SAP Data Hub – Turning Disparate Data into Valuable Assets for Digital Innovation

Daniel Rutschmann
Sr. Director SAP Data Hub Solution Management, SAP
Session ID ASUG84047

SAP SAPPHIRENOW | ASUG ANNUAL CONFERENCE

May 7 – 9, 2019

About the Speakers

Daniel Rutschmann

- Sr. Director, SAP Data Hub Solution Management, SAP

- I have worked in the software and technology industry for about two decades. Over the years, I took on different roles from consulting to solution management and business development. In my current role I'm part of the solution management and go-to-market team focused on SAP Data Hub.

- In my free time I enjoy woodworking and being outdoors with our dogs.

Key Outcomes/Objectives

1. Get an understanding of SAP Data Hub

2. Learn how customers innovate with SAP Data Hub

3. Hear about a new announcement regarding SAP Data Hub

asug

Enterprises are struggling to maximize value from their existing data assets
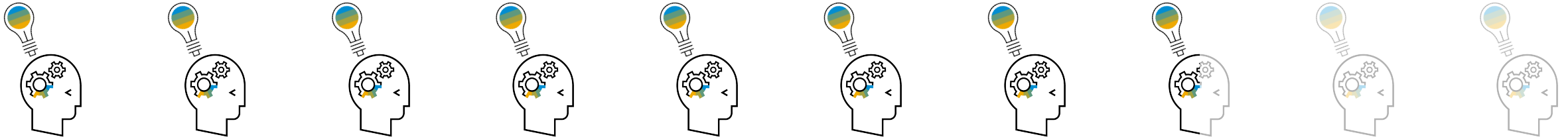
86% of enterprises claim that they are not getting the most out of their data
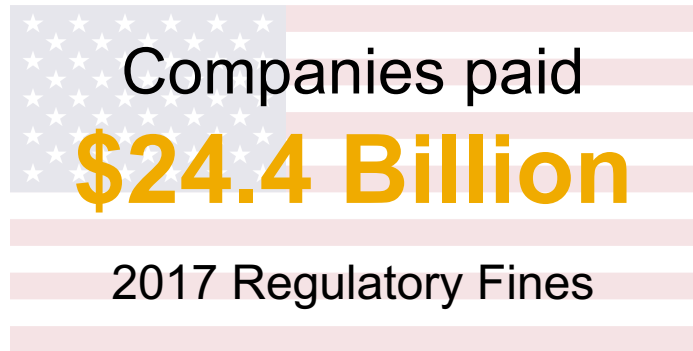
Enterprises are struggling with trusted data when using intelligent algorithms

84% are concerned about the quality of the data
using to feed intelligent algorithms

$9.7M is the average financial impact of poor data quality
on organizations per year

Enterprises are struggling with compliance

Companies paid
$24.4 Billion
2017 Regulatory Fines
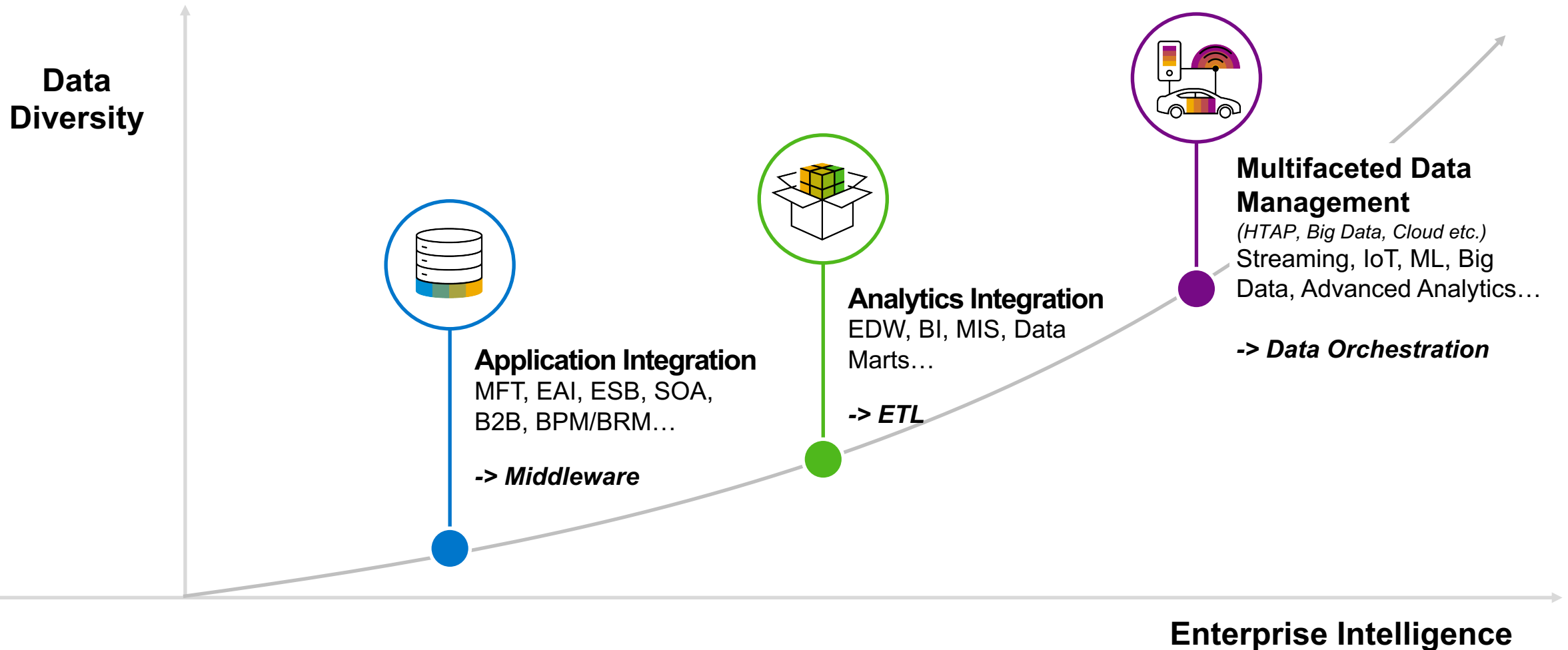
Companies paid
€866 Million
2017 Regulatory Fines

Regulatory compliance can cost as much as **$7k per employee**

$14.5M is the average annual cost for organizations that experience data protection non-compliance problems

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**.

Applications

Data cataloging

Cloud Datastores

Speech recognition

Data masking

ETL

Data Warehouses

Data ingestion

Event Stream processing

Video processing

Data replication

Data cleansing

Geospatial processing

Data quality

Streaming analytics

Databases

ELT

Data profiling

Time series

Graph processing

Image processing

Text analytics

Data Marts

Meta Data management

Machine learning

Third-Party Data

Customer Experience

Manufacturing & Supply Chain

Digital Core

People Engagement

Network & Spend Management

The evolution of data management in enterprise landscapes

Data Diversity

Application Integration
MFT, EAI, ESB, SOA, B2B, BPM/BRM…

-> Middleware

Analytics Integration
EDW, BI, MIS, Data Marts…

-> ETL

Multifaceted Data Management
(HTAP, Big Data, Cloud etc.)
Streaming, IoT, ML, Big Data, Advanced Analytics…

