ASUG
Americas' SAP Users' Group

Accelerate Digital Transformation @ General Mills
Douglas Maltby, Solution Architect, General Mills
Prasanthi Thatavarthy, Sr. Director Dev, T&I Big Data, SAP
Session ID #83511

General Mills and SAP Data Hub

Accelerate Digital Transformation @ General Mills

Session: 83511

About the Speakers

Douglas Maltby

- Solution Architect, General Mills

- Background: At GMI for 13 years with 23 years of SAP experience.

- Fun Facts: 7 marathons, lived in Shanghai for 6 months.

Prasanthi Thatavarthy

- Sr. Director, SAP

- Data Quality & Big Data at SAP for 15 yrs.

- 0 marathons but starting on tiny 5Ks, worked in Yokohama, Japan for 2 yrs. ☺

asug

Key Outcomes/Objectives

1. Share General Mills' EDW Modernization & "Big Data" integration journey to date

2. ETL is easy, Integration is hard

3. SAP Data Hub enables Enterprise "Data Ops"
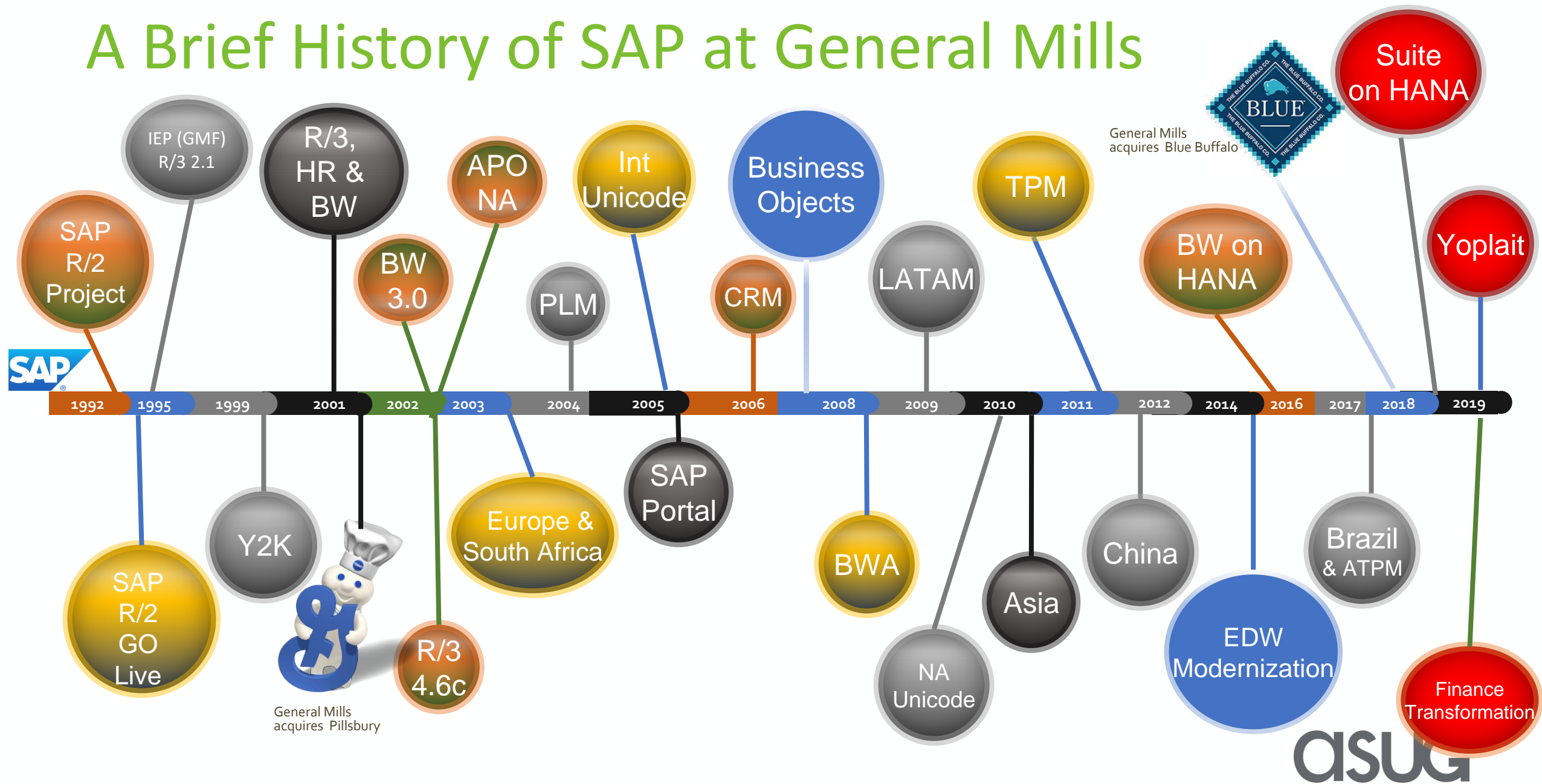
ASUG

Agenda

- Intro: General Mills and our Strategic Data Platforms
- Our Journey with HANA and Hadoop
  - EDW Modernization and the Data Lake
- Enterprise Capabilities Desired
  - Orchestrate and Integrate HANA & Data Lake assets
  - Pipelining: Real-time replication from SAP sources simultaneously to HANA & Data Lake -> a Kafka backbone?
  - Governance/Data Catalog: Enterprise Find > Trust > Connect
  - Data Science ML/AI capabilities
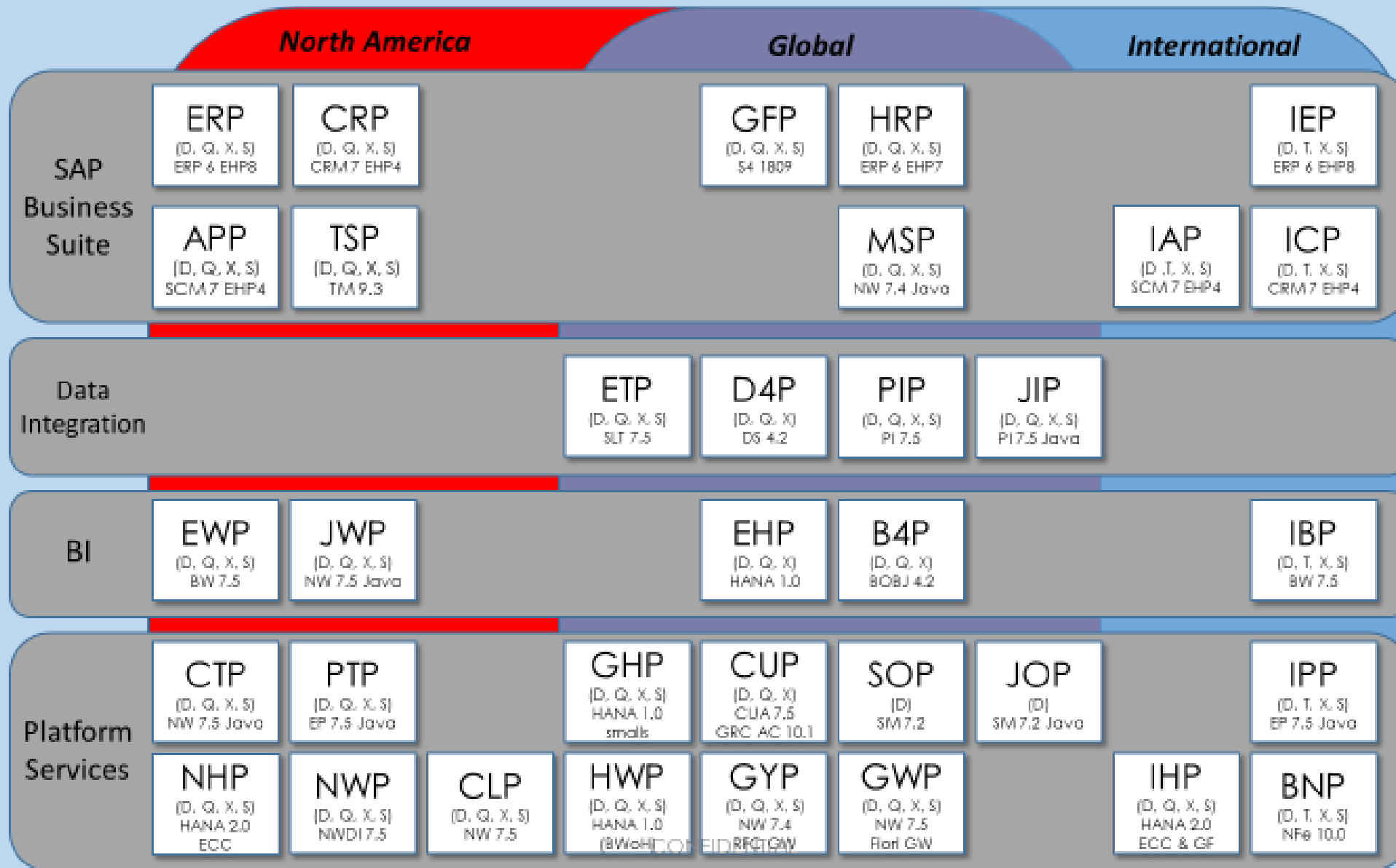- Data Hub Capabilities, GMI PoC, and SAP Roadmap

