DataHub: a radical shift in data management

Nicola Bertini, Sr Presales Specialist, SAP Cloud Platform, Big Data Service, EMEA South Champion
ENTERPRISE DATA LANDSCAPES ARE GROWING INCREASINGLY COMPLEX

- Proliferating Data Lakes
- Enterprise Applications (ERP, CRM, HR)
- Mobile Apps
- Multiple EDWS, Data Marts
- Master Data Management
- 3rd Party Data
- Cloud Apps
- BI and Data Visualization
- Cloud Storage
- Data and Systems of Partners, Customers
Organizations struggle to design a business-relevant infrastructure that is both effective and efficient at mediating differing semantics (e.g., governance) to support data sharing and integration.”

Gartner, Use a Data Hub Strategy to Meet Your Data and Analytics Governance and Sharing Requirements, Andrew White, Ted Friedman, February, 2nd 2017
THE EVOLUTION OF DATA INTEGRATION IN ENTERPRISE LANDSCAPES

DATA HETEROGENEITY

OLTP TO OLTP
- MFT, EAI, ESB, SOA, B2B, BPM/BRM ...

OLTP TO OLAP
- ETL, EDW, BI, MIS, Data Marts ...

HETEROGENEOUS TO HETEROGENEOUS
- HTAP, Big Data etc.
- Streaming, IoT, ML, Big Data ...
CHALLENGES TO UNITING THE LANDSCAPE
Overcoming silos, complexity to drive better operations and insight

- Lack of Enterprise Governance, Security of Big Data solutions
- Missing Link between Big Data and Enterprise Data
- Limited Tools = High Effort to productize complex data scenarios across data landscape
SAP DATA HUB – FREEDOM OF DATA IN A COMPLEX WORLD

Simplify, understand, and govern the flow of data at scale

GOVERNANCE & OPERATIONS
- Unified data landscape view
- Data access and security
- Data quality profiling
- Lineage and impact analysis

DATA PIPELINES
- Meta data repository and catalog
- Data pipeline modeling
- Data model lifecycle management

Data Sharing & Orchestration
- Connectivity management
- Workflow definition
- Scheduling and monitoring
- Enterprise application integration

© 2018 SAP SE or an SAP affiliate company. All rights reserved. | PUBUC
Define data-driven processes across complex enterprise landscapes

SAP Data Hub will allow you to manage data:
- Cloud or on-premise: Access on-premise, cloud, or hybrid data sources
- Across a breadth of data stores: SAP data stores or third-party (Amazon, Hadoop)
- Across a breadth of endpoints: Connect to SAP and non-SAP applications, data stores, and analytic solutions as endpoints
- By leveraging robust enterprise integration capabilities

SAP DATA HUB AS DATA SHARING AND PIPELINING MANAGER

APPLICATIONS
- Enterprise - IoT, CRM, ERP
- Mobile

ANALYTICS
- Dashboards
- Standard and ad hoc reporting

DATA STORES
- Business Warehouse
- On-premise data stores
- Cloud and hybrid stores

SAP DATA HUB

User Experience
- SAP Data Hub Modeler
- Self Service Data Prep
- SAP Data Hub Cockpit

Security
- Data Discovery & Governance
- Data Refinery & Orchestration
- Data Ingestion & On-Boarding

Data Hub Distributed Native Processing

Extensions
SAP DATA HUB: CENTRALIZED GOVERNANCE OF THE DISTRIBUTED LANDSCAPE

Visibility, control, and power, without mass centralization of data

- Single system view – for data governance, monitoring, orchestration
- No centralization of data – no mass data movement to a single data store
- Distributed “push down” native processing – executes pipeline activities quickly, where the data resides

Pipelines
Governance
Sharing

Big Data Lakes
Customer Systems and Data
3rd Party Data
BL and Data Visualization
Cloud Storage
EDWs, Data Marts
Cloud Applications
Partner Systems and Data
Enterprise Applications
SAP Data Hub
SIMPLER, MORE SCALABLE APPROACH TO MANAGING THE DATA LANDSCAPE

EASE THE COST AND EFFORT OF DATA MANAGEMENT AND INTEGRATION
- Connect quickly to your diverse, distributed data landscape
- Easily include more data sources and targets
- Empower more users to manage data and data pipelines with a single, easy-to-use solution

RESPOND QUICKLY TO OPPORTUNITIES TO CREATE AND IMPROVE DATA FLOWS
- Identify opportunities to connect systems and information to gain new insights
- Accelerate and automate data processes; remove unnecessary process redundancies
- Improve the effectiveness of data results by resolving data quality issues or friction points
ACCELERATE AND EXPAND YOUR DATA PROJECTS

