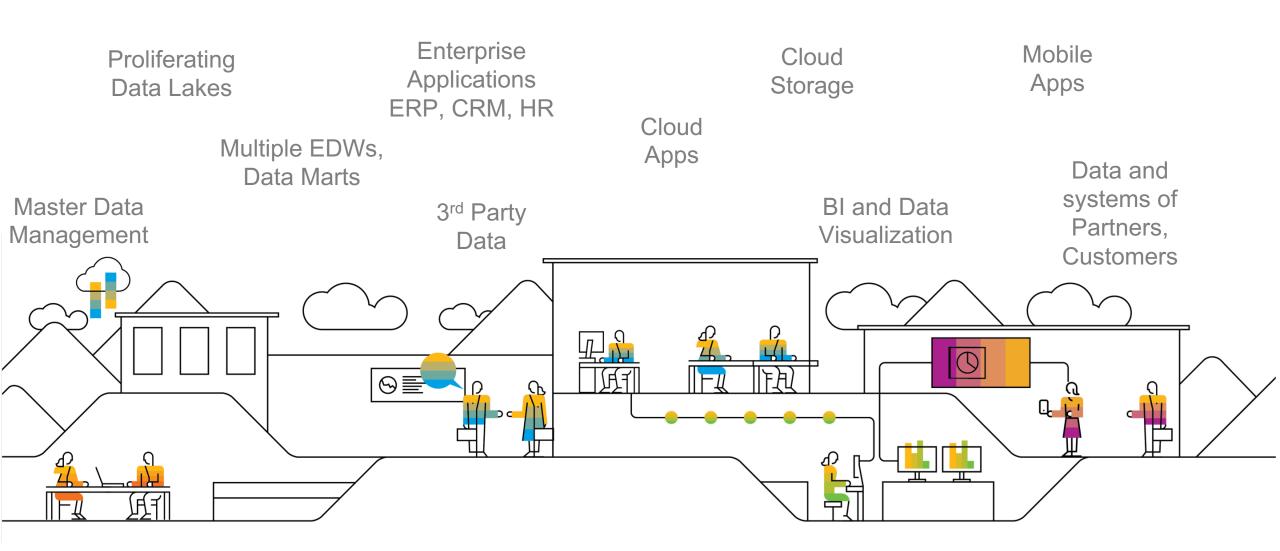
Freedom of Data in a Diverse Landscape





Enterprise data landscapes are growing increasingly complex



Modern Landscapes

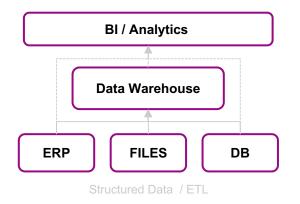
"Big" Data is transforming customer landscapes

From centralized, relational, on premise DWH approaches...

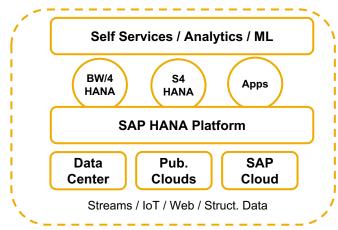
...to modern distributed (cloud) Data Platforms

Characteristics

App.Server & DB OS & Hardware ETL Driven Structured Data







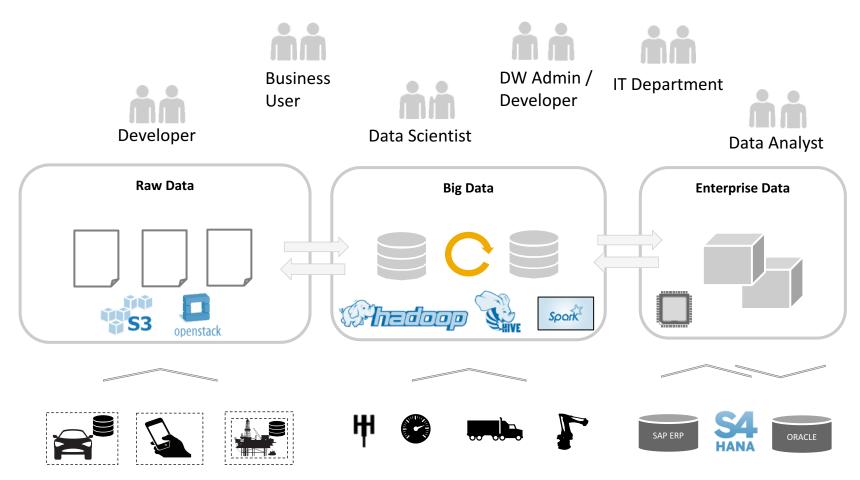
Characteristics

Serverless Computing Containerized Software Distributed Data Data Driven Any Format