- One of the world's largest food companies
- Products marketed in more than 100 countries on six continents
- 40,000 employees
- $15.7 billion in fiscal 2018 net sales

A Heritage of Innovation and Brand Building

Betty Crocker created

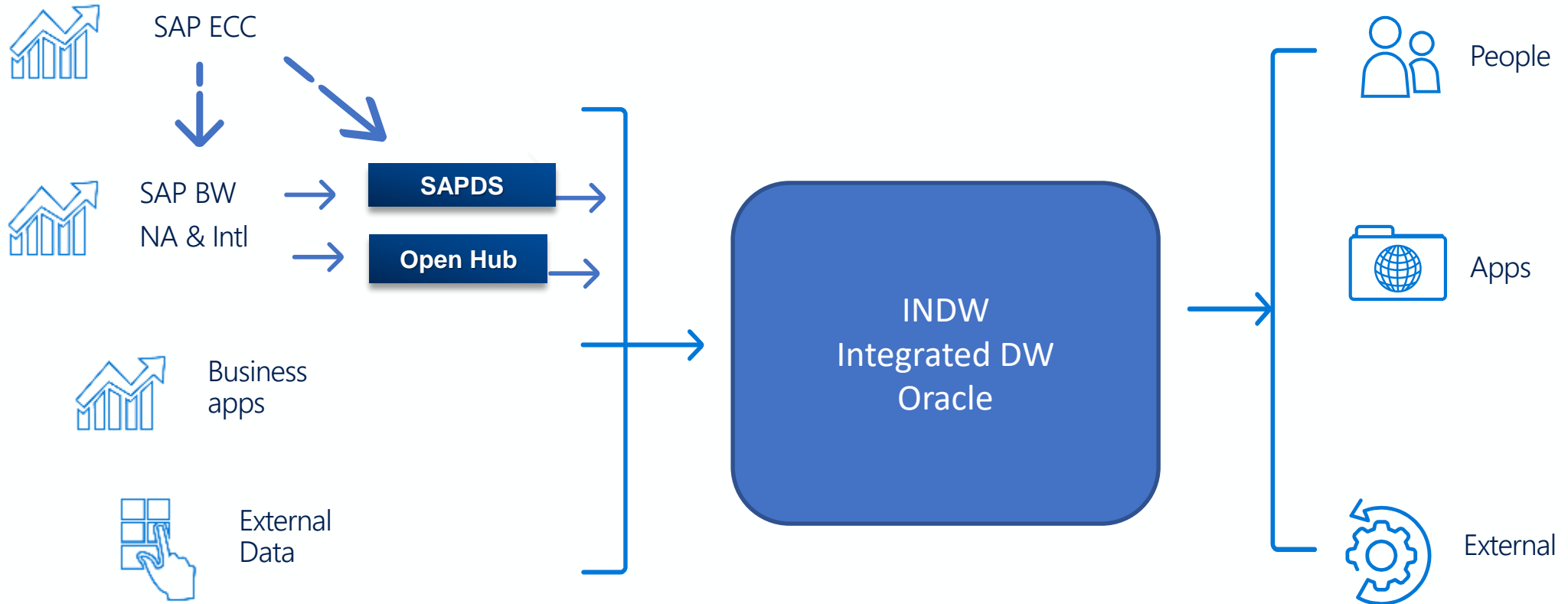U.S. licensing rights to Yoplait are acquired

