BARC

BARC Spotlight

SAP Data Integration – Turning Challenges into Opportunities

Authors: Timm Grosser, Thomas Zeutschler

Publication: August 23, 2024

Provided by

Qlik®

Index

SAP Data Integration – Turning Challenges into Opportunities

Introduction

In today's rapidly evolving digital landscape, businesses are increasingly recognising the urgent need to modernise their data platforms to maintain their competitive advantage. Specifically, maximising the value of SAP data is essential for businesses looking to optimise supply chains, modernise finance functions and empower employees with better analytics and insights.

"Seamless Data Pipelines should be a top priority on your agenda."

This phrase encapsulates the essence of modern data management. A seamless integration process is a prerequisite for long-term competitiveness, enabling businesses to be data-driven, data-connected, and data-intelligent. Achieving this requires a structured and well-managed approach, offering significant returns in business agility, value generation, and reduced costs. The necessary technologies for this transformation are already available, making it an achievable goal. We should modernize our SAP integration and it's worth it... then next chapter shows why.

Opportunities in SAP Data Integration

Reducing Complexity and Increasing Efficiency

The traditional setup of SAP data integration often involves multiple, redundant interfaces. Traditionally, SAP data integration setups often suffer from having numerous redundant interfaces, which complicate the overall system architecture and elevate maintenance costs. By shifting from multiple, fragmented SAP access interfaces to a more streamlined model, organizations can significantly simplify their data integration landscape. Modernization is achieved through abstraction. 1:1 connections require a lot of effort, as a connector for each case needs to be developed and maintained. However, by using connectors that be can fed with different parameters depending on the use case, several cases can be handled with one connector.

As a result, there are fewer interfaces and changes and maintenance can be done more efficiently – thus saving man power and cutting costs. This approach encapsulates complexity within the SAP systems, making them more robust and adaptable to changes. Furthermore, eliminating these redundant interfaces not only reduces the overhead associated with managing them but also leads to substantial cost savings. The streamlined setup allows for more efficient and manageable data integration processes, improving overall system efficiency.

Improving Data Availability and Accessibility

The adoption of advanced data platforms that offer high-performance, low-cost data processing and analysis at scale can greatly enhance data availability. This is particularly advantageous compared to relying on multiple, specialized tools for different integration tasks such as ETL, data preparation, enterprise application integration, and data orchestration. Tools like SAP Data Intelligence and SAP Data Services facilitate unified data management, enabling seamless handling of data pipelines across both SAP and non-SAP systems.

Enhancing API Access and System Integration

Open interfaces (APIs) play a vital role in providing access to specific data points, thus facilitating seamless system integration and enabling comprehensive business process orchestration and automation. By introducing additional APIs that encapsulate SAP logic under a user-friendly interface, businesses can significantly broaden access to SAP data and metadata. This enhanced accessibility allows for better context and meaning to be attached to the data, enabling both humans and machines to understand and utilize it more effectively. This is also a prerequisite for AI and ML. The result is improved business agility, support for a broader range of use cases, increased trust in the data, and the establishment of a more robust data foundation.

Recommendations for Modernizing SAP Data Integration

To modernize SAP data integration, businesses should focus on standardizing their data provisioning and utilization processes, ensuring a seamless and efficient flow of information across systems. This involves conducting a detailed assessment of the existing data integration landscape to identify inefficiencies and potential areas for improvement.

Additionally, by offloading data to analytical systems like data lakes or data warehouses, organizations can reduce costs and improve processing speeds and create the basis for implementing further analytical use cases by combining SAP and non-SAP data, which can be made available openly to other tools.. Implementing continuous data integration, where feasible, further enhances real-time data availability, supporting more dynamic and responsive business operations. The following recommendations outline key strategies for achieving these goals and leveraging advanced integration capabilities to streamline and enhance SAP data management.