Key Drivers:

- Challenges for traditional architectures due to multi-structures, large data volumes, landscape scale outs
- Growing Cloud / Data Lake / IoT Adoption

Modern Landscapes – Example



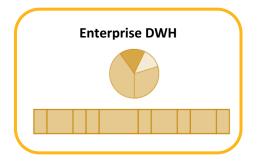
Key Challenges & Questions

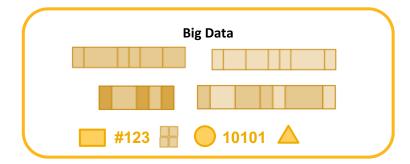
- Which technology for which data?
- How can collaboration between different groups with individual requirements be facilitated?
- How can huge and evergrowing data volumes be handled?
- Rapidly changing landscape –
 where is the next opportunity?

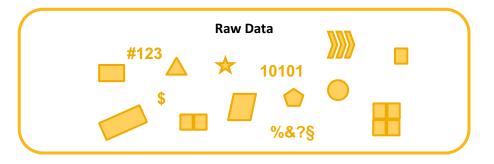
Another set of silos are not the answer!!!

Modern Landscapes

Data Perspective as biggest challenge







Analytical Modeling

Joins / Unions De-normalization

Matching / Duplicate Check

Meta Data Extraction and Generation

Enrichment

Cleansing

(Re-) Formatting

Anonymization / Masking

Parsing

Filtering

Search

Data Validation

Data Stream

Refining Insights out of diverse Data

- Structuring the unstructured.
- Handling large volumes of data efficiently
- ETL or DWH are not the answer.
 - Data formats & granularity
 - Data Streams and flexible structures
 - Deploy logic at the data, not data at the logic
- Need to ensure integrated analytics across the enterprise.
- Early insights on each level of data needed
- Automation across layers needed

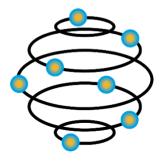
Challenges

Overcoming silos, complexity to drive better operations and insight



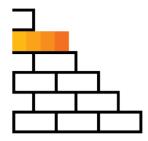
Missing Link

between Big Data and Enterprise Data



Lack of Enterprise Readiness

of Big Data solutions



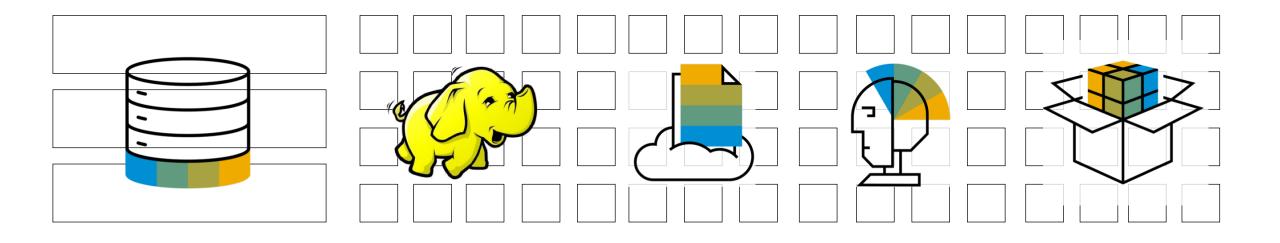
Limited Tools = High Effort

to productize complex data scenarios across data landscape

New challenges require new technologies

Distributed systems in a distributed landscape

Existing Systems



Cloud Storage

(i.e. AWS S3)

Machine

Learning

(Python, Spark, Tensorflow)

Containers

(Kubernetes, Docker)