Cheeri Oats debut

Yoplait acquisition

Cadwallader Washburn builds first flour mill

CPW joint venture launched

| 1866 | 1869 | 1903 | 1921 | 1924 | 1928 | 1941 | 1961 | 1977 | 1984 | 1990 | 2001 | 2011 | 2012 | 2014 | 2016 |

Green Giant founded as Minnesota Valley Canning

General Mills stock trades

James Ford Bell Research Center opens

Häagen-Dazs goes international (Japan)

Yoki acquisition

General Mills acquires Annie's

Charles Pillsbury invests in first Minneapolis mill

Wheaties launches as Whole Wheat Flakes

General Mills acquires Pillsbury

A Brief History of SAP at General Mills

General Mills acquires Blue Buffalo

General Mills acquires Pillsbury

Timeline: 1992 · 1995 · 1999 · 2001 · 2002 · 2003 · 2004 · 2005 · 2006 · 2008 · 2009 · 2010 · 2011 · 2012 · 2014 · 2016 · 2017 · 2018 · 2019

- SAP R/2 Project
- IEP (GMF) R/3 2.1
- R/3, HR & BW
- BW 3.0
- APO NA
- Int Unicode
- Business Objects
- CRM
- TPM
- BW on HANA
- Suite on HANA
- Yoplait
- PLM
- LATAM
- SAP R/2 GO Live
- Y2K
- R/3 4.6c
- Europe & South Africa
- SAP Portal
- BWA
- NA Unicode
- Asia
- China
- EDW Modernization
- Brazil & ATPM
- Finance Transformation

asug

SAP Instance Landscape
Updated Feb 2019

	North America			Global			International

SAP Business Suite

ERP (D, Q, X, S) ERP 6 EHP8
CRP (D, Q, X, S) CRM7 EHP4
GFP (D, Q, X, S) S4 1809
HRP (D, Q, X, S) ERP 6 EHP7
IEP (D, T, X, S) ERP 6 EHP8

APP (D, Q, X, S) SCM7 EHP4
TSP (D, Q, X, S) TM 9.3
MSP (D, Q, X, S) NW 7.4 Java
IAP (D, T, X, S) SCM7 EHP4
ICP (D, T, X, S) CRM7 EHP4

Data Integration

ETP (D, Q, X, S) SLT 7.5
D4P (D, Q, X) DS 4.2
PIP (D, Q, X, S) PI 7.5
JIP (D, Q, X, S) PI 7.5 Java

BI

EWP (D, Q, X, S) BW 7.5
JWP (D, Q, X, S) NW 7.5 Java
EHP (D, Q, X) HANA 1.0
B4P (D, Q, X) BOBJ 4.2
IBP (D, T, X, S) BW 7.5