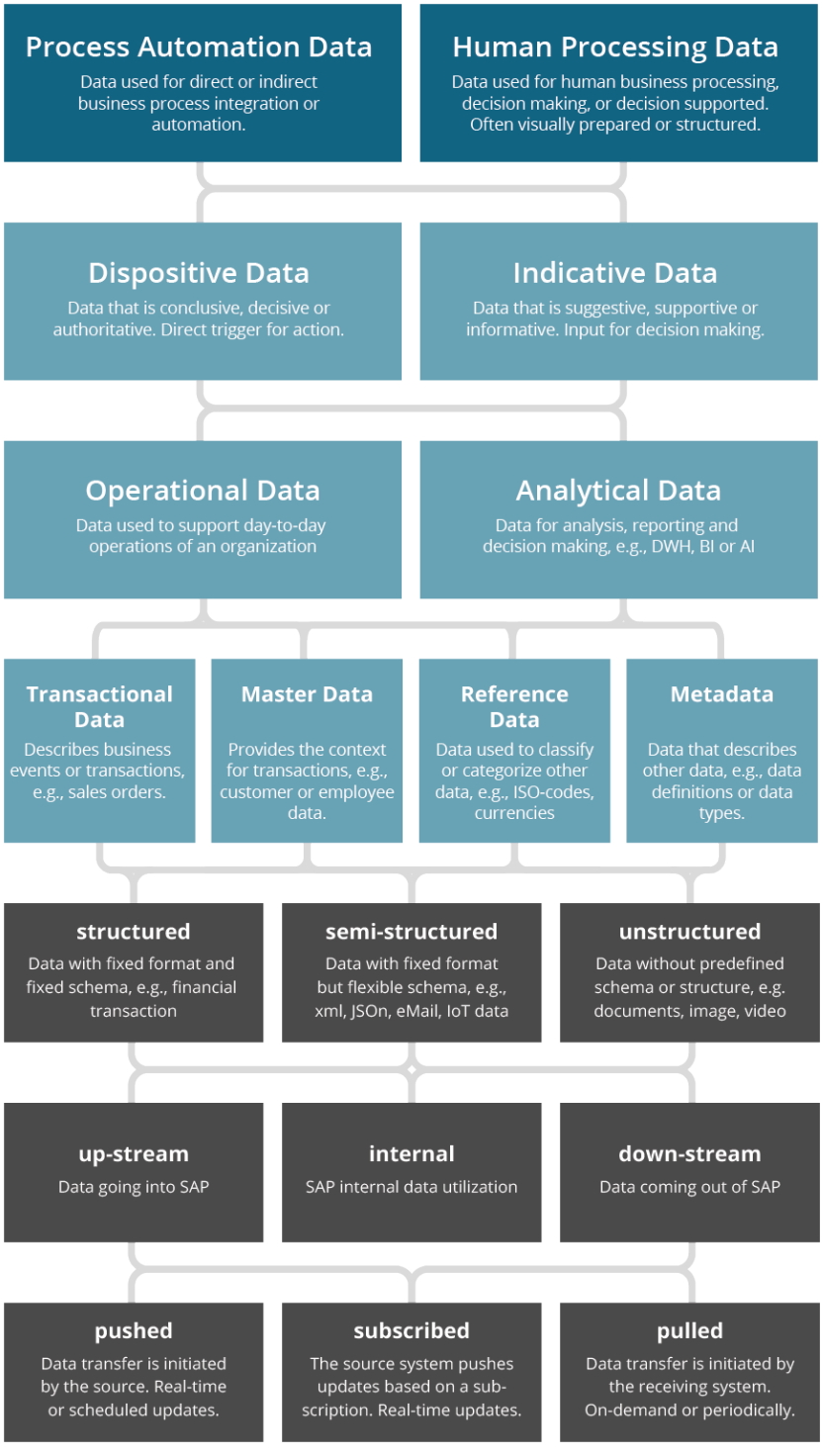
To fully leverage SAP data integration, consider the following recommendations:

- **Standardize Data Provisioning and Utilization**: Create a clear, standardized connection between data provisioning and utilization to minimize complexity and cost.
- **Assess and Modernize Existing Infrastructure**: Conduct a thorough assessment of the current SAP data integration landscape and identify areas for improvement.
- **Offload Data to Specialized Systems**: Moving data to specialized management systems can reduce costs and improve speed. Consider using data lakes or data warehouses for downstream applications.
- **Implement Continuous Data Integration**: Where possible, shift from batch processing to continuous data integration for real-time data availability.
- **Streamline and Eliminate Redundant Interfaces**: Identify and eliminate redundant interfaces to reduce system strain and costs.

- Leverage Modern Data Integration Capabilities: Utilize advanced features of BI systems and dedicated data integration platforms to enhance data lineage, security, and governance. Also consider modern non-SAP tools in the evaluation. Depending on the intended use, these can help to significantly improve SAP data integration and make it more efficient.

