ASUG 84185
Overview of SAP Data Hub
Michael Eacrett, SAP Labs
May, 2019
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

Overview
- Motivation
- Strategy & Key Features
- Architecture

Use Cases

Demo

Additional information
Enterprises are sitting on a data treasure trove which they can’t unlock

**Data Transparency**
Difficulties in managing and orchestrating data across distributed data landscapes

74% of enterprises say their data landscape is so complex that it limits agility

**Data Innovation**
Complexity in synthesizing data from SAP and non-SAP landscapes for machine learning training, and IoT use cases

86% of enterprises claim that they are not getting the most out of their data

**Data Compliance**
Enforcing a multitude of corporate and regulatory data policies is becoming a burden and risk for enterprise IT

4% of annual turnover can be fined by the EU for violating GDPR regulations
Data deluge: Distributed landscapes, diverse data, manual siloed operations

Manual, single step systems and processes for moving data are unable to deliver actionable information instantly and automatically.
SAP Strategy – Deliver the Intelligent Enterprise

THE INTELLIGENT ENTERPRISE features 3 KEY COMPONENTS:

1. Intelligent Suite
2. Digital Platform
3. Intelligent Technologies
Digital Platform: Unlock data-driven intelligence and innovation

Unified data management to capture real-time value from different types of data

Best-in-class digital platform for new app development, extensions, and integration

Next-generation data management expands SAP HANA in-memory database to address structured and unstructured data use cases and external data

SAP HANA powers SAP applications as the foundation of high-performance data warehousing and analytics

SAP Data Hub provides data orchestration and metadata management across heterogeneous data sources

Platform for extending the business processes of the Intelligent Suite and enabling new innovations

Delivering deep data and process integrations through APIs and microservices

Marketplace for ecosystem to build new innovations leveraging APIs and business services
SAP Data Hub – Unified Data Integration for the Intelligent Enterprise

Data-driven Applications

- IoT
- Machine Learning
- Analytics / BW
- ...

SAP BW

- ABAP Integration*

Cloud Data Integration*

Workflow

SAP HANA Integration

SAP Data Hub

- Metadata Management
- Orchestration & Monitoring
- Pipelining & Processing
- Ingestion & Integration
- Distributed Runtime

SAP Applications

- SAP BW
- SAP S/4HANA
- SAP Fieldglass
- GIGYA
- SAP Ariba
- SAP Concur
- Business Apps
- Business Services
- BW Process Chains
- Data Services Jobs
- HANA Flowgraphs

SAP HANA

Distributed / External Data Systems

- Streaming (e.g. IoT)
- Cloud Storages
- Hadoop / HDFS
- REST APIs
- Databases
- Public Clouds
- 3rd party apps
- SAP API Business Hub*
- SCI for process integration*
- SAP Event Bus*

Connectors (open & native protocols)

*This is the current state of planning and may be changed by SAP at any time without notice.
SAP Data Hub

Product highlights

SAP Data Hub is a multifaceted solution that provides data integration, data orchestration, data refinement and data governance across the diverse and distributed data landscape.

**Data Integration**
- Landscape management
- Enterprise application integration
- Multiple methods of data ingestion (streaming, batch, replication, etc.)

**Data Orchestration**
- Workflow scheduling and monitoring
- Remote task execution
- Data transfer between systems

**Data Refinement**
- Data pipeline modeling
- Complex data processing (e.g. multimedia, machine learning, etc.)
- Distributed query processing for big data

**Data Governance**
- Metadata management and cataloging
- Visual data lineage and Impact analysis
- Profiling, discovery, access and security

---

© 2019 SAP SE or an SAP affiliate company. All rights reserved. | PUBLIC
Architecture overview of SAP Data Hub

**Connected Systems**

- SAP S/4HANA
- SAP BW/4HANA
- SAP Data Services
- SAP LT Replication Server
- SAP HANA
- Databases
- SAP Cloud Applications (API-driven)
- Open connectivity for third-party & open source

**SAP Data Hub System Management (based on SAP HANA)**

- Multi-Tenancy
- User & Access Management
- Content Lifecycle Management
- Cluster Management
- Diagnostics

