SOFTWAREMILL

How retailers can keep up with consumers

LPP S.A. Case Study

CLIENT PROFILE

LPP S.A. is one of the biggest retailers in Europe with a total of more than 1700 stores located in 20 countries. Their goal is to become a company whose brands (Reserved, Cropp, House, Mohito and Sinsay) are globally recognisable. The ever-changing customer preferences and behaviours, as well as ubiquitous digitalisation are both triggering transformation in retail trade, therefore LPP S.A.'s strategy includes the development of an online store chain.

PROJECT OVERVIEW

INDUSTRY
FASHION RETAIL

CHALLENGE
PROCESSING 600 EVENTS PER SECOND IN REAL TIME

PROJECT DURATION
3 MONTHS

GOAL OF THE PROJECT
IMPLEMENTING ENGINE FOR REAL-TIME DATA PROCESSING

TECHNOLOGIES
Apache Kafka, Kafka Connect, Kafka Streams, Apache Beam, Google Cloud Platform (Dataflow, BigQuery, CloudSQL), Snowplow, Kubernetes, Strimzi

BENEFITS

- TAILORED PRODUCT RECOMMENDATIONS IN REAL-TIME
- REAL-TIME INSIGHTS INTO CUSTOMERS ONLINE BEHAVIOUR
- AGGREGATED ORDERS ENABLE REAL-TIME MARKETING & SALES DECISIONS

BACKGROUND

Running online stores for five brands, targeting different groups of customers, causes an immense volume of data arriving at high velocity. LPP is aware of the challenges of the eCommerce market caused by various data sources and siloses. The company approached SoftwareMill to help them streamline data pipelines and in result, provide customers with a unique experience. A tailored interaction, that is the key to customers' heart.

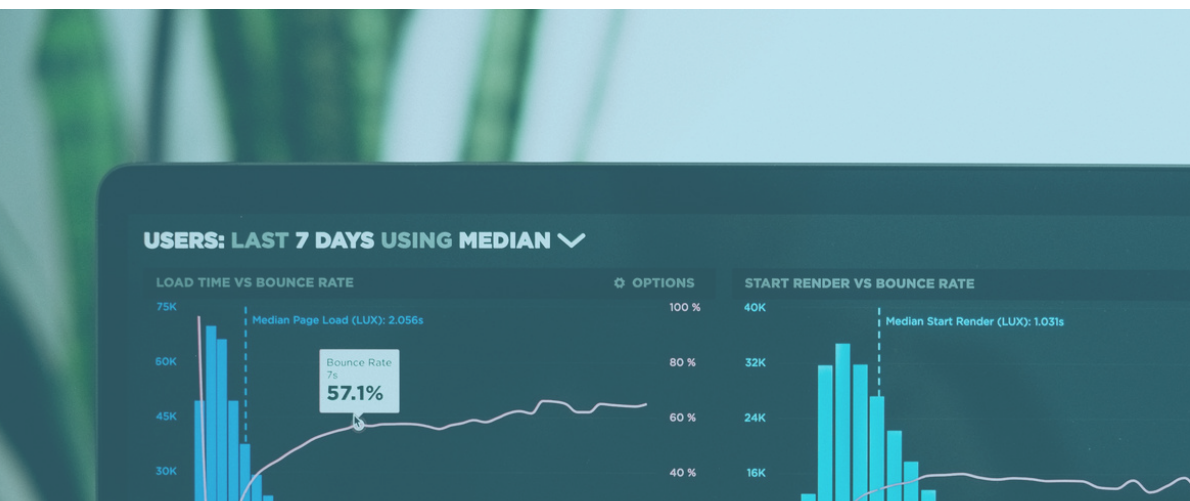
The goal of the project was to enable tailored product recommendations in real-time, by tracking online customers behaviour with no delay. Such mechanism, of metrics and graphs aggregating orders and customers' behaviour in real-time, can significantly increase sales by enabling the right and timely marketing decisions.

Companies that are unable to access data struggle to effectively measure marketing ROI. One of the benchmarks and pioneers among successful retailers is Amazon, claiming to generate 35% revenue with its recommendations [after McKinsey report]. Delivering personalisation to individuals at scale requires a combination of the right data and the right technology. Relying on Google Analytics to track customers online behaviour is not enough, as we have to deal with up to 72h of delay, while the data can be processed. Additionally, online data have to be catered for veracity to extract the real value out of it.