Data Type Classifications Relevant For SAP Data Integration

All directly supported by SAP through various data integration technologies and methodologies.



Modern business processes are increasingly distributed, automated and executed in real-time. In addition, business process expertise is too valuable to be spent on the execution of processes. **The future goals of data management are to feed process automation** and to provide information to humans to **shape and manage business processes**, ideally not run them.

The goal of modern data management is to **generate direct input for decision making**. For machines as for humans, the more decisive data is, e.g., a simple yes or no, the more valuable and usable it is. Modern statistics and AI helps to transform indicative data into dispositive data. Nevertheless, indicative data is crucial for context and to enable automation.

The more modern data analysis and AI is utilized for automated or semi-automated decision support and data analysis, the more the logical and physical **distinction between operational and analytical data becomes obsolete** and obstructive.

The rising need for data and business **process automation requires proper Metadata**. Therefore, data integration technologies need to be able to actively manage and use metadata. Metadata leverages to **find, understand, share and protect** data and make it accessible for machines and humans.

Structured data is still the most valuable type of data for analytical purposes and AI. **But the advances in classic and generative AI** will likely make the accessibility and utilization of unstructured data (text, images, video) even more relevant. Modern data integration technologies are going to support both at scale.

Data needs to be forwarded to as many business processes as possible to leverage its full potential and value. A trend is to simply **down stream all relevant SAP data on a granular level**. Low-cost data storage (e.g. Data Lake) is used increasingly for data offloading. Requires highly efficient and scalable data integration technologies.

Robust connectors & SAP Change Data Capture (CDC) enables constant pushing of data into DWH & BI systems. Offering always-up-to-date analytical data, reduces strain on SAP Systems and better governance.

SAP S/4 Data Integration Big Picture

With focus on for DWH, BI, analytics and AI use cases.

1	2	3	4

A. Data Within SAP Systems

Using "Sales Data" related tables and views as an example

SAP Core Data Services (CDS)

1	2	3	4	
x	x			**Consumption View** prefiltered, ready to use business views C_SalesOrderItemQuery
x	x	x	x	**Composite View** bringing related artefacts together I_SalesOrderItemCube
x	x	x	x	**Basic View** often individual business artefacts I_SalesOrder I_SalesOrderItem
x	x	x	x	**Internal Tables**, technical, most detailed, sometimes needed VBAK VBKD VEDA

Here you **talk to a well-designed service for data access.** Depending on requirements, various kinds of predefined views with different levels of details are available through **Virtual Data Models** or **VDMs. This is the way to go!**

I_ = **Interface** View, C_ = **Consumption** View

HANA Database

Physical HANA Tables, not intended for any use

VBAK VBKD VEDA

Here you would **talk to a database.** Technically possible, neither supported nor recommended by SAP. **Don't do this!**

B. SAP Data Access Concepts and Technologies

Data access to SAP systems is "off" by default; what is a good idea to secure and govern your most valuable data. So, you need to explicitly activate and manage data access on a per data object and user in SAP, what is cumbersome at best. For more agility, data can be off-loaded and authorizations moved to access layer.

1	2	3	4	
	x	x	x	**Connectors & Extractors** Delivering larger chunks of data or mass data for DWH and BI purposes and real-time data. High throughput. BW Extractor API SAP and non-SAP integration tools and connectors
x	x	x	x	**Queries (SQL)** Allowing direct queries tables and views in SAP. **Most flexible, optimal for interactive end-user** access through frontends. GraphQL SAP Query SAP Open SQL SAP and non-SAP frontend tools
x	x		x	**Reports & Function Calls** Data is requested by a function call that delivers a **predefined dataset,** often combined with some **specific parameters** for the function or report, e.g. product name, period etc. Very robust, **easy to use.** SAP Report BAPI/RFC ODP /oData SAP and non-SAP integration tools

Overall, Data Extraction from SAP is not an easy task. So many options, settings and technologies. **Modern Data extraction technologies hide the complexity** from users.

SQL is the most flexible way to extract data from SAP. But using **SQL requires a deep understanding** of tables and views. Users could get overwhelmed. GraphQL is even more complicated.

