

CASE STUDY

DATA INTELLIGENCE— SCANNERS DEVELOPED TO GATHER METADATA FROM VARIOUS TOOLS

Client Background

Our client, a privately held computer software company, develops and sells enterprise software to international clients. The company creates and deploys software solutions that reduce cost, mitigate risk, and improve service delivery throughout the IT lifecycle. The client provides comprehensive solutions including:

- Metadata and content management
- Systems solutions—a wide variety of products including performance management, workload automation, application management, and service quality and control.

softserve

Business Challenge

During the client's product evolution, the amount of data used and provided grew dramatically. As a result, the data became cumbersome and unmanageable—it was time consuming to get it into a usable format.

- Data lakes contain a huge variety of information. The type, creation date, owner, initial size, etc. could be determined. However, the data was so complex that analyzing separate chunks while ignoring the relationship between it did not give the complete picture.
- Even if the data was quickly found and analyzed and then put in the proper context, it was still probable the data was inconsistent—caused by its mutability. With time some dependencies between the data elements disappeared making the data useless.

Project Description

The client collaborated with SoftServe to develop scanners to gather metadata from various tools—Hive, HDFS, Falcon, Flume, Sqoop—used for transformations, migrations, and storing data in Apache Hadoop.

The scanners could gather metadata using REST or Java API and then transform the data into a specific format for the metadata repository. Also, the scanners created links between metadata items in the metadata repository scanned from different tools. All scanned, imported, and linked data was used for further analytical processes and lineage creation.

Value Delivered

As a result of the collaboration, the client received the following value:

- **X-platform.** Scanners developed by SoftServe gathered metadata from Apache Hadoop framework components enabling compatibility with the most popular Hadoop distributions—Cloudera, Hortonworks, IBM BigInsights, and Talend Studio.
- **Modular.** Each scanner is a standalone application and can be used separately or in combination with other tools. This approach allowed our client to select the exact tool combination for their needs and guaranteed time and money savings.
- **Flexible.** Scanners produced structured data output (XML) so various programming languages could be used for additional data processing—analyzing, reconstructing relationships, and visualizing the metadata structure and lineage.

ABOUT US

SoftServe is a digital authority that advises and provides at the cutting-edge of technology. We reveal, transform, accelerate, and optimize the way enterprises and software companies do business. With expertise across healthcare, retail, media, financial services, software, and more, we implement end-to-end solutions to deliver the innovation, quality, and speed that our clients' users expect.

SoftServe delivers open innovation—from generating compelling new ideas, to developing and implementing transformational products and services.

Our work and client experience are built on a foundation of empathetic, human-focused design that ensures continuity from concept to release.

We empower enterprises and software companies to (re)identify differentiation, accelerate solution development, and vigorously compete in today's digital economy—No matter where you are in your journey.

Visit our [website](#), [blog](#), [Facebook](#), [Twitter](#), and [LinkedIn](#) pages.

NORTH AMERICAN HQ

Tel: +1 866 687 3588 (USA)

Tel: +1 647 948 7638 (Canada)

EUROPEAN HQ

Tel: +44 (0) 800 302 9436

info@softserveinc.com

www.softserveinc.com

softserve