# Extracting SAP Data to Snowflake

Continuus worked with a manufacturing organization to extract SAP and Salesforce data to Snowflake. The organization aims to leverage Snowflake's capabilities for advanced analytics and reporting across multiple data systems. The focus of the engagement was on tool performance, methodology, and benefits derived from the process.

## AT A GLANCE

### ABOUT THE CLIENT

Mid-sized multinational manufacturing company in Wisconsin, using SAP ERP and Salesforce to manage operations and generate vast transactional and operational data.

### PROBLEM

The company needed to centralize its SAP and Salesforce data in Snowflake for real-time analytics, reporting, and improved business intelligence. They also sought to optimize storage costs, reduce data management complexity, and prepare for AI-enabled scenarios.

### SOLUTION

Continuus implemented a data extraction process using Informatica Mass Ingestion to extract and transform SAP data, and a Snowflake Connector for loading it into Snowflake. The project involved mapping SAP data to Snowflake, cleaning and aggregating the data, and automating the extraction and loading process for real-time updates.

### OUTCOME

The project improved reporting speed, with reports now generated in minutes instead of hours. It reduced data storage costs by 30% and enhanced cross-departmental collaboration by centralizing data. The company now has a scalable, efficient data warehouse in Snowflake, supporting real-time analytics and future AI initiatives.

## PROBLEM

This manufacturing organization utilizes SAP ERP for managing its operations, which generates vast amounts of transactional and operational data. Business intelligence teams were looking for ways to integrate this data into a centralized data warehouse for analytics, reporting, and business intelligence purposes and, eventually, AI-enabled scenarios.

Objectives for this project included:
- **Centralized Data Repository:** Create a unified data warehouse in Snowflake for all analytics and reporting needs.
- **Real-time Analysis:** Enable near real-time data analysis to support decision-making.
- **Cost-Effective Solution:** Optimize storage costs and reduce the complexity of data management.

## SOLUTION

### STEP 1: TOOL SELECTION

After an analysis of a variety of tools that could help the company accomplish its objectives, the engagement team decided on:
- **Informatica Mass Ingestion:** To extract and transform SAP data.
- **Snowflake Connector:** To load data into the Snowflake environment.

### STEP 2: DATA EXTRACTION

1. **Connection Setup:** Established a connection to the SAP system using Informatica.
2. **Data Mapping:** Mapped the data fields from SAP to corresponding fields in Snowflake. This involved determining which tables and columns were necessary for analytical purposes.
3. **Incremental Extraction:** Implemented incremental extraction methods to ensure only new or updated records are pulled into Snowflake, reducing load times and resource consumption.

## SOLUTION CONT'D

### STEP 3: DATA TRANFORMATION

1. **Data Cleaning:** Addressed discrepancies in data formats and eliminated duplicates.
2. **Data Aggregation:** Prepared data to meet the analytical needs by summarizing and aggregating as necessary.
3. **Data Enrichment:** Joined SAP data with other internal datasets to enhance analysis capabilities.

### STEP 4: DATA LOADING INTO SNOWFLAKE

1. **Staging:** Initially loaded the data into a staging area within Snowflake for validation.
2. **Final Load:** Once validated, the data was moved to the final tables designed for reporting and analysis.
3. **Automation:** Set up scheduled jobs to automate the extraction and loading process, ensuring up-to-date information in Snowflake.

## CHALLENGES FACED

- **Data Volume:** Managing large datasets resulted in extended loading times.
- **Data Quality Issues:** Inconsistent data formats from various SAP modules required extensive data cleaning.
- **Team Skills:** Need for additional training on Snowflake and ETL tools to optimize the extraction process.

## OUTCOME

### BENEFITS ACHIEVED

- **Enhanced Reporting:** Stakeholders can now generate real-time reports, improving decision-making capabilities.
- **Cost Efficiency:** By leveraging Snowflake's capacity to scale, the organization optimized storage costs compared to traditional data warehouses.
- **Increased Collaboration:** Centralized data storage improved cross-departmental collaboration, allowing teams to share insights more effectively.

### METRICS

- **Reporting Speed:** Reports that previously took hours to generate now take minutes.
- **Cost Reduction:** Data storage costs decreased by 30% compared to previous infrastructure.

### NEXT STEPS

- **Continuous Monitoring:** Set up performance monitoring to optimize ETL processes and improve data quality.
- **Explore Advanced Features:** Investigate Snowflake's capabilities for predictive analytics and machine learning integrations.
- **Feedback Loop:** Regularly gather feedback from stakeholders to identify additional data requirements or adjustments needed in existing processes.

### CONCLUSION

The extraction of SAP data to Snowflake has proven to be a transformative project for the organization, aligning with its objectives of data centralization and enhanced analytics capabilities. Future initiatives may include exploring advanced analytics tools within Snowflake and integrating additional data sources for a comprehensive data ecosystem.

Get in touch to learn more.

**sales@continuus-technologies.com**