Discover Data Integration Aspects of the SAP Data Hub

Axel Schuller

SAP Data Intelligence Product Management
Disclaimer

The information in this presentation is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. Except for your obligation to protect confidential information, this presentation is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has no obligation to pursue any course of business outlined in this presentation or any related document, or to develop or release any functionality mentioned therein.

This presentation, or any related document and SAP’s strategy and possible future developments, products and or platforms directions and functionality are all subject to change and may be changed by SAP at any time for any reason without notice. The information in this presentation is not a commitment, promise or legal obligation to deliver any material, code or functionality. This presentation is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This presentation is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this presentation, except if such damages were caused by SAP’s intentional or gross negligence.

All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.
Agenda

Data Management in the Intelligent Enterprise
- Use Cases

SAP Data Hub and SAP Data Intelligence

The approach for EIM and beyond:
- Native Data Integration
- Choreographed workflows
- Metadata Management
Challenges in Today's Enterprise Landscapes

The result is landscape complexity

- Applications
- Data Warehouses
- Databases
- Cloud Datastores
- Data Marts
- Third-Party Data
- Data profiling
- Data cleansing
- ETL
- Video processing
- Geospatial processing
- Time series
- Text analytics
- Machine learning
- Speech recognition
- Data cataloging
- Event Stream processing
- Data ingestion
- Data replication
- Data masking
- Data quality
- Streaming analytics
- Graph processing
- Image processing
- Metadata management

Customer Experience

Manufacturing & Supply Chain

Digital Core

People Engagement

Network & Spend Management
SAP Strategy for Diverse Landscapes
Outlook for data integration and processing portfolio

- SAP LT Replication Server
- Data Quality Management (DQaaS)
- CPI – Data Services (CPI-DS)
- Agile Data Preparation (ADP)
- Smart Data Integration (SDI)
- Smart Data Quality (SDQ)
- SAP Data Services (DS)

SAP Data Hub
SAP Data Intelligence
Data Integration Technologies by Purpose

ETL / Data Quality
- On Premise Sources
- Cloud Sources
- SAP Data Services
- SAP Information Steward
- Batch CPI-DS Batch SAP IBP

ETL / Replication
- ABAP sources
- On Premise/Cloud Sources
- Real-time Replication
- SLT ABAP Targets
- Real-time Replication
- Batch SAP HANA SD/SDQ
- Batch Virtualization

Governance, Pipeline, Orchestration
- Intelligent Enterprises
  (SAP Leonardo IoT, SAP Hybris, SAP Concur)
- BW
- LT
- DS
- Open Source

Complex Event Processing
- SAP HANA
  - Streaming Analytics
  - Incoming Streams

Virtualization
- Remote Systems
  - Table/View
  - Query Data Remotely
  - SAP HANA SDA Virtual Tables
Guiding Principles for EIM Portfolio

INTEGRATE
Integration of all tools to enable new use cases and scenarios.

EMBED
Bring critical capabilities into a single innovation platform.

REUSE
Protect existing landscapes by leveraging investments customer made.
SAP Data Hub

Key capabilities

SAP Applications

- SAP S/4HANA
- SAP NetWeaver + DMS Addon
- SAP BW
- SAP S/4HANA Cloud
- SAP C/4HANA
- SAP Concur
- BW/4 HANA
- SAP Fieldglass
- SAP BusinessObjects
- SAP HANA
- SAP Analytics Cloud

Distributed & External Data Systems

- Standard Connectors (open & native protocols)
- 3rd Party Connectors
- SAP Cloud Platform Connectors
- SAP Open Connectors
- SCI for process integration
- SAP API Business Hub
- REST APIs
- Public Clouds
- Databases
- Hadoop / HDFS
- Streaming (e.g. IoT)

Data Governance

- Data Discovery
- Data Profiling
- Metadata Cataloging

Data Orchestration & Monitoring

- Connection Management
- Workflows
- Scheduling

Data Pipelining & Processing

- Data ingestion
- Data Processing
- Data Enrichment

*This is the current state of planning and may be changed by SAP at any time without notice.
Use **SAP Data Hub** to reimagine your business processes using …

**Business Application Transformation**
Streamline innovation initiatives around Business Applications, supporting enterprise transformation programs

- Customer Risk Intelligence with S/4 HANA Cloud
- Timesheet Analytics with Fieldglass
- […]

**IoT Ingestion and Orchestration**
Transform IoT event streams into enterprise-ready data, and derive actionable insights

- Predictive maintenance
- Product as-a-service
- Intelligent logistics
- Smart manufacturing
- […]

**Data Warehousing**
Build a multifaceted data warehouse, across diverse and distributed data assets

- Consumer 360 view
- Marketing campaign effectiveness
- Renewable energy production simulation
- […]

**Data Science and Machine Learning**
Streamline data science and machine learning, from modeling and development to operations, across all enterprise data assets

- Smart energy management
- Fraud detection
- Customer churn prediction
- […]

**Information/ Data Governance**
Understand, govern and secure your data to deliver trusted insights

- Proactive Information/ Data Governance
- Digital Data Stewardship
- […]

© 2019 SAP SE or an SAP affiliate company. All rights reserved. | PUBLIC
What is SAP Data Intelligence?