WORK ACROSS A DIVERSITY OF SYSTEMS AND DATA

- Open architecture to work across a broad array of data sources: cloud or on-premise, Big Data or enterprise data, SAP or non-SAP (third party)
- Common orchestration of data and processes to overcome system boundaries
- Accelerated development time with visual modeling

MAGNIFY THE IMPACT OF YOUR DATA

- Ensure data quality with predefined operations to cleanse, process, and transform data
- Process large volumes of data where they occur to gain the insights you need
BUILD AGILE, DATA-DRIVEN APPLICATIONS AND PROCESSES

CREATE POWERFUL DATA PIPELINES
- Leverage distributed Big Data processing
- Take advantage of serverless computing paradigms for radical scale
- Incorporate powerful third-party libraries such as TensorFlow

BUILD DATA STREAMS TO ENABLE DATA-DRIVEN PROCESSES
- Establish data streams to react to data changes instantly
- Embed complex algorithms into existing data flow by leveraging any code or script
ENTERPRISE VISIBILITY AND GOVERNANCE ACROSS THE CONNECTED DATA LANDSCAPE

UNDERSTAND THE OVERALL DATA LANDSCAPE IN A MODERN, UNIFIED COCKPIT
- Describe and organize the landscape within a single environment
- Leverage a single entry point for monitoring, scheduling, and landscape health

RAPIDLY ITERATE AND ADVANCE YOUR DATA MODELS
- Build a repository of reusable and extensible metadata models

ENSURE DATA INTEGRITY AND SECURITY
- Enforce access policies to dynamically secure data from source to destination
- Mask and anonymize data to protect sensitive sources
Data lakes

Enterprise data warehouse

BIG DATA WAREHOUSING

- Modern, open, and hybrid data warehouse tailored for all kinds of data volumes and formats
- Increases the effectiveness of a data warehouse
- Embraces data lakes for high-volume data operations
- Aligns processes across data lakes and your data warehouse
BUSINESS BENEFITS & VALUE DRIVERS

INNOVATION FOR ENTERPRISE INFORMATION MANAGEMENT (EIM) SOLUTIONS

- Manage multiple SAP Data Services instances simultaneously
- Helps EIM reach into the Big Data world by adding data lake access
- Brings SAP Data Services into a complete context, due to pipelining capabilities
- Goes beyond read and write to include processing and scheduling
- Makes self-service data prep more robust, goes beyond SAP HANA sources
Manage multiple SAP Data Services instances simultaneously

Helps EIM reach into the Big Data world by adding data lake access

Brings SAP Data Services into a complete context, due to pipelining capabilities

Goes beyond read and write to include processing and scheduling

Makes self-service data prep more robust, goes beyond SAP HANA sources
Big Data warehouse use case

- Acquiring new data sources: previously siloed data, social media, new third-party databases
- Example: Merger success depends on the ability to integrate data from two organizations quickly and efficiently
- SAP Data Hub assists in data discovery, semantic analysis, accelerating data pipelines, and auditing the flow of data

EIM use case

- Move to real-time monitoring demands ability to pull data in real time from multiple sources and audit the completeness of the data.
- SAP Data Hub can automate and audit data aggregation

IoT use case

- To drive product development and marketing, manufacturers can combine:
  - Real-world performance information from Internet-enabled devices, like appliances
  - Customer demographics
  - Supply chain information
- SAP Data Hub unites data from high-volume stores such as Amazon S3, as well as data from SAP data management solutions and enterprise applications
TOP 5 REASONS FOR CHOOSING THE SAP DATA HUB SOLUTION

01 UNIVERSAL
Universal view of the enterprise and Big Data: Get a consolidated view of all data from all data sources, covering business processes and applications

02 INTELLIGENT
Intelligent discovery of data relationships: Improve data quality through cleansing and get a graphical view of data correlations and lineage across your enterprise

03 EFFICIENT
Efficient data enrichment: Employ distributed data pipeline processing and refinement using a variety of computation techniques such as OLAP, graph, time series, and machine learning

04 SCALABLE
Scalable data operations (DataOps) management solution: Orchestrate data end to end, process data where it is located, and avoid expensive data movement

05 COMPLIANT
Optimal compliance and data governance across the enterprise: Maintain your security policy dynamically in one place and help ensure that policy measures are in place to meet regulatory and corporate requirements
Thank you.

Nicola Bertini
SAP Senior Presales Specialist – SAP Cloud Platform, Big Data Service, EMEA South Champion
SAP Italia