-> Data Orchestration

Enterprise Intelligence

A master conductor orchestrates diverse sounds from disparate sources

into beautiful music

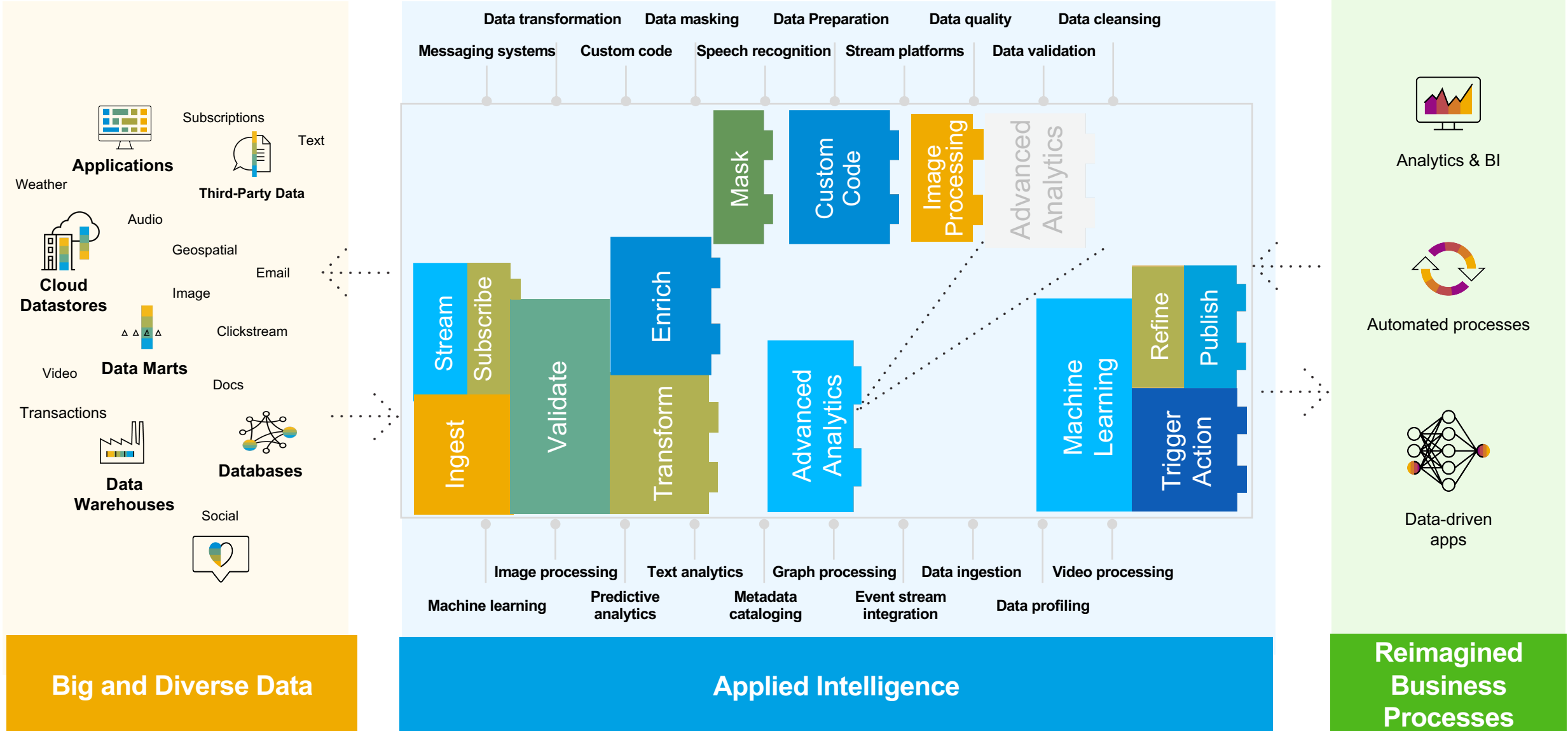# SAP Data Hub orchestrates diverse data from disparate sources
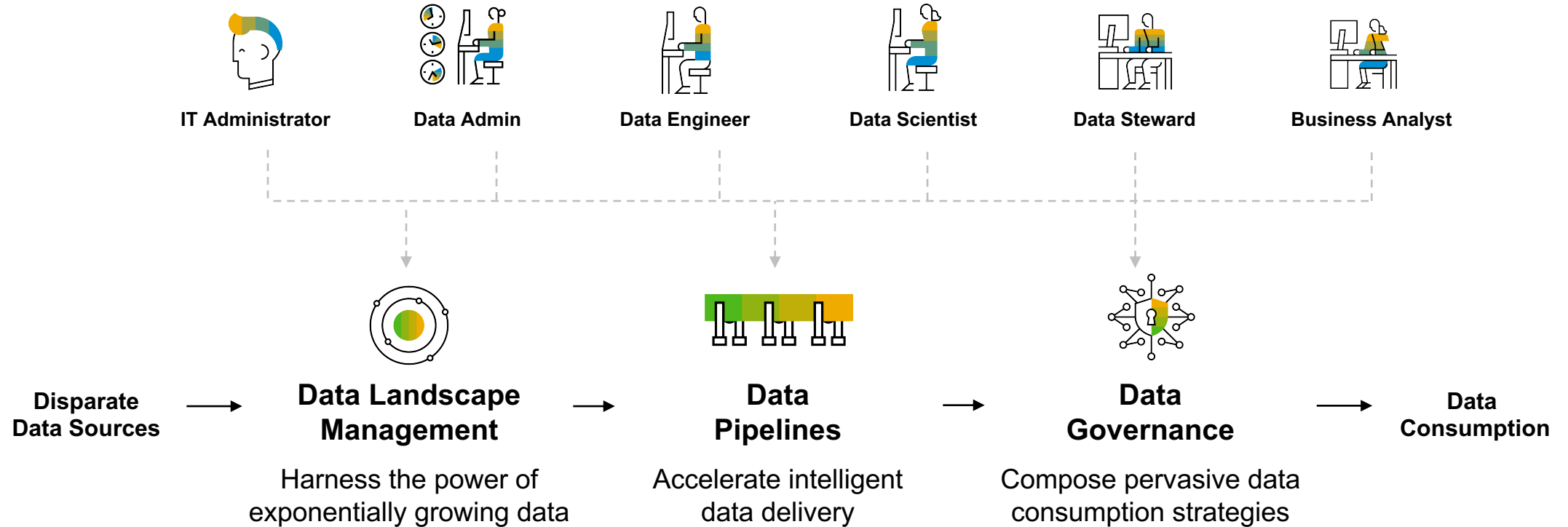

into refined,
governed,
intelligent insights.