A big advantage of SAP is **the number of predefined reports** available for almost every business questions (6000+ in S4). The big challenge: How to know what is relevant or valuable to users. Future of ODP interface using RFC?

C. Use Cases

Some typical use cases

1 Business Process Automation 2 User Self Service Analysis 3 DWH, Data Vault 4 Analytics, AI

Final Thoughts

Effective SAP data integration is not just a technical necessity but a strategic imperative for organizations aiming to maximize the value derived from their data assets. As businesses increasingly rely on data to drive decision-making and operational processes, the complexity of managing SAP data can no longer be a bottleneck. By rethinking and optimizing data integration strategies, companies can achieve several critical benefits, including significant cost savings, enhanced data accessibility and quality, and improved overall operational efficiency.

The journey towards an modernized SAP data integration landscape begins with a clear understanding of current systems and processes. Organizations must assess their existing infrastructure, identify redundant interfaces, and streamline their data flow to eliminate inefficiencies. This process involves not just technical changes but also a shift in organizational mindset towards valuing data as a key asset. By leveraging modern data integration tools and technologies, such as data lakes and advanced BI systems, businesses can offload data processing tasks, thus reducing the burden on their core SAP systems. This not only cuts costs but also frees up resources, allowing for greater innovation and agility. Moreover, adopting continuous data integration practices ensures that businesses have real-time access to critical data, empowering them to respond swiftly to market changes and emerging opportunities. In conclusion, a well-structured and efficient SAP data integration strategy is essential for organizations to stay competitive, innovate continuously, and make data-driven decisions that propel growth and success.

Terms, Facts & Insights

SAP Data Integration Tools and Interfaces

SAP offers various technologies for data integration, each suited for different purposes.
Some of the key tools include:

- **SAP Data Intelligence**: Primarily used for building and orchestrating data pipelines within SAP Business Technology Platform (BTP).
- **SAP Data Services**: Manages ETL processes for SAP and non-SAP systems.
- **SAP HANA Smart Data Integration**: Handles data pipelines within SAP HANA.
- **SAP Integration Suite**: Integrates on-premise and cloud applications in BTP.
- **SAP Process Integration / Orchestration**: Supports service-oriented architectures.
- **SAP System Landscape Transformation Replication Server**: Facilitates real-time data replication from SAP or third-party systems into SAP HANA.

SAP also provides various interfaces such as RFC, BAPI, IDoc, and CDS Views, which can be accessed through protocols like oData, SOAP, REST, and HTTP/S. The choice of interface depends on the specific use case, spanning batch to real-time scenarios, and the performance and complexity of required transformations. This diversity of tools and interfaces is a significant driver of complexity in SAP Data Integration, alongside the enormous data model complexity in the SAP source systems themselves.

Glossary

Data Lakehouse
A data lakehouse is an architectural concept that combines a data warehouse with a data lake. Raw data is stored permanently and cost-effectively in the data lake. Only structured data that is required for certain analyses is loaded into the data warehouse. The data lakehouse concept is now recommended by all cloud hyperscalers with different names.

Data Warehouse
A Data Warehouse (DWH), or Enterprise Data Warehouse (EDW), is a central data repository for reporting and analyzing integrated data from multiple sources. It stores current and historical data to support BI applications like reporting and OLAP. DWHs may include Data Marts and extract data from operational systems such as marketing and sales. While DWHs can be used as data sources for other systems, they should not be misused as data hubs, as varying service levels can complicate adjustments.

SAP data cataloging & lineage
Context information about data is needed to find, understand, analyze and protect data for both humans and machines. Catalogs help to inventory disparate SAP data and make it much more intuitive to navigate. It can provide useful information that helps to get a better understanding about data provenence, applied logic and rules, usage of the data, dependencies and also quality of information. This information help to streamline different use cases from migration (e.g. understand table dependencies), bi or analytical use cases (find right data or queries to build analytical applications) to governance use cases (e.g. monitor data usage, observe policies, understand data quality). If a catalog encompasses SAP and non-SAP systems metadata it can be a good way to integrate both worlds for better decision making.

SAP Connector
A connector provides performant and easy access to a source system and ideally encapsulate all technical complexity. Why to mention it here: Connectors can differ in various ways like performance of data extraction, supported integration styles and extraction methods (e.g. read/write, ACID compliance, support of bulk load or mass data extraction), profiling capabilities, sample extraction, depth of access to source systems data objects, and more. To have a reliable, performant connectors that encapsulated SAP data and provides robustness against changes in source system as well as supports needed integration styles and depth of data access is crucial. Finally user need trusted data, in the right format, in the right time at scale.

