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

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). 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.

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.

Industry
Telecommunications

Location
Global

Technologies in Use
• Hadoop Platform: CDH
• Hadoop Components: HBase, HDFS, Scribe, Sqoop
• Data Warehouse: Teradata, Oracle, MySQL

Business Applications Supported
• Geospatial application development
• Content/engagement optimization
• Network sessionization

Big Data Scale
• 100+ TB structured data
• Multiple PB multi-structured data
• Thousands of users in multi-tenant environment

Hadoop Impact
• Enables unprecedented scale and flexibility to build 3D digital maps of the globe
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

About Cloudera
Cloudera, the leader in Apache Hadoop-based software and services, enables data driven enterprises to easily derive business value from all their structured and unstructured data. Cloudera’s Distribution including Apache Hadoop (CDH), available to download for free at www.cloudera.com/downloads, is the most comprehensive, tested, stable and widely deployed distribution of Hadoop in commercial and non-commercial environments. For the fastest path to reliably using this completely open source technology in production for Big Data analytics and answering previously un-addressable big questions, organizations can subscribe to Cloudera Enterprise, comprised of Cloudera Manager software and Cloudera Support. Cloudera also offers training and certification on Apache technologies, as well as consulting services. As the top contributor to the Apache open source community and with tens of thousands of nodes under management across customers in financial services, government, telecommunications, media, web, advertising, retail, energy, bioinformatics, pharma/healthcare, university research, oil and gas and gaming, Cloudera’s depth of experience and commitment to sharing expertise are unrivaled.