Data Lakes

Data Warehouses

Third-Party Data

Databases

Applications

Data Marts
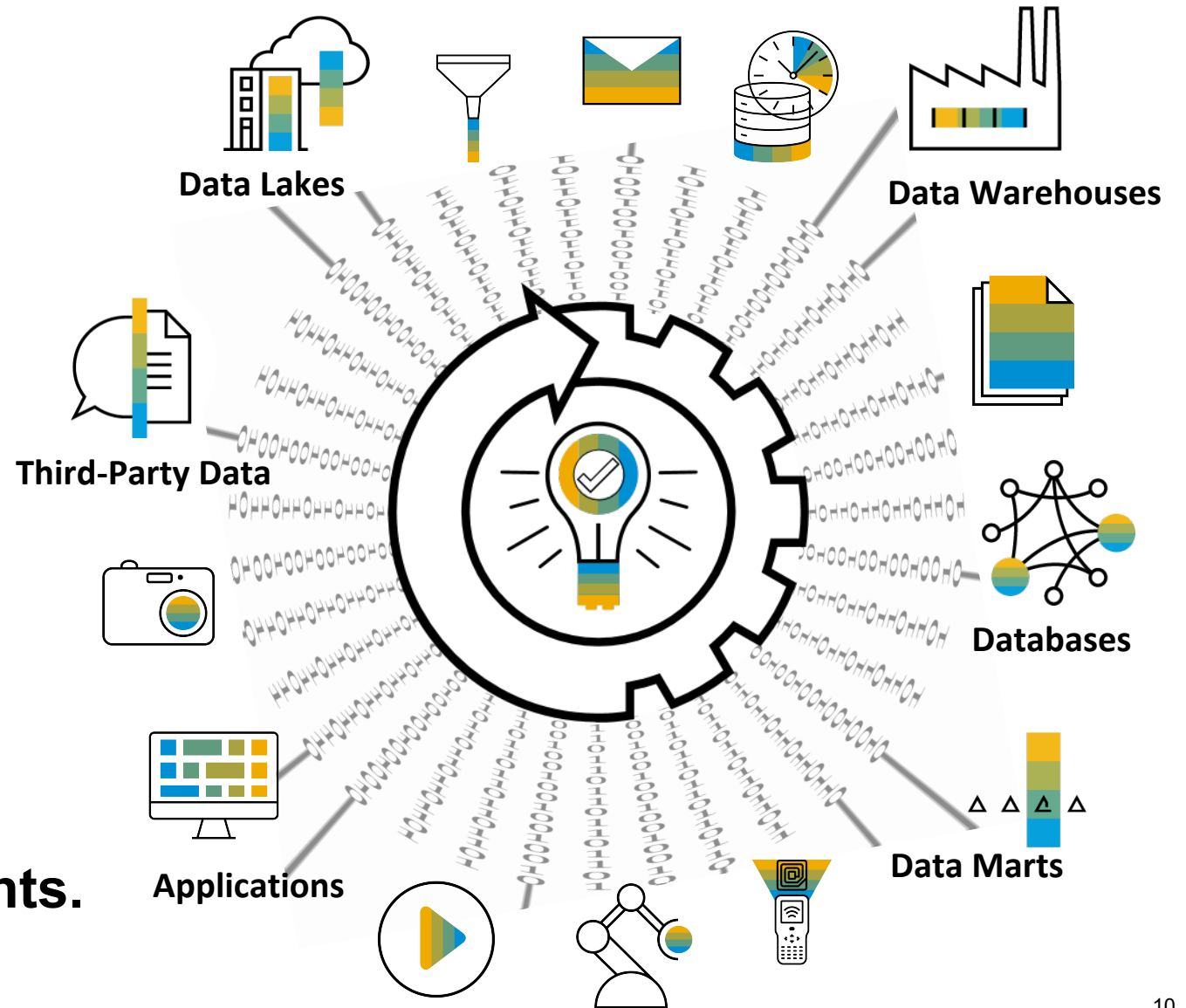
SAP Data Hub extracts value from distributed data assets
Streamlining data-driven innovation for the intelligent enterprise
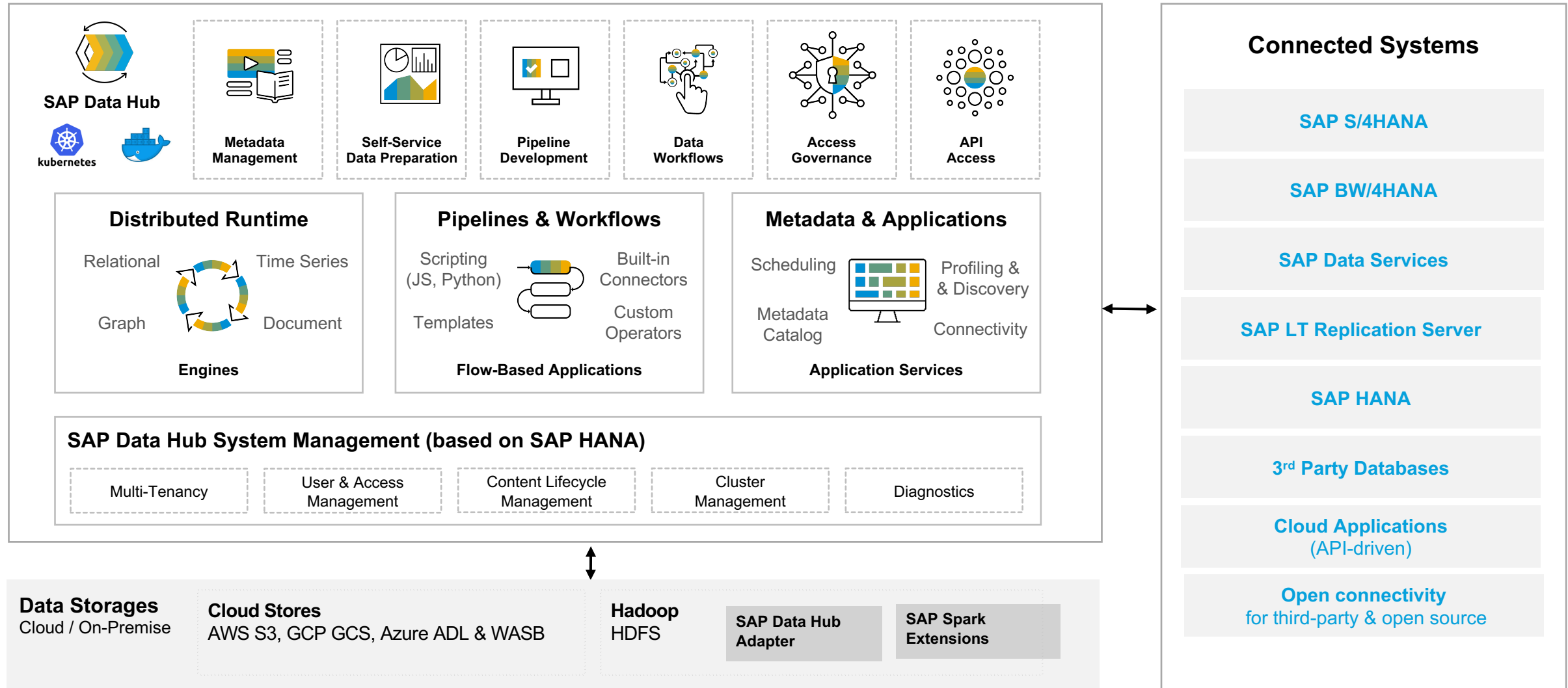
Delivering intelligent data for the intelligent enterprise

Right data to the right users with the right context at the right time

IT Administrator   Data Admin   Data Engineer   Data Scientist   Data Steward   Business Analyst

Disparate
Data Sources → **Data Landscape
Management** → **Data
Pipelines** → **Data
Governance** → Data
Consumption

Harness the power of
exponentially growing data

Accelerate intelligent
data delivery

Compose pervasive data
consumption strategies

Architecture Overview of SAP Data Hub

SAP Data Hub

kubernetes

| Metadata Management | Self-Service Data Preparation | Pipeline Development | Data Workflows | Access Governance | API Access |

Distributed Runtime

Relational   Time Series

Graph   Document

Engines

Pipelines & Workflows

Scripting (JS, Python)   Built-in Connectors

Templates   Custom Operators

Flow-Based Applications

Metadata & Applications

Scheduling   Profiling & & Discovery

Metadata Catalog   Connectivity

Application Services

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

SAP Data Hub Adapter

SAP Spark Extensions

Connected Systems

SAP S/4HANA

SAP BW/4HANA

SAP Data Services

SAP LT Replication Server

SAP HANA

3rd Party Databases

Cloud Applications
(API-driven)

Open connectivity
for third-party & open source

SAP Data Hub
Cloud Deployments

Simplified deployment of SAP Data Hub in Cloud environments and on-premise

- All necessary components are fully containerized and delivered as a Docker images

- Decoupling data processing (in Kubernetes) and data storages (any supported cloud stores, databases or Hadoop)

- Deployment in multiple Kubernetes managed environments*

  - Leveraging managed cloud Kubernetes Services in AWS, Microsoft Azure, Google Cloud Platform,