ETL/ELT
ETL is a classic data warehouse loading process in which the data is retrieved from the source system (extract), checked and converted in a second step (transform) and loaded into the data warehouse in a final step. In contrast to the ETL process, the data is loaded directly into the data warehouse after extraction. The transformation takes place within the DBMS. Particularly relevant in the context of big data and data lakes as well as data warehouse automation (DWA).

Data Virtualiziation
Data virtualization enables users to access, query, and manipulate data from multiple sources as a unified database without moving the data. It offers advantages such as unified access, real-time integration, and cost savings by reducing physical data storage. Challenges include potential performance issues and complexity in integrating diverse data sources.

A use case example is a financial services company using data virtualization to combine customer data from CRM, transaction systems, and external credit scores, providing a real-time, 360-degree view of customer profiles for better decision-making and personalized service.

Data Preparation
Data preparation is the data preprocessing for data discovery and data science, typically performed by business users or data scientists in a decentralized manner. It primarly provides low-code/no-code interfaces with lightweight or well-guided transformation and mapping functionality. It can also support data quality remediation use cases to prepare a data set for furhter usage. The goal is to provide data for answering specific questions. This distinguishes data preparation from the generic processing in data warehousing (ETL).

Data Streaming
A data stream is a continuous flow of data records of the same type in a fixed chronological order, potentially with varying data input rates. Using data streaming, sensor data, for example, can be stored long-term as a time series at any desired resolution, enabling comprehensive analyses based on synchronized timestamps and comparable events.

Data Ingestion
Data ingestion is the process of collecting and importing data from various sources into a storage system, such as a data warehouse or data lake, that resides typically in the cloud. It involves extracting, transforming, and loading (ETL) data to ensure it is accessible for analysis and decision-making. Data ingestion can be done in real-time or batch processing, and it supports multiple data formats and sources, including databases, files, and streaming data, to provide a unified view of information. It differs from ETL as it is dedicated into loading data into a target system (mainly in the cloud) and as it do not focus on applying complex transformation logic to mass-data.

Data Orchestration
Data pipeline orchestration is the automated process of managing and scheduling data pipelines across multiple systems and tasks. It involves coordinating the sequential execution of data ingestion, transformation, and processing stages to ensure data flows smoothly from source to destination. Orchestration tools provide monitoring, error handling, and dependency management, allowing data engineers to automate repetitive tasks and optimize performance. This ensures data pipelines run efficiently and reliably, supporting timely data delivery for analytics and business intelligence.

Delta Loading
Refers to an essential capability of SAP BW and S4/HANA systems, that supports incremental data loads, which only contain new or changed records since the last load. Prerequisite for data warehousing at scale from within (BW, Datasphere) and/or outside the SAP Ecosystem (BI-Systems, DWH, Data Lakehouse), as full-loads, especially daily, may take much to long (days) and create high system strain from such prolonged workloads.

CDC (Change Data Capture)
Modern pattern for continues data integration, aka “streaming”. Changes, e.g., new postings, are automatically captured by the SAP system/database and immediately forwarded to subsequent processes within or outside an SAP system. This could be for operational or analytical purposes (to trigger some process) or for continuous data extraction into BI or DWH systems. This approach helps to reduce peak system strain on SAP side caused by larger delta or full loads. In addition, some receiving data management systems are able present changes from the underlying SAP system in near-real-time, data is always in sync.

About Boon Solutions

Boon Solutions is committed to delivering innovative data solutions tailored to address the unique challenges of your business. With expertise in AI, machine learning, business intelligence, data integration, analytics and automation, we help organisations unlock the full potential of their data assets.

As a Qlik Elite Partner with a proven track record of delivering bespoke strategies aligned with business objectives, we work to empower organisations to enhance operational efficiency, drive innovation, and achieve sustainable growth.

Discover how we can transform your data into actionable insights. Visit boon.com.au

Contact info

Boon Solutions Pty Ltd
Level 6
12 St Georges Terrace
Perth WA 6000
Australia

Phone. +61 08 6102 3206
Email info@boon.com.au

boon.com.au
boon.com.au/sap