



PagesJaunes Uncovers New Insights with Cloudera

Overview

Solocal Group (LOCAL) is a leader in digital local communication in Europe, with revenues of €936 million in 2014. It employs 4,800 people in France, Spain, Austria, and United Kingdom including 2,300 sales advisers to support its 700,000 local and national advertising customers. Solocal Group operates three complementary business lines, content and service provision, media, and advertising representation, through 17 brands (including Solocal Group, PagesJaunes, Mappy, ComprendreChoisir, and A Vendre A Louer).

PagesJaunes, a Solocal Group company, works with small and micro-businesses to help them develop and grow. And to create successful new services for its users and advertisers, the company needs a constant pulse of the marketplace. To more effectively explore data and quickly uncover new insights, the company is moving from siloed data management environments to an [enterprise data hub](#) (EDH) powered by Cloudera.

This new open, scalable, and flexible data discovery and analytics environment combines unstructured and structured data from a variety of internal and third party sources to help staff develop a deeper understanding of consumer behavior, market needs, and business risk.

By creating a single platform for storing and analyzing data, and making it available to users across the company, the organization can gain new business intelligence to improve every aspect of operations—sales, marketing, CRM, human resources, finance, and supply chain.

The company's first goal was to more precisely target its sales efforts. By optimizing ETL processes, sales data is now refreshed in minutes, instead of weekly. As a result, PagesJaunes staff can more rapidly respond to changing sales conditions and identify which products to sell to whom and when.

Key Highlights

Industry

- Media and Entertainment

Locations

- Headquarters: Paris, France

Business Applications Supported

- Enterprise data hub empowering BI infrastructure
- Functional data marts across Sales, Marketing, CRM, Human Resources, and Supply Chain

Impact

- Reduces time to refresh sales data from weekly to every 5 minutes
- Enables Sales to identify customers most likely to purchase
- Inspires new collaboration between business leaders and BI

Technologies in Use

- Hadoop Platform: Cloudera Enterprise, Data Hub Edition
- Hadoop Components: Apache Hive, Apache Spark, Cloudera Manager, Impala, MapReduce
- ETL Tool: Syncsort
- BI and Analytics Tool: Zoomdata
- Data Warehouse: IBM Netezza

Big Data Scale

- 80 TB analyzed every month

The Challenge

“BI [business intelligence] is not a tool; it’s not a report; it’s not a database,” said Abed Ajraou, head of Business Intelligence at Solocal Group. “BI is a complete environment that creates intelligence with data.”

It was with this vision that Mr. Ajraou began his work to transform a rigid legacy environment into a flexible platform would allow staff to exploit and quickly analyze any information—unstructured and structured, from both internal and third party sources.

The company’s legacy environment, built on data silos, could not support either the volume or variety of data required.

“It’s complicated to maintain a data warehouse with 50 terabytes (TB) of data,” he said. “When you need to add a column or table, it becomes even more difficult. We also have to consider unstructured data from logs, blogs, and social networks. Traditional databases are designed for structured data.”

Additionally, Mr. Ajraou sought to give executives and business staff easy access to the latest data. But to do so, it needed to reduce extract, transform, load (ETL) cycle delays. “With traditional BI, we had to stop the database, collect the data, do the transformations and aggregations and complete the reports,” said Mr. Ajraou. “The world is going real-time and we need to operate our business at the same pace.”

The Solution

“Our conviction was that [\[Apache\] Hadoop](#) could solve most of our BI pains,” said Mr. Ajraou. “We’re using an enterprise data hub from Cloudera as our extended data warehouse. We found Cloudera to be very easy to use relative to the alternatives, and it was easy to build the clusters using [Cloudera Manager](#). We also felt that the Cloudera platform offered better integration with other products to facilitate administration and use of Hadoop, including simplifying access and security.”

Cloudera serves as the company’s data discovery and analytics environment, combining structured data—such as website log, mobile log, finance, supply chain, and CRM data—with unstructured data from social networks, content management systems, and videos.

The team uses [Syncsort](#) ETL to transform data into functional data marts for each division. Data marts that may require modification over time run on [Hive](#). [Impala](#) is used to allow business users to directly query the data in Hadoop using SQL queries.

“I was very impressed the first time I saw Impala,” Mr. Ajraou said. “Impala gives users a simple way to access the data, such as the latest sales data, while helping us ensure good performance. We use Impala in a number of different business cases. For example, in one case, we are using Impala with a Java website to determine the digital visibility of a company in France.”

The team will soon use [Apache Spark Streaming](#) to incorporate streaming and real-time data into the data marts. “Spark Streaming will help us move more toward real-time so we can better track KPIs [key performance indicators] and more quickly evaluate new business opportunities,” said Mr. Ajraou.

Business users can explore and visualize the data stored in the Cloudera EDH using

Zoomdata's dashboards. "Zoomdata connects to every stack in Hadoop—including Impala and Spark—so the dashboards are updated as the data changes," said Mr. Ajraou.

For Mr. Ajraou, the Cloudera enterprise data hub will enable the organization to answer questions that it couldn't before.

"Finance teams are able to predict which transactions are at risk," said Mr. Ajraou.

"Marketing can see which new product has the greatest chance to succeed. Human Resources are able to correlate data from LinkedIn to forecast if a key employee is planning to leave. There are so many possibilities that we couldn't even imagine before."

Additionally, the flexibility of the platform enables BI staff to continually test new theories with the data.

"We don't know what the data is going to tell us," said Mr. Ajraou. "With this architecture, and leading tools, such as Impala, we can build an application very quickly to test if it works, and then if it works, we can iterate."

According to Mr. Ajraou, this work has inspired greater collaboration between BI and business leaders.

"In traditional BI, the questions come from the business," said Mr. Ajraou. "But now, we can give answers to questions that the business didn't even know to ask. This is helping them see what's possible and they're now coming back to us with new ideas."

Impact: Refreshing Sales Data Every Five Minutes

PagesJaunes today uses the Cloudera enterprise data hub to provide its executive team with more up-to-the minute sales figures. The data is refreshed every five minutes as opposed to weekly, which enables executives to take immediate action to keep sales targets on course. The impact has been so great that the solution was recognized for its tremendous value and innovation during an internal competition.

"From a BI perspective, the data collection and calculations for this use case were relatively simple, but the impact was big," said Mr. Ajraou. "Every person on the executive team uses Hadoop every day, without knowing it."

Impact: Identifying Top Prospects to Increase Sales Success

Which clients are most likely to buy which products? How can we better target our offers?

For sales reps, having the "inside scoop" on their clients can help them more quickly and effectively close sales. By bringing together external data about company revenue, web traffic metrics, and sales data, PagesJaunes can precisely target which customers in which industries to approach with which products.

"We're able to coalesce the right data with the right person at the right time so we can tell sales reps where to spend their time," said Mr. Ajraou. "It's very important to be data-driven. It's amazing what you can find in the data—there are a lot of new ideas and business areas to explore."

"We're able to coalesce the right data with the right person at the right time so we can tell sales reps where to spend their time."

Mr. Abed Ajraou, Head of Business Intelligence,
Solocal Group

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 1,400 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

© 2015 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.

cloudera-casestudy-pagesjaunes-101