  - Support for Private Cloud & on-premise installations, e.g. Red Hat, SUSE, Cisco, Virtustream etc.
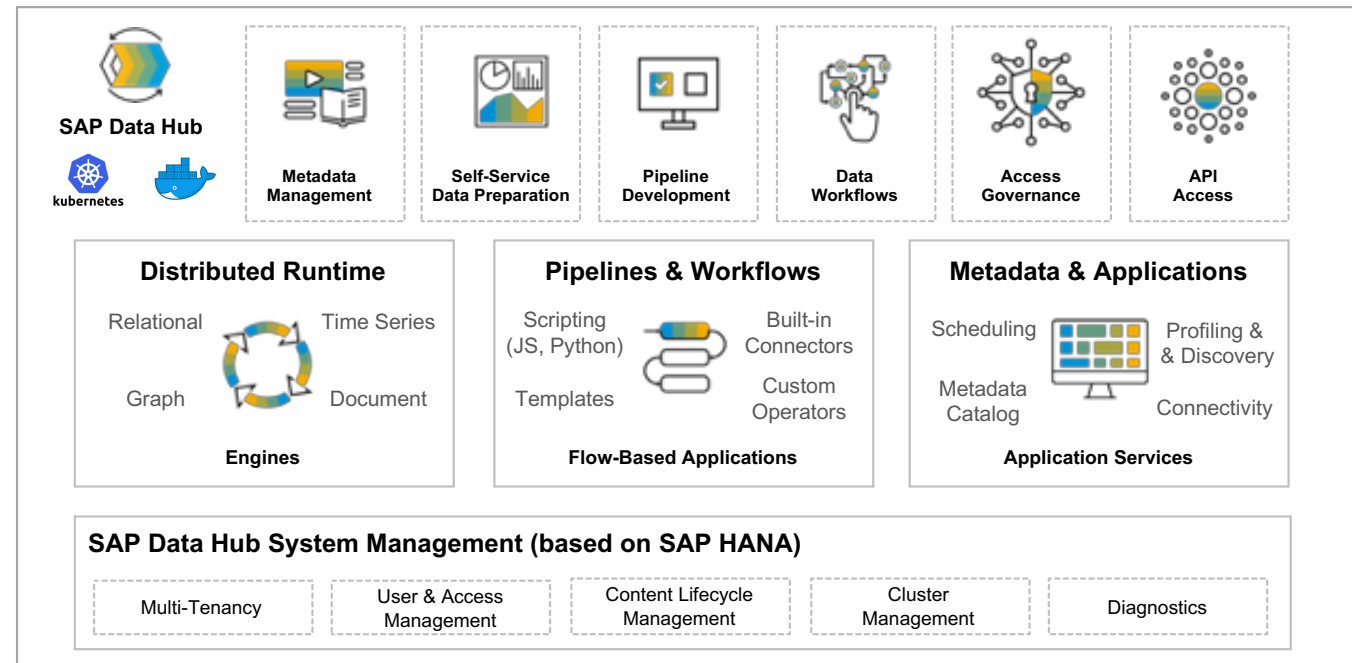
* See Product Availability Matrix for detailed version dependencies

SAP Strategy – Deliver the Intelligent Enterprise

THE INTELLIGENT ENTERPRISE

features **3 KEY COMPONENTS:**

1 Intelligent Suite

2 Digital Platform

3 Intelligent Technologies

Customer Experience

Manufacturing & Supply Chain

Digital Core

People Engagement

Network & Spend Management

Intelligent Suite

Intelligent Technologies

AI/ML | IoT | Analytics

Digital Platform

Data Management

Cloud Platform
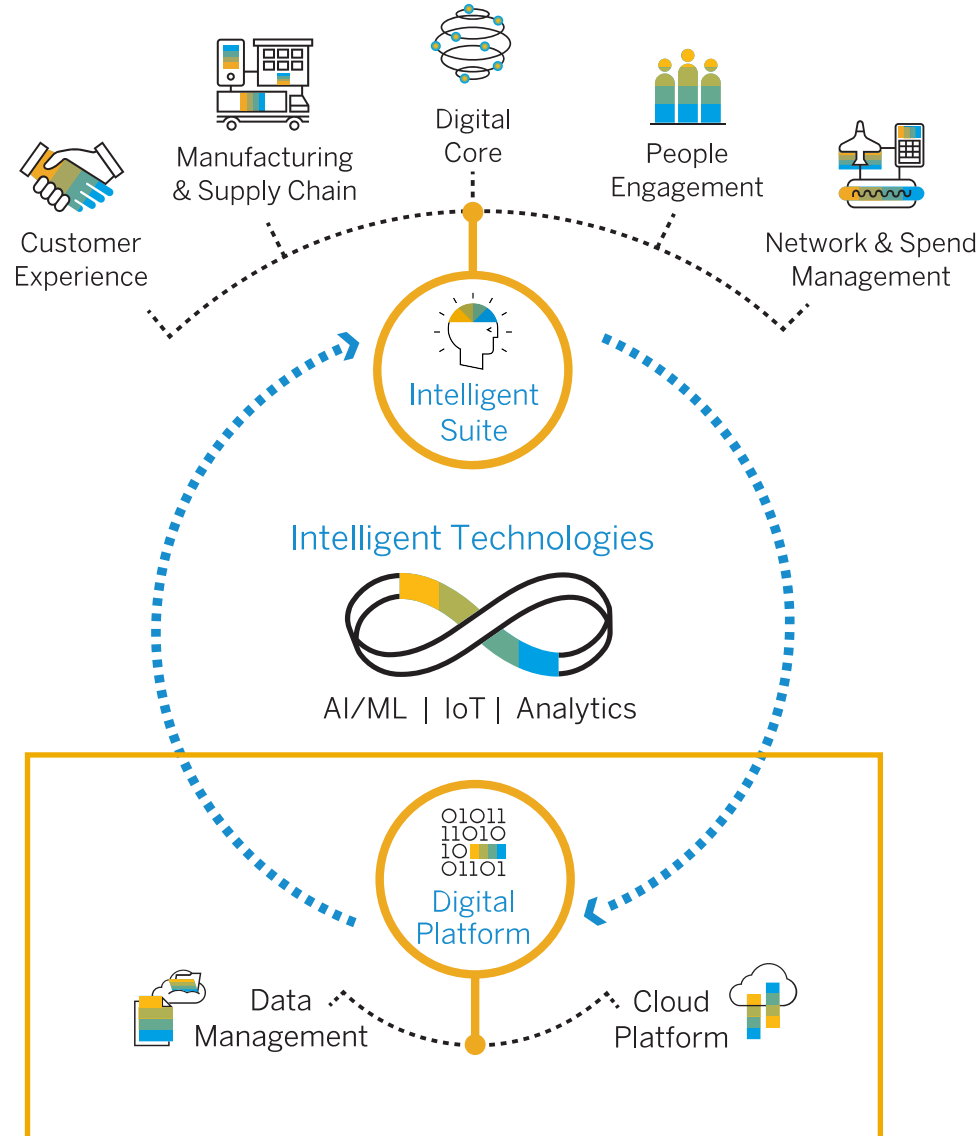
2 Digital Platform

SAP DATA HUB

Discovers, refines, governs, orchestrates any type, variety and volume of data across distributed landscapes

Use Cases and Customer Success Stories

Use SAP Data Hub to reimagine your business processes using …

Business Application Transformation

Streamline innovation initiatives around Business Applications, supporting enterprise transformation programs

IoT Ingestion and Orchestration

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

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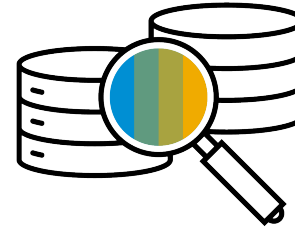
Build a multifaceted data warehouse, across diverse and distributed data assets

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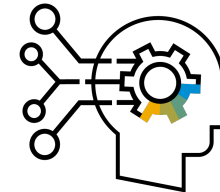
Streamline data science and machine learning, from modeling and development to operations, across all enterprise data assets

Metadata Catalog and Governance

Automate metadata discovery, understand data sources, prepare and govern according to data standards, privacy, and security

IoT Data Ingestion and Orchestration and SAP Data Hub

Integrate and process disparate data from messaging systems and high-volume cloud storages

DATA STREAMS

SAP Data Hub

IoT App

SAP HANA

Data Lake

- 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

Customer-driven Product Insight through IoT ingestion
The problem

Home4You manufactures **Smart Dishwashers** that collect sensor and machine data.
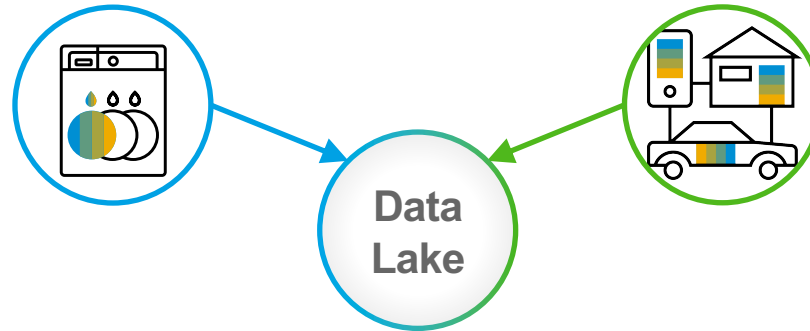
They also maintain an **IoT Collection** of user behaviors.

