SFR Implements an Enterprise Data Hub to Improve the Customer Journey

Overview

With revenues of more than US$13 billion in 2013, SFR is the second largest telecommunications operator in France. It serves more than 21 million mobile clients and delivers high-speed wired internet to 5.2 million households. In addition to consumers, SFR has 160,000 business, government, and community clients. Founded in 1987 and headquartered in Paris, France, the company's approximately 9,000 employees provide a wide portfolio of mobile, fixed, Internet, and television services.

SFR believes the proliferation of telecom networks and digital technologies allows people to enrich their lives through easy interactions with the world around them. As consumers and businesses transition to the digital world, SFR is optimizing their journey through the use of an integrated Big Data ecosystem that combines a Cloudera-powered enterprise data hub (EDH) with an enterprise data warehouse (EDW).

The Challenge

With today's tidal wave of mobile device activity—for work, play, financial, social, and many other interactions—SFR was challenged to create a mechanism capable of collecting and storing the huge magnitude of data generated by subscribers; volumes in excess of a billion events a day are not uncommon.

As Francois Nguyen, SFR's director of systems integration decisions and relationship marketing said, "Big Data is our way of 'Google-ising' the company. We wanted to have a shared and detailed view into the customer journey, so we could drastically improve it." To deliver on this goal, SFR would require the ability to provide employees across the organization with a 360-degree customer view that is accessible for real-time search, reporting, and analysis. This would ultimately help the company improve its quality of customer support, network operations, and would drive a better customer experience.

SFR's data warehouse has served the company well for ten years, containing data on products, device usage, invoices, contracts, price plans, and call detail records (CDR). But to truly understand the customer journey, SFR recognized it needed to bring in multi-structured data from new sources, such as customer behavior across SFR's many channels.

Key Highlights

Industries

- Telecommunications

Locations

- Headquartered in Paris, France
- Pan-European network

Business Application Supported

- Enterprise data hub for 360-degree customer view

Impact

- Enhanced customer experience
- Optimized IT ecosystem - data warehouse environment lasts 3x longer

Technologies in Use

- Hadoop Platform: Cloudera Enterprise, Data Hub Edition
- Hadoop Components: Apache Flume, Apache Sentry (incubating), Cloudera Impala, Cloudera Search, HDFS
- NoSQL Database: MongoDB
- BI Tools: IBM SPSS, Qlik, Microsoft BI
- ETL Tool: Syncsort DMXh
- Servers: HP

Big Data Scale

- 200 TB on 12 CDH nodes today, and rapidly growing
- Users across 9000+ employee base may access the centralized EDH

The company started researching up-and-coming technologies adopted by the likes of Facebook and Google, and identified Apache Hadoop as a potential platform that could complement the data warehouse by supporting large-scale data ingest, processing, and management of multi-structured data. Four commercial Hadoop vendors made it to SFR's short list for evaluation. Through the evaluation, three key criteria surfaced that led SFR to ultimately decide on Cloudera:

- Dedication to open source: Cloudera's Hadoop distribution (CDH) is 100% open source.
- Unique open source capabilities: Cloudera offers several benefits on top of the core CDHplatform, such as support for Cloudera Search and Impala, which were key differentiators.
- Enterprise-readiness: Data security is very important to SFR, which Cloudera addresses through tools like Apache Sentry (incubating), providing fine-grained authorization and role-based access control all through a single system.

The SFR team was also impressed by Impala's easy integration with a variety of BI tools, which would help to accelerate adoption of the new platform throughout the enterprise.

Solution

By complementing its data warehouse infrastructure with Cloudera Enterprise, Data HubEdition, SFR is delivering the 360-degree view that will help the company optimize the customer journey. Many of SFR's employees now have a self-service discovery environment enabling query and exploration of a single, centralized data store using familiar tools.

SFR's multi-tenant enterprise data hub processes data from numerous data sources, using Apache Flume for real-time data ingest, and stores the data in HDFS and Impala. SFR uses Syncsort DMXh to batch process massive files.

MongoDB is also interconnected with the Cloudera data hub, "exposing" the data processed in Cloudera using the MongoDB connector on both sides. The Cloudera platform is dedicated to collecting data and doing all the analytics, while MongoDB feeds SFR's front-office tools.

Users have the ability to access the data in Cloudera via a variety of tools that support their diverse needs. Impala is leveraged by SQL and BI users who have direct interaction with the EDH through the tools they're comfortable with: IBM SPSS (accommodating up to 500 users), Qlik, and Microsoft BI. Cloudera Search will be the day-to-day tool that makes it easy for employees to explore the data in Hadoop without running SQL queries or MapReduce commands.

Impact: A Better Customer Experience

With its enterprise data hub, SFR is empowering employees across the country to operate based on a centralized, real-time customer view that spans many devices and data sources.

For the first time, SFR has the capacity to ingest, store, and analyze log data, from which the company may glean previously hidden customer insights by combining that data with other data sets.

Impact: An Optimized IT Ecosystem

SFR's integration of its Cloudera EDH with the existing data warehouse has resulted in a best-of-breed Big Data ecosystem that's delivering maximum operational efficiency. By

offloading large-scale data ingest, processing, and exploration of multi-structured data sets from the data warehouse to Cloudera, the data warehouse now has more capacity to focus on what it was built for: high-performance analytics and access to the company's "first-class data."

This also elongates the shelf life of the data warehouse. "Instead of upgrading our environment every three years, the system will deliver optimal performance for eight or nine years now," said Nguyen.

Cloudera Manager and Sentry play key roles in streamlining the operational efficiency of SFR's data hub, offering robust system management and security.

> "We wanted the benefits of open source plus the enterprise-ready capabilities, hence we chose Cloudera"
>
> Francois Nguyen, Director of Systems Inegration Decisions and Relationship Marketing, SFR

About Cloudera

Cloudera is revolutionizing enterprise data management by offering the first unified platform for big data, an enterprise data hub built on Apache Hadoop. Cloudera offers enterprises one place to store, process and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. Only Cloudera offers everything needed on a journey to an enterprise data hub, including software for business critical data challenges such as storage, access, management, analysis, security and search. As the leading educator of Hadoop professionals, Cloudera has trained over 40,000 individuals worldwide. Over 800 partners and a seasoned professional services team help deliver greater time to value. Finally, only Cloudera provides proactive and predictive support to run an enterprise data hub with confidence. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production. www.cloudera.com.

cloudera.com

1-888-789-1488 or 1-650-362-0488
Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA

cloudera®