like, ‘Which feature did they go to after this one?’ and ‘Where did they seem to get lost?’ This has all been enabled by Hadoop, and we wouldn’t have gotten our Big Data platform to where it is today without Cloudera’s platform, expertise and support.”