Data Lake

Feeding this information to the **Data Lake** will help them build better dishwashers

But there are some problems:

6 million+ devices generated **>16 TB** streaming data a day
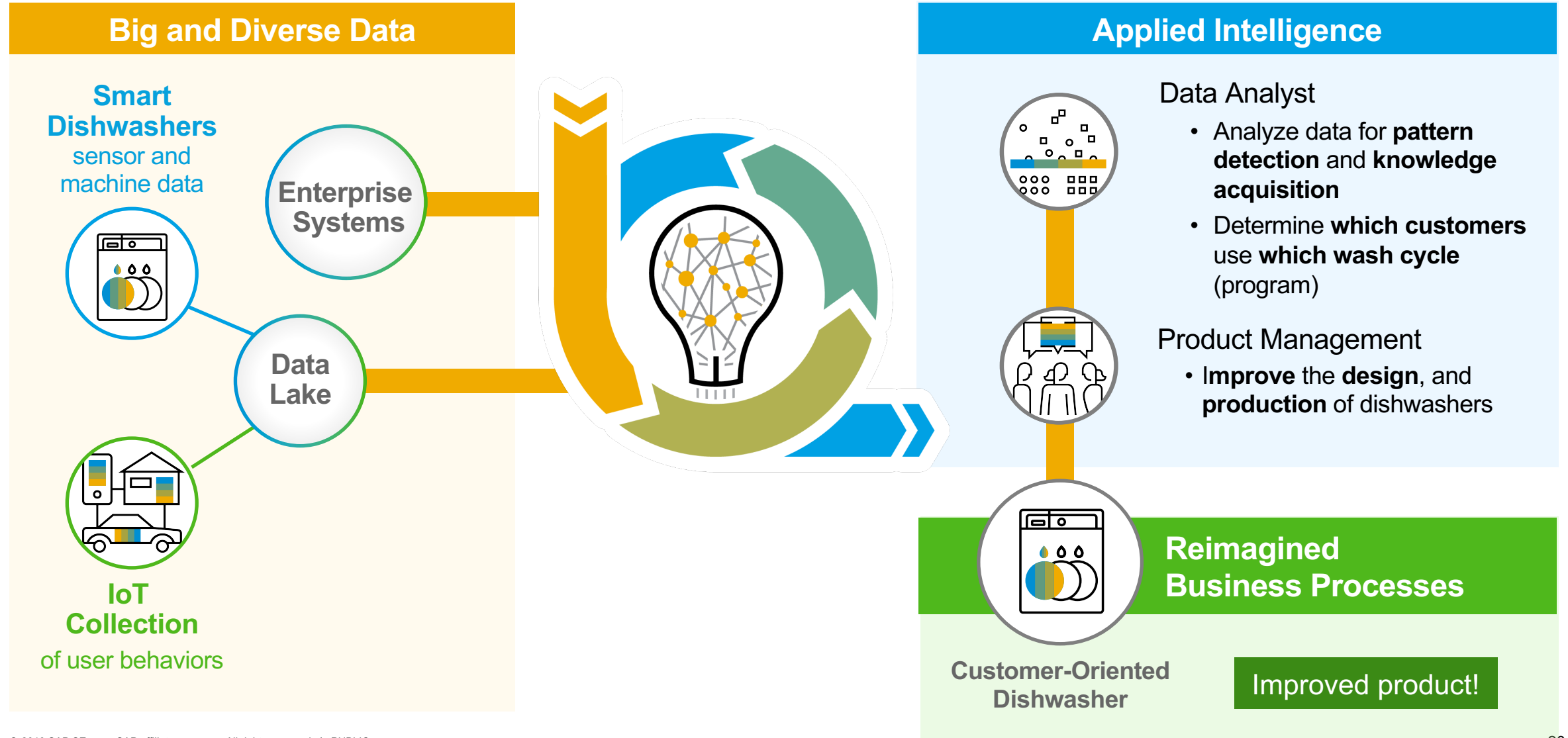
The data collected isn't easily understood

The enterprise is fragmented. There are operational silos. No feedback loop between data scientists and operational systems.

Incoming data **isn't quality checked**

Enterprise data **isn't connected**.

Customer-driven Product Insight through IoT ingestion

The solution with SAP Data Hub

Big and Diverse Data

Smart Dishwashers
sensor and machine data

Enterprise Systems

Data Lake

IoT Collection
of user behaviors

Applied Intelligence

Data Analyst
- Analyze data for **pattern detection** and **knowledge acquisition**
- Determine **which customers** use **which wash cycle** (program)

Product Management
- **Improve** the **design**, and **production** of dishwashers

Reimagined Business Processes

Customer-Oriented Dishwasher

Improved product!

Customer-driven Product Insight through IoT ingestion

The architecture and its capabilities
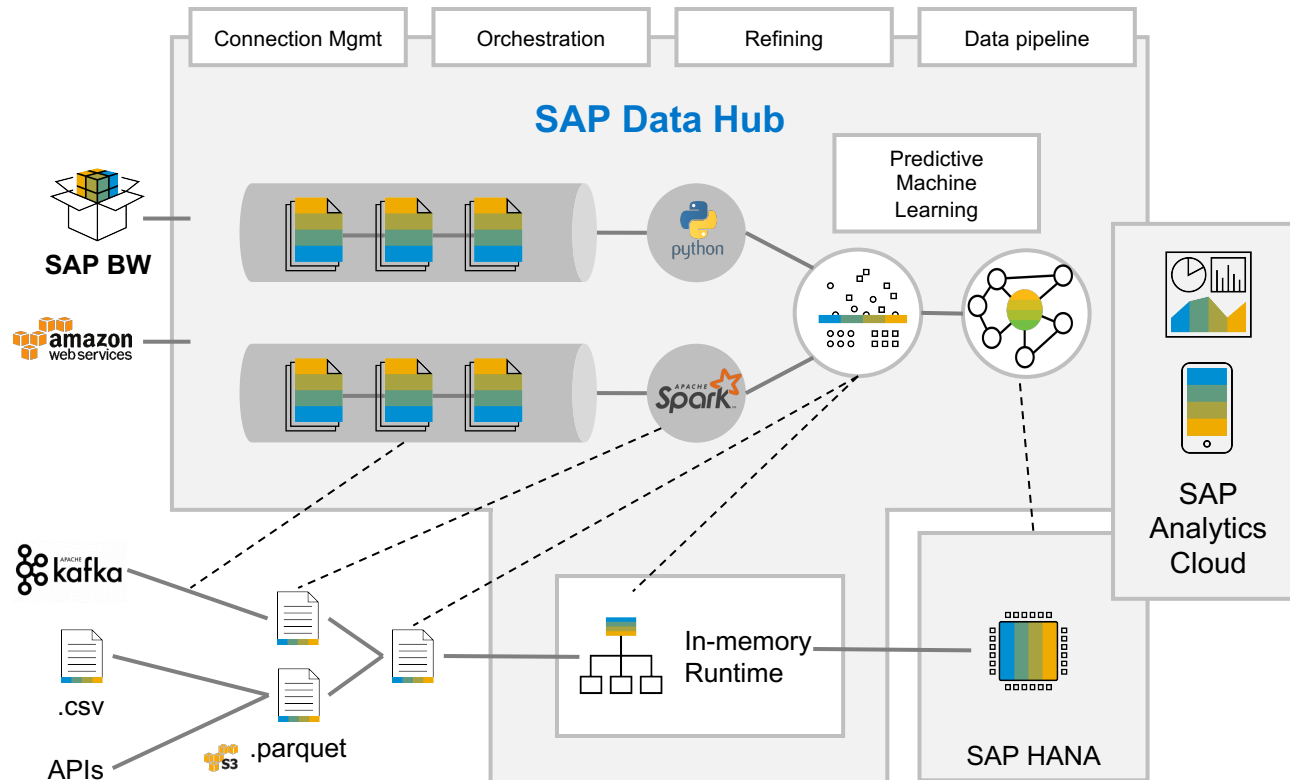
DISCOVER
- Connect any distributed and heterogeneous enterprise data
- Ingest data via streaming, batch, replication