Create powerful data pipelines to leverage your data projects and to orchestrate the data processing.

Harness the advanced machine learning content to accelerate and scale and automate your Data Science projects.

Manage metadata across a diverse data landscape and create a metadata repository.

SAP Data Intelligence is a mPaaS comprehensive solution to deliver data-driven innovation and intelligence across the enterprise, unifying scalable enterprise AI and intelligent information management.

- Access & connect data
- Govern & discover data
- Prepare & label data
- Build scalable & flexible data pipelines
- Deploy & integrate intelligent applications
- Monitor & orchestrate the lifecycle
SAP Data Hub: Orchestration and Monitoring

Data Workflows

- Central management of all system connections
- **Data Workflows**: Definition of cross-system processes, e.g.
  - Trigger Execution of SAP BW process chains
  - Transfer data from from SAP BW
  - Execute remote SAP Data Services jobs
  - Submit Spark jobs to Hadoop clusters
- **Scheduling** of Data Workflows
- Extensive **Monitoring**
SAP Data Hub: Pipelining & Processing

Pipeline Modeler

- **Data Pipelines**: Model flow-based applications
  - Operators (independent computation units)
  - Data (messages) flows between operators
  - Containers constitute the operators’ execution environments

- Over 250 pre-defined Operators in different categories such as Connectivity, Processing, Data Quality, Computer Vision, Machine Learning

- Development of custom Operators e.g. in Python, JavaScript, Go and R

- Content Lifecycle Management
Complementary solutions: SAP Data Services and SAP Data Hub

SAP Data Services
Moving application data from transactional sources to data warehouses, including data quality processes and built-in Data Integration transformations

Key use cases
- BI / Traditional Data Warehousing, Data Migration, Data Quality

Characteristics
- ETL in a standalone heterogeneous landscape
- Centralized, on premise & server-based infrastructure
- Relational data focused
- Advanced data transformations & processing (e.g. Join, SQL, DQ…)

SAP Data Hub
A pipeline-driven data integration, operations and governance solution for disparate kinds of data (structured, unstructured, streaming, cloud etc.), supporting both integration and processing in a distributed fashion

Key use cases
- Data science & Machine Learning, Big Data Warehousing, IoT, Data Tiering, Data Network

Characteristics
- Support multiple data ingestion methods
- Pipelining and orchestration of data processes in complex landscapes
- Cloud & on-premise deployment, distributed data processing & serverless computing via Kubernetes
- Big data focused (tables/views, object storages, any data formats), for both data at rest and data in motion
- Complex data refinery & processing (e.g. ML, Predictive, Image, custom code)
SAP Data Services and SAP Data Hub
Interoperability: Example for combined scenario

1. Ingest large volumes of data (e.g. distance, pace, heartrate, location…) from machine sensors by using MQTT/Kafka operator (SAP Data Hub)

2. Refine data according to purpose and store it in data stores (SAP Data Hub)

3. Acquire additional relevant structured data (e.g. customers, sales, behavioral, demographic) into data stores by remotely orchestrating DS jobs (leveraging existing SAP Data Services investments)

4. Apply machine learning algorithms (e.g. classification, clustering, identifying outliers, etc.) on the data to discover new insights about user characteristics (SAP Data Hub)

5. Invoke process chain to ingest the results into SAP BW/4HANA for further data analysis and reporting (SAP Data Hub)
Complementary solutions: SAP Cloud Platform Integration and SAP Data Hub

**SAP Cloud Platform Integration**
Provides service on cloud platform that facilitates the integration of business processes and data across on-premise and cloud applications

**SAP Data Hub**
A pipeline-driven data integration, operations and governance solution for disparate kinds of data (structured, unstructured, streaming, cloud etc.), supporting both integration and processing in a distributed fashion

**Key use cases and characteristics**
- Integration between transactional (cloud and on-premise) applications
- A2A / B2B, B2G / Master Data Synchronization
- API focused (synch., async., business events)*
- Message-based processing (mapping, routing, monitoring)
- Transactional integrity (reliable messaging and error handling)
- Pre-packaged integration content for SAP/Non-SAP scenarios

* Also table access possible by CPI (like CPI-DS) and API level access possible by SAP Data Hub, but not key focus

**Interoperability**

**Data processing and integration style**

- IoT Data Stream
- ML/Predictive
- Unstructured Data

SAP Cloud Platform Integration

<table>
<thead>
<tr>
<th>SAP Cloud Apps</th>
<th>Third-Party Cloud Apps</th>
<th>Business Partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP Applications</td>
<td>Third-Party Systems</td>
<td>Public Authorities</td>
</tr>
</tbody>
</table>

SAP Data Hub

<table>
<thead>
<tr>
<th>Enterprise Data</th>
<th>Data Warehouse</th>
<th>Data Lake/Object Storages</th>
</tr>
</thead>
</table>

© 2019 SAP SE or an SAP affiliate company. All rights reserved. | PUBLIC
Direction SAP Data Services & SAP CPI-DS with SAP Data Hub

- Run SAP DS/CPI-DS with SAP Data Hub
- Reuse connectivity
- Share infrastructure

- Orchestration and landscape monitoring for all data flows:

Monitor Mockup