CONSIST PROJECT REFERENCE

Green light for data – creating a big data lake in just 90 days

A big player in logistics – the story in brief

The logistics company: a holding company with three divisions, 15 subsidiary companies and 29 billion metric tons of freight moved per year in more than 23,000 transport units – in short, with a huge amount of infrastructure data. A group that is now increasingly supporting the big data ambitions of its business units, and helping them to establish data science throughout the company. Because enormous data volumes offer enormous potential for further development of individual tasks, business processes, business units or even entire companies.

Therefore, various data tools and technologies were already in place in the different divisions. As a result three different business intelligence platforms existed. This caused a high degree of complexity, high integration costs and long implementation intervals. There was therefore a wish for consolidation towards a central big data platform. That's why, throughout the group, big data – and consequently also data science – became a key topic for the associated competence center.

In order to get all group units involved in this process by mutual agreement, a pilot project was initiated, to impressively demonstrate the advantages of a single data platform. In only 90 days, a roughly-described conceptual toolset should be transformed into a working pilot application with real data. Consist supported this process both architecturally and „hands on" with the practical implementation. Together with the agile group team, they developed the pilot „Freight_Wagon_Monitoring_ETA (Estimated Time of Arrival)".

The task

- Establishing a group-wide concept for dealing with big data challenges – defining the methodology, technology and infrastructure
- Setting up an on-premises big data lake and piloting an initial use case on this basis
- Determining the path to an enterprise platform (enterprise data lake)

The challenge

- The very different structure of data lakes requires important decisions when selecting the technical platform, the nature of the distributions, and the number of nodes for scaling the platform.
- The clarification of these issues in the context of mutual dependencies, their quick resolution in the group environment, and the immediate setting up of a pilot required an experienced team.
- There was also severe time pressure due to upcoming first real trials and test runs in parts of the group, as well as increased demand for such systems.

The solution with Consist

- Development of a concept based on the industry-standard Lambda architecture, to serve as the basis of a group-specific enterprise data lake for batch and real-time processing:
- Definition of a suitable framework, to be configured and evaluated in the context of a test version
- Implementation of an ingestion pipeline for processing the data

Technologies used

- Ingestion pipeline via Logstash, Kafka and Flume (using Consist interceptors): preparation of the relevant data and final, well-partitioned conversion into HDFS.
- Aggregation and analysis of the data using Hive
- Transferring the analysis results into a relational database via Sqoop
- Coordinating the timing of the individual processes using Oozie

Particular strengths of Consist

- Holistic know-how in big data technologies, coupled with professional rail expertise
- Efficient, technologically-sound implementation of the project on time and within budget
- Together with the agile team of the group competence center, an appropriate technical and professional environment for the data lake could be created very quickly.

Customer benefits

- High level of acceptance for a central group-wide big data platform
- Data lake with pilot use case, which can serve as the nucleus for the product platform, and at the same time be developed further in a production environment
- Avoiding the higher time requirements and costs of further partial solutions, thanks to consolidation into a central big data platform

CONSIST
Business Information Technology

IT that works.

Your contact:
info@consist.de
+49 (0) 431 / 3993-672
Consist Software Solutions GmbH
Falklandstr. 1-3, 24159 Kiel, Germany
www.consist.de