REFINE
- Cleanse, validate, enrich data
- Machine learning algorithms
- Relate IoT data to structured data

GOVERN
- Catalog metadata from IoT, BW, and other applications
- Data lineage to data origins
- Data anonymization for privacy

ORCHESTRATE
- Define, execute and control modular data pipelines
- Enable real-time action and decision making faster with less costs

Data Science & Machine Learning Data Management and SAP Data Hub

Prepare data from all sources to increase effectiveness of AI/Machine learning algorithms

SAP Data Hub

Data Prep   Machine Learning

App

App
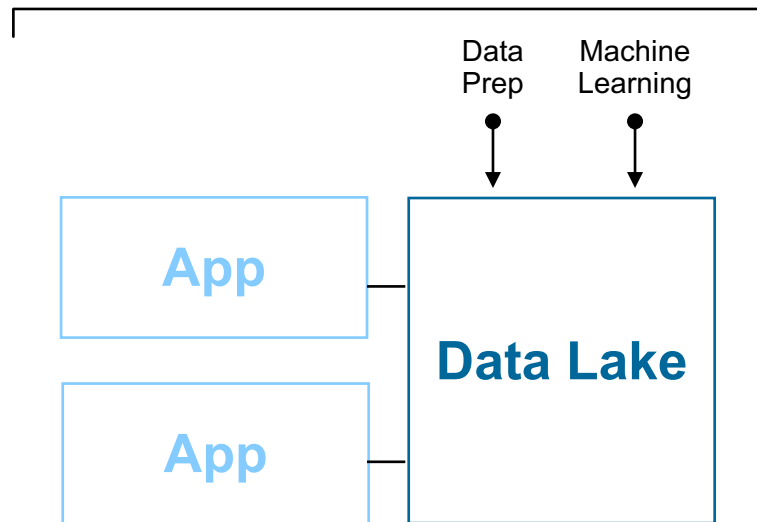
Data Lake

DISCOVER

- Integrate variety of data sources with open data landscape management
- Ingest any type of data with one tool

REFINE

- Process machine learning models leveraging SAP and non-SAP engines within the same tool
- Quickly operationalize machine learning outcomes to enterprise processes

GOVERN

- Monitor, profile, index, publish metadata
- Prepare data with profiling and the fact sheet

ORCHESTRATE

- Define, execute and control modular data pipelines
- Manage high volume streams
- Leverage both SAP and open-source processing engines
- Bring IT and Data Science Teams together

Predictive Quality in Manufacturing Plant through Data Science
Improving Quality Management in Car Manufacturing

Predictive Quality in Manufacturing Plant through Data Science

The prototype without Data Hub: could not go beyond experimental stage

Apache kafka — Event Streaming

Enterprise Data Center

SAP HANA

ERP Data

Manufacturing plant

Raw Material — Molding Press — Pressure Sensor

IR Camera

Cloud Data Lake

hadoop

R — Offline Image Processing

Not ready for production

Smart Quality App
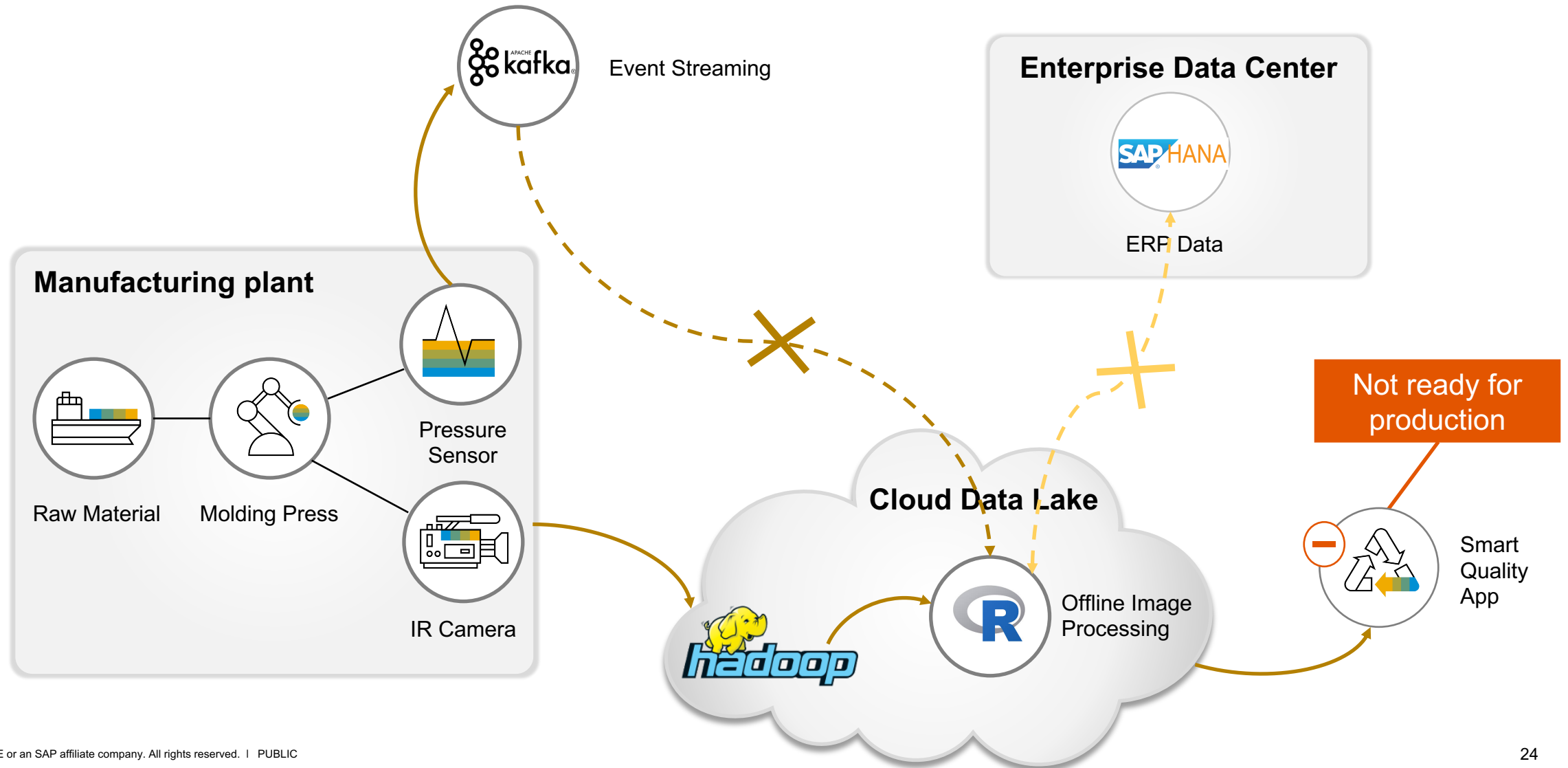
Predictive Quality in Manufacturing Plant through Data Science

The project with Data Hub: Data Science streamlined across the landscape!

Enterprise Data Center

SAP HANA — ERP Data / PAL engine

Correlate Data

Smart Quality App

Manufacturing plant

Raw Material

Molding Press

Pressure Sensor

IR Camera

Event Streaming — kafka

Cloud Data Lake

hadoop

python — Image Processing

Production ready!