CHALLENGES

LPP S.A. collects online customers' data from multiple sources. The two biggest inputs are: orders, stored in databases, and customers' online behaviour recorded with Google Analytics. On an ordinary day there are up to 600 events per second that need to be processed. Big data can play a pivotal role in generating more sales for any eCommerce business. If a customer has spent time browsing products on the site, retailer has to understand what they're looking for and respond with the right recommendations before they move on to another site. Working on a BigData project is about to tame the 5 Vs: volume, variety, veracity, velocity and last but definitely not least: value. The value extracted from data, allows retail businesses reap the rewards of better customer experiences and bigger profits.

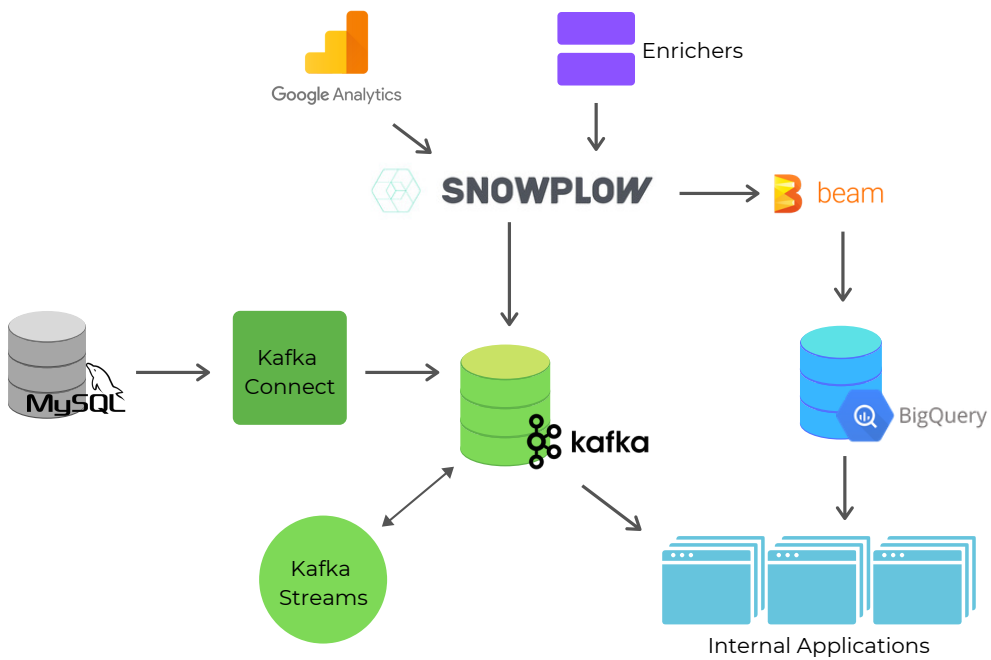
The existing solution was based on batch processing moving data between different data siloses. Although this approach allowed to analyse customers' online behaviour and to display personalised recommendations as well as to prepare personalized email campaigns the day after the visit, it missed the opportunity to interact with the customer during their current shopping session. Processing data, arriving at high velocity, required us to use battle-proven tools for stream processing. Massive volumes demand not just huge storage capabilities, but also the ability to scale during peak hours. Due to the nature of the online retail business our solutions have to work 24 hours a day and therefore must be deployed on a system featuring self-healing, as well as load balancing. Finally it is required to monitor the infrastructure health and performance and capture metrics exposed by the delivered applications to derive business value in real-time.



SOLUTION

The first milestone was to move data from the databases, transform it and push into Google BigQuery for further analysis. It was achieved with Kafka Connect and the Debezium connector to pull customers' and ordering data from relational databases into Kafka. The next step was to transform the data using Kafka Streams and push it to BigQuery via the WePay Kafka Connect connector. Google Analytics data was collected, processed or enriched and finally pushed into BigQuery with Snowplow, an event data collection platform. The infrastructure was built up on Google Cloud Platform leveraging Google Kubernetes Engine and the underlying Compute Engines. The Kafka cluster was deployed by Kubernetes with the Strimzi operator. The monitoring and alerting infrastructure has been setup with the Prometheus operator, Grafana and AlertManager.

The goal of the second milestone was to prepare data for a recommendation engine. Based on an order history and buyers' paths a built in-house application propose products to be displayed on the currently visited page. Furthermore, graphs and reports display current state of orders, as well as cumulative summaries of: total orders, abandoned baskets, most-viewed products, etc. These reports no longer have to run expensive queries against Google BigQuery, but are fed in real-time via stream processing applications build up on Kafka Streams and Apache Beam.