Hadoop/Spark

Benefits

Accelerate and scale your data projects

Build agile, data-driven applications

Gain fast, relevant business insights

Enterprise visibility and governance





Communicates with pilots

Monitors airport activities

Knows weather conditions

Knows passenger lists

Influences ground processes

Leverages different tools / technologies

Oversees all type of vehicles

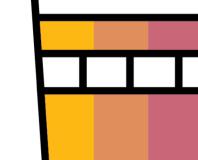
Schedules all startings

Orchestrates landing strip

Knows traffic

Schedules all landings

Knows crews / passengers / destinations

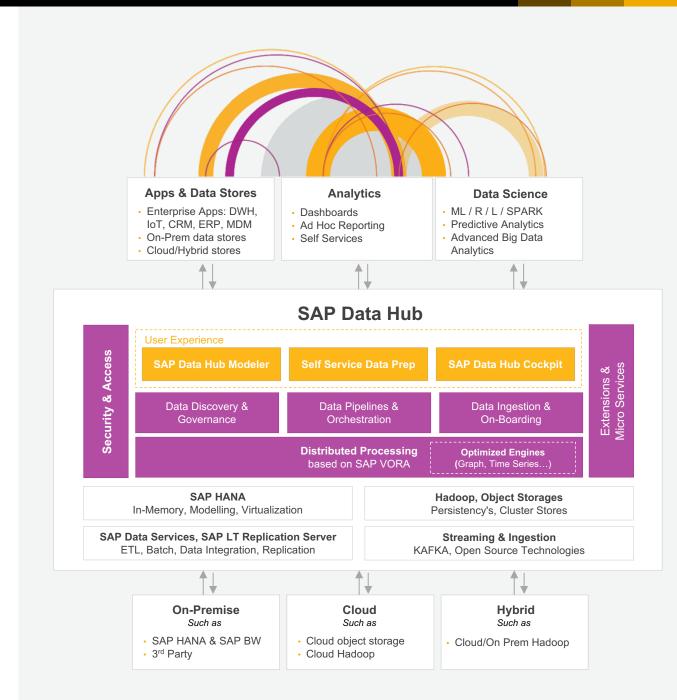


SAP Data Hub

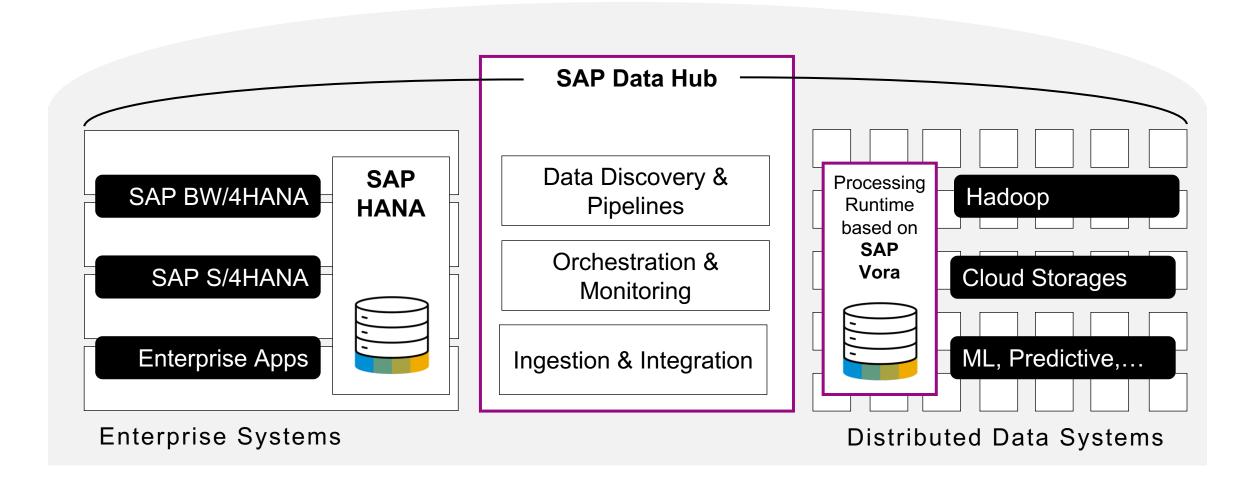
Overview

Define data driven processes across complex enterprise landscapes

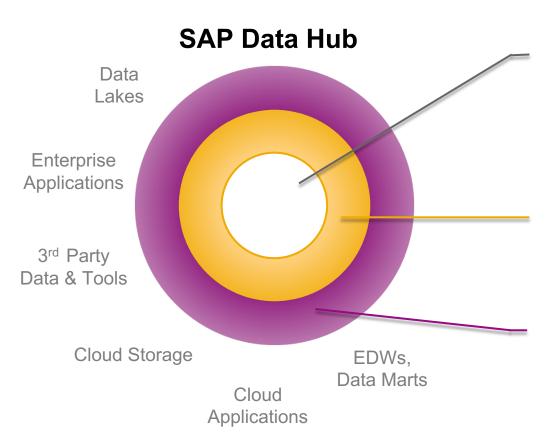
- Access on-premises, cloud, or hybrid data sources – SAP or non-SAP (Amazon, Hadoop)
- Leverage robust enterprise integration capabilities
- Connect easily to SAP data management and application solutions as data sources
- Connect to SAP and non-SAP applications and analytic solutions as endpoints