Data Hub pipelines
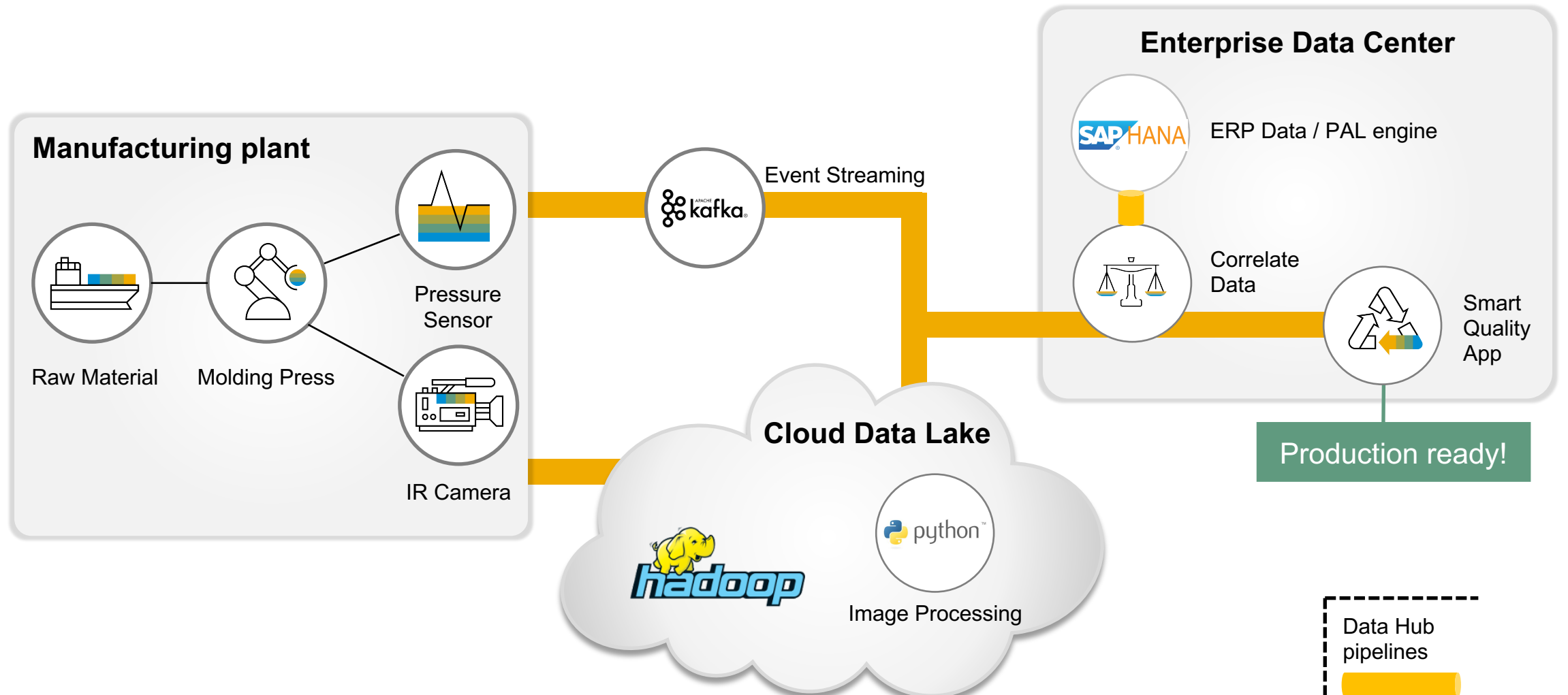
Summary
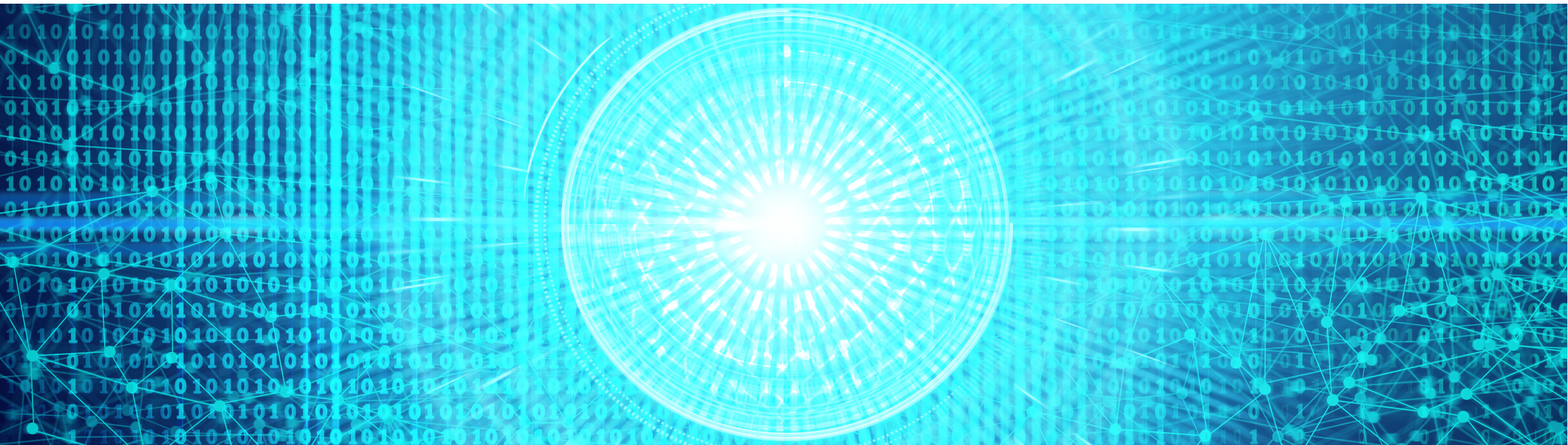
SAP Data Hub Simplifies Distributed Data Assets

Data Lakes

Databases

Data Warehouses

Data Marts

Third-Party Data

Internal Applications

External Applications
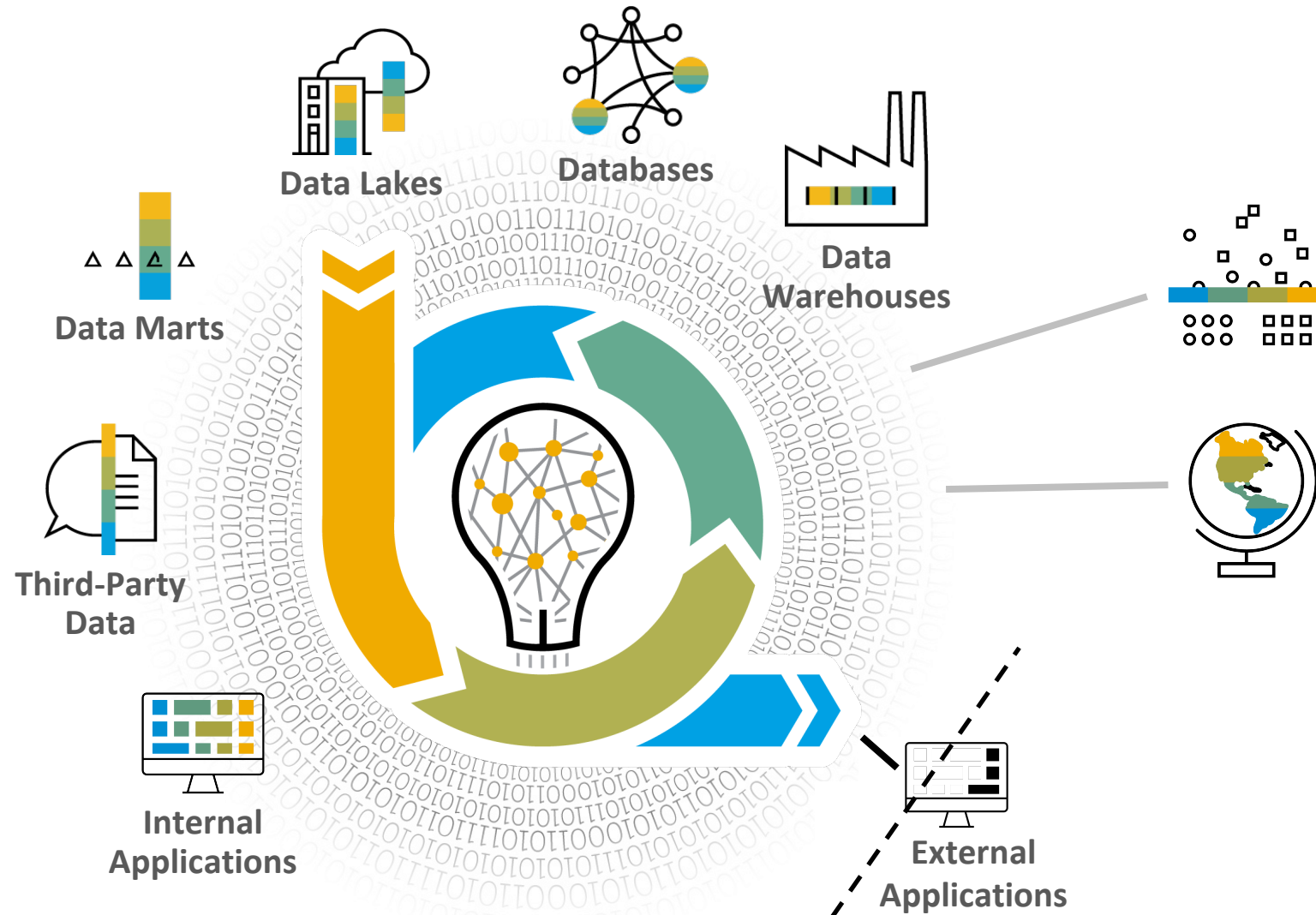
No lift and shift.
No storing.
No point-to-point.