**Data Storages**

- Cloud / On-Premise
- Cloud Stores
  - AWS S3, GCP GCS, Azure ADL & WASB
- Hadoop
  - HDFS (optional)
- SAP Data Hub Adapter
- SAP Vora Spark Extensions

**Distributed Runtime**

- Relational App Server
- Streaming Tables SQL

**Pipelines & Workflows**

- Scripting (JS, Python)
- Built-in Connectors
- Custom Operators
- Flow-Based Applications

**Metadata & Applications**

- Scheduling
- Metadata Catalog
- Profiling & Discovery Connectivity
- Application Services

**SAP Vora Engines**

- SAP Data Hub Adapter
- SAP Vora Spark Extensions

**Open connectivity**

- for third-party & open source
Use Cases
Data science & machine learning data management
Prepare data from all source to increase effectiveness of AI/Machine learning algorithms

- Integrate variety of data sources with open data landscape management
- Process ML models leveraging many SAP and non-SAP engines within the same tool
- Quickly and safely operationalize ML outcomes back into enterprise processes
- Dramatically decrease the time needed to collect, refine, and orchestrate data
Data science & machine learning data management

Example – Machine Learning Model Training Pipeline
IoT data ingestion and orchestration
Integrate and process disparate data from messaging systems and high-volume cloud storages

- Tackle the challenge of integrating and analyzing vast quantities of raw data and events from disparate semi-structured sources with low-level semantics and no business context

- Solve the point-to-point challenge of distributed heterogeneous environments spanning messaging systems, cloud storages, SAP data management solutions, and enterprise apps

- Event-driven pipelines scaling to executions of many pipelines in parallel, at any time
IoT data ingestion and orchestration
Example – IoT Ingestion Pipeline
Intelligent data warehouse
Rapidly integrate and leverage new data sources

• Acquire new data sources with previously siloed data from traditional data warehouses, data marts, enterprise applications, and Big Data stores

• Combine all types of sources including structured and unstructured data, and enable a large variety of processing on them

• Seamlessly process large data sets across highly distributed landscapes and close to the data source, moving only high-value data
Governance and data cataloging
Understand and secure your data

- Crawl through data stores to gather valuable metadata and store it in a centralized information catalog
- Profile source data to gain a deeper understanding of the data to create meaningful data pipelines
- Move to centralized data access and control for all orchestration, data refinement, scheduling, and monitoring
Demo
Further information

Public SAP Web sites
SAP Community: https://www.sap.com/developer/topics/data-hub.html

SAP training and certification opportunities
openSAP – Freedom of Data with SAP Data Hub: https://open.sap.com/courses/hub1
SAP Data Hub video on SAP HANA Academy: https://www.youtube.com/saphanaacademy
SAP Data Hub Channel – https://www.youtube.com/channel/UCLMsNLj0GF0nEw3iixeSEMw/featured

ASUG+Sapphire:
Sapphire show floor: SAP Data Hub booth and “pop up” demo booth

ASUG SAP Data Hub Sessions:
ASUG84186  Deep Dive Into SAP Data Hub: The What, When, Why, and How
ASUG83241  Roadmap: SAP Data Hub
ASUG84047  SAP Data Hub: Turning Disparate Data into Valuable Assets for Digital Innovation
ASUG82702  SAP Data Hub: The Journey to a Productive Use Case at BASF
ASUG84223  KAESER KOMPRESSOREN: "Smart Air" Strategy Execution
ASUG83238  SAP's Strategy for Big Data and Enterprise Information Management
ASUG83511  Accelerate Digital Transformation at General Mills
Take the Session Survey.

We want to hear from you! Be sure to complete the session evaluation on the SAPPHIRE NOW and ASUG Annual Conference mobile app.
Presentation Materials

Access the slides from 2019 ASUG Annual Conference here:
http://info.asug.com/2019-ac-slides
Q&A

For questions after this session, contact us at DataHub@sap.com.
Let’s Be Social.

Stay connected. Share your SAP experiences anytime, anywhere.
Join the ASUG conversation on social media: @ASUG365
#ASUG
Your Turn!

Contact information:

Michael Eacrett
SAP Data Hub Product Management
Michael.Eacrett@sap.com