SAP Data HubUnifying Data Silos

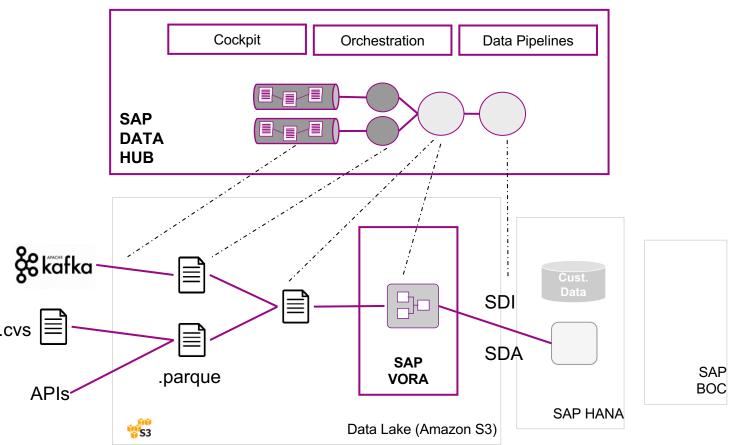


SAP Data HubKey Functions & Capabilities



- Data Pipelines flow based applications consisting of reusable and configurable operations, e.g. ETL, Preparation, Code Execution, Connectors
- Workflows orchestrate processes across the data landscape, e.g. executing data pipelines, triggering SAP BW Process Chains, SAP Data Services Jobs and many more.
- **Governance** metadata repository of information stored in the connected landscape. Offering discovery, profiling and search capabilities.

SAP Data Hub – Streaming and Processing IoT Data



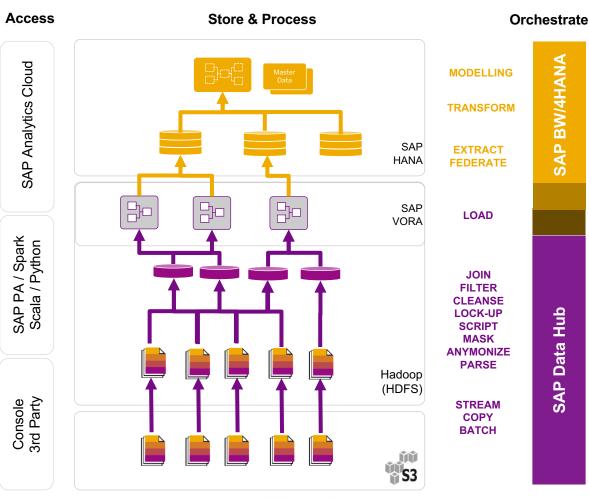
How to understand customer behavior and to drive insightful decisions? - Customer Example

- Smart appliances sending sensor data which need to be processed
- ~6 million devices, 16 TB data volume per day

Solution with SAP Data Hub

- Refine business value from data ingestion to enterprise applications
- Visual modeling environment
- Governance and Data Management
- Orchestration and scheduling to define automated data driven processes

SAP Data Hub – Analyzing Web Log Data with Data Warehouse



Example Scenario

- Combine refined big data with enterprise data and corporate master data
- Extract or federate data into SAP BW/4HANA