No matter the type.

No matter the location.

No Silos.

Key Take-Aways

SAP Data Hub …

- Enables you to extract value from distributed data.

- Connects assets and transforms them into business insights.

- Discovers, refines, governs, and orchestrates any type, variety, and volume of data across your distributed data landscape.

- Applies machine learning and intelligent technologies to reimagine your business process.

- Leverages SAP engines and open source engines across distributed infrastructures, minimizing data movement.

- Accesses structured, unstructured, streaming data across all data types.

- Enables you to get the most out of your data, simplify visibility across your landscape, provide trust in intelligent data, with governance, security, compliance.

How Does a Global Manufacturer Evolve to Compete in a Service-Oriented World?

Delivering "air as a service" to a global customer base

Air compressors built by Kaeser Kompressoren SE are in constant use, many in some of the world's most demanding manufacturing environments. Proactive field support helps ensure customer satisfaction, but truck rolls are expensive. Gathering sensor data remotely from the Internet of Things (IoT) might work to minimize on-site visits, but only if that data could be processed and combined with customer data from existing business systems. If it could interact efficiently with all this data, Kaeser could seize new opportunities to meet customer needs for products, services, and responsiveness around the world.

THE BEST RUN SAP

Picture Credit | Kaeser Kompressoren SE, Coburg, Germany. Used with permission.

ASUG84223: KAESER KOMPRESSOREN: "Smart Air" Strategy Execution

New Forrester Study

FORRESTER®

Forrester Opportunity Snapshot: A Custom Study Commissioned By SAP | April 2019

Hybrid Data Management Drives Innovation And Growth

Utilize A Modern Data Architecture That Enables Enterprise Data Fabric

GET STARTED ▶

"Hybrid data management enables the data fabric architecture by delivering trusted and integrated data to support business apps and insights by seamlessly accessing, orchestrating, modeling, governing, and sharing any type of data, whether on-premises or cloud."

Source: Forrester Opportunity Snapshot – Hybrid Data Management Drives Innovation And Growth

Get Started with SAP Data Hub

Find out more about SAP Data Hub
sap.com/datahub

Try SAP Data Hub,
developer edition

Try SAP Data Hub,
Trial edition

Take free SAP Data Hub Course

Read State of Big Data Report

New Announcement

NEW: SAP Data Intelligence

SAP Cloud Platform (Kubernetes Infrastructure)

SAP Data Intelligence

Machine Learning Services

ML Notebooks

Data Science Workflows

ML Operations

Core Services
Model Training Service
compute custom ML models

Model Inference Service
query custom trained ML models

Functional Services
Pre-trained Services
Face Recognition, Translation, OCR etc.

Customizable Services
Customizable Image Classification, Customizable Text Classification etc.

Other Use Case Specific Services

Data Management and Orchestration Services

Data Governance

Metadata Management

Access Governance

Data Preparation & Labeling

Integration & Orchestration

Visual Pipeline Modeling

Data Workflows

API Access

Foundational Capabilities

Distributed Runtime

System Management

Pipelines & Workflows

Metadata & Applications

Data Store

External systems

Data Warehouse as a Service

HANA as a Service

S/4HANA

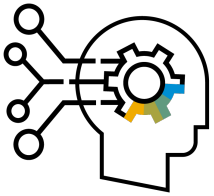**BW/4HANA**

3rd Party Databases

Other solutions, repositories and data stores

SAP Data Intelligence

Comprehensive solution for scaling artificial intelligence, extracting value from distributed data, and embracing open technologies

Scale Artificial Intelligence
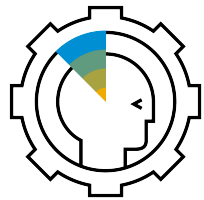Design, deploy, and deliver AI across the enterprise

Extract Value From Distributed Data
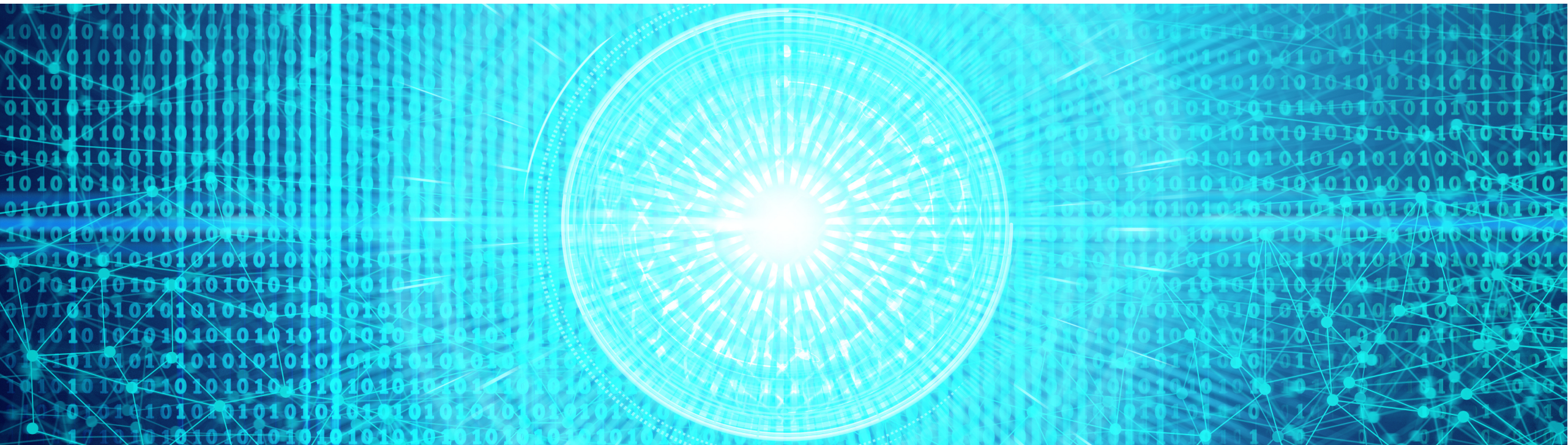Connect data assets and transform them into business insights

Embrace Open Technologies
Take advantage of the latest technologies

Connect with Us and Q&A

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

ASUG

Q&A

For questions after this session, contact contact me at
daniel.rutschmann@sap.com.

ASUG

Let's Be Social.

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

ASUG

Thank you