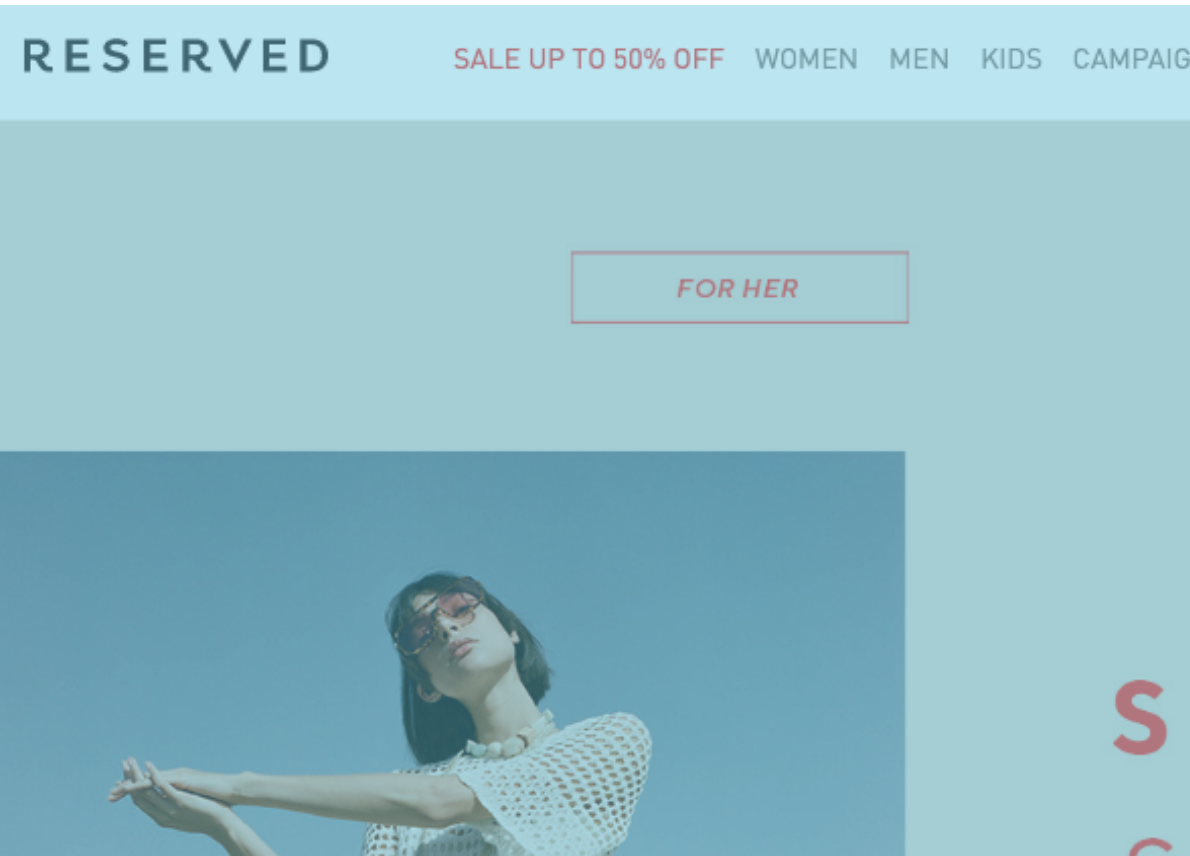


RESULTS

During three months we transformed the existing batch-based process into a data streaming platform built up on mature and popular open source tools. Since our customer relied already on Google Cloud Platform, we took advantage of the available services, like BigQuery, Dataflow, CloudSQL, Memorystore and many more.

Seven streaming applications have been delivered as well as devops scripts setting up Kubernetes and all necessary tools. As a result, our customer was able to attach custom dashboards to display the sales volume as well as feed their recommendation engine in a cost effective way without latency.

Thanks to the deployed solutions, LPP S.A. can implement an all-encompassing real-time marketing strategy from square one - from real-time product recommendation on-site to multichannel customer-centric approach, also known as Customer360, in the nearest future.



"They [SoftwareMill] stuck to the three-month timeline, doing everything to deliver a ready solution. (...) thanks to the infrastructure, we're able to compete with market leaders."

Szymon Chojnacki, Big Data Architect at LPP S.A

GOT AN IDEA?
WE'LL MAKE IT HAPPEN

contact@softwaremill.com
www.softwaremill.com

We are SoftwareMill, a Poland(EU)-based consulting & custom software development company, delivering services remotely, worldwide for 10 years. Being experts in Scala (Akka, Play, Spark), Java, Kotlin we specialize in blockchain, distributed, big data systems, machine learning, IoT, and data analytics.

We believe that focus on quality, self-improvement and a true engineering approach can result in systems that do their job, bring value to clients, help them scale and grow.