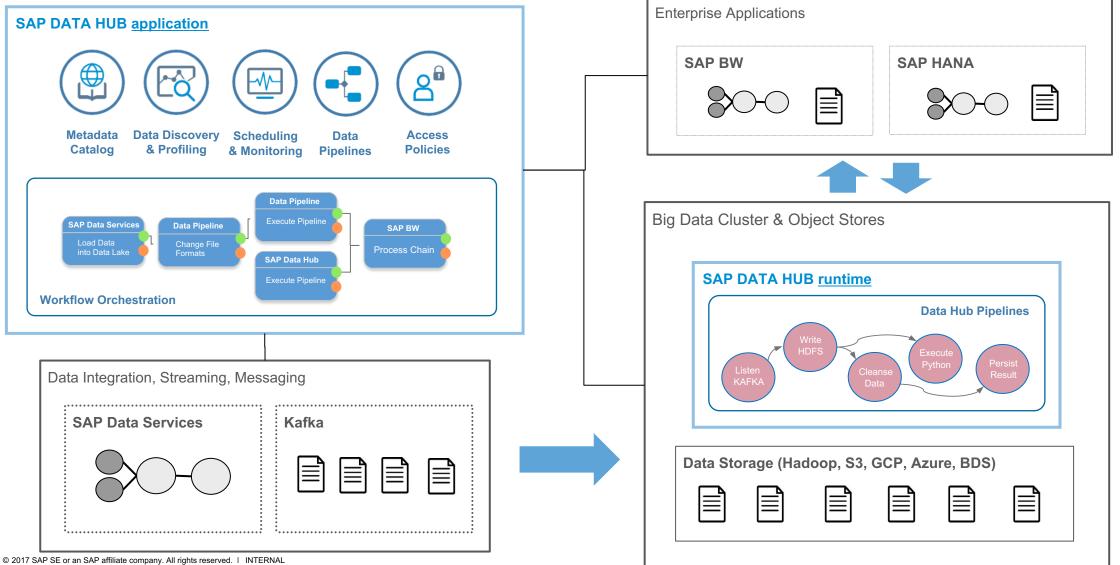
- Ingest Data into S3 as Landing Zone for data
- Orchestrate and schedule all related processes
- Implement transformations and data pipelines
- Harmonize data structures and look up of reference data
- Execute operations on large data volumes







Data Orchestration & Data Pipelines



External Materials

Main SAP Data Hub Page

https://www.sap.com/products/data-hub.html

Franz Faerber (EVP P&I Big Data and SAP Vora) Keynote

http://events.sap.com/sapandasug/en/session/32274

Press release and video from Greg McStravick

https://news.sap.com/new-sap-data-hub-tames-data-landscape/

TechEd 2017 Demo

https://www.youtube.com/watch?v=y3XIF-s75Bk

SAP Data Hub Youtube Channel

https://www.youtube.com/channel/UCLMsNLj0GF0nEw3iixeSEMw

Product road map & vision

RTC Scope

Data Pipelines & Processing

- Modelling of Data Pipelines and Flows (e.g. Data Quality, Prep., embedded Scripts) for Big Data
- Distributed execution via SAP Vora
- Content Lifecycle & Management

Orchestration

- Workflows across landscape
- Remote scheduling for SAP BW/4HANA & SAP Data Services

Data Discovery

 Visually browsing for data in the landscape, data profiling with predefined KPIs to check quality

Hub Management

- Cockpit with unified monitoring for complete landscape
- Define Zones & logical systems
- Adapter framework for Connections
- Access control management

2017 - Planned innovations

Orchestration

 Remote Process scheduling for native SAP HANA scenarios

Security

Policy Creation & Security Logging

Meta Data Governance

Object Lineage & Impact Analysis

Self Services with SAP ADP

- Self Service Data Prep with SAP Agile Data Preparation
- Integration with SAP BO Cloud

Data Pipelines & Streaming

- Streaming integration (Kafka)
- Further operations: Delete, SAP DLM, SAP BW data
- Embedding SAP Predictive Analytics models & ML
- Extensible operator concepts
- Container based delivery

2018 - Product direction

SAP Data Hub in the Cloud

SAP Cloud Platform (PaaS)

Data Ingestion & On-Boarding

 Enhancement of connection framework with inbuild data loading via SAP Data Services

Open Source Integration

 Embracement and Integration of key technologies (e.g. Ranger/Atlas)

Meta Data Governance

- Data Lineage for complete Big Data Stack
- Search and crawling for sources

Integration with Enterprise Applications

 Process orchestration with integration to IoT, S/4HANA, Master Data Management etc.

2019 – Product vision

SAP Data Hub as a Service

 Offering as a Service in SAP Cloud Platform (SaaS)

Enterprise Information Management

- Unifying existing capabilities and data integration portfolio
- Definition of governance rules

Business & Industry Content

 Delivery of content for business scenarios and industry use cases

Integration with Enterprise Applications

- Further end to end scenarios
- Interfaces and integration support for 3rd Party applications

Big Data Applications

Enabling Data driven & automated Big Data Enterprise applications

This is the current state of planning and may be changed by SAP at any time.

Thanks!