Platform Services

CTP (D, Q, X, S) NW 7.5 Java
PTP (D, Q, X, S) EP 7.5 Java
GHP (D, Q, X, S) HANA 1.0 smalls
CUP (D, Q, X) CUA 7.5 GRC AC 10.1
SOP (D) SM7.2
JOP (D) SM7.2 Java
IPP (D, T, X, S) EP 7.5 Java

NHP (D, Q, X, S) HANA 2.0 ECC
NWP (D, Q, X, S) NWDI 7.5
CLP (D, Q, X, S) NW 7.5
HWP (D, Q, X, S) HANA 1.0 (BWoH)
GYP (D, Q, X, S) NW 7.4 RFC GW
GWP (D, Q, X, S) NW 7.5 Fiori GW
IHP (D, Q, X, S) HANA 2.0 ECC & GF
BNP (D, T, X, S) NFe 10.0

General Mills Technology Framework

Technology 2020

Purpose

We serve the world by making food people love

Goal

Create market leading growth to deliver top tier shareholder returns

Priorities

Drive More From Core	Fund Our Future	Reshape Portfolio for Growth	Build Advantaged & Agile Organization

Mission

Our #AwesomePeople lead the delivery of connected, secure, trusted technology solutions and capabilities that drive business value

Strategy

Champion technology to drive Consumer first in a mobile world

Lead with Security	Operational Excellence	Drive action through connected data	Simplify the employee Experience

ASUG

Initial State – Enterprise Data

EDW Modernization Program Background

- **Purpose: EDW Modernization and later Agile Big Data Integration**
  - BWA Obsolescence – HP BWA Hardware July 2016
  - TCO: Performance, Agility and Productivity
  - DLM: NLS for TCO and Data Aging – IQ
  - SDA: Data federation and virtualization with key enterprise assets (DSR, Nielsen, Orchestro, etc.)
  - BW4: Positioned for BW/4HANA and Big Data federation and integration

- **Duration: 3 years** for remediation, TDI Setup, BW Upgrades, Migration & Optimization

- **Infrastructure:**
  - **TDI** Scale Out (9+1TB nodes) on HP with RHEL
  - MDC: Int'l (4+1), NA (5+1)

- **BW Versions**:
  - Chose HANA migration in lieu of annual BW upgrade in 2015. DMO not used.
  - Int'l: BW 7.4 SP08   Interim: BW 7.5 SP04   Now: BW 7.5 SP12
  - NA: BW 7.4 SP10   Interim: BW 7.5 SP07   Now: BW 7.5 SP12
  - HANA SPS 11 rev 112.03   Interim: HANA SPS 12 rev 122.06   HANA SPS 12 rev 122.20

asug

EDW Modernization Program

	F'15 (Platform Setup)	F'16 (Platform Migration)	F'17 (Process Optimization)

BOBJ Explorer — BWA>HANA Migration · Explorer Blade Decommission

SAP HANA Platform — SAP HANA TDI & NLS Discovery · SAP HANA Platform Setup

REFR3 — REFR3 Discovery · REFR3 Remediation

BW Remediation — BW on HANA Discovery · BW Remediation · BEx 3.5 EOS · HP BWA EOES · BWA Decommission

NA BW (EWP) — ABAP/JAVA Stack Split · BW 7.40 SP 10 Upgrade · BPC 7.5 EOS · BW on HANA DB Migration · BW 7.5 SP 7 Upgrade · BW on HANA Optimization

Intl BW (IBP) — BW 7.40 SP8 Upgrade · BW on HANA DB Migration · BW 7.5 SP 4 Upgrade · BW on HANA Optimization

We are beyond our original 3yr program plan

Program Projects

Legend:
- Project In Flight (yellow)
- Planned project (dark green)
- Key Constraints (blue)
- completed (green)

Future State – Enterprise Data

SAP BW powered by SAP HANA (SAP BW/4HANA)

Data Storage

Search

Data Lake (Hadoop)

Governance

Advanced Analytics

Business apps

External Data

Sensors and devices

People

Apps

Automated systems

FIND ➤ TRUST ➤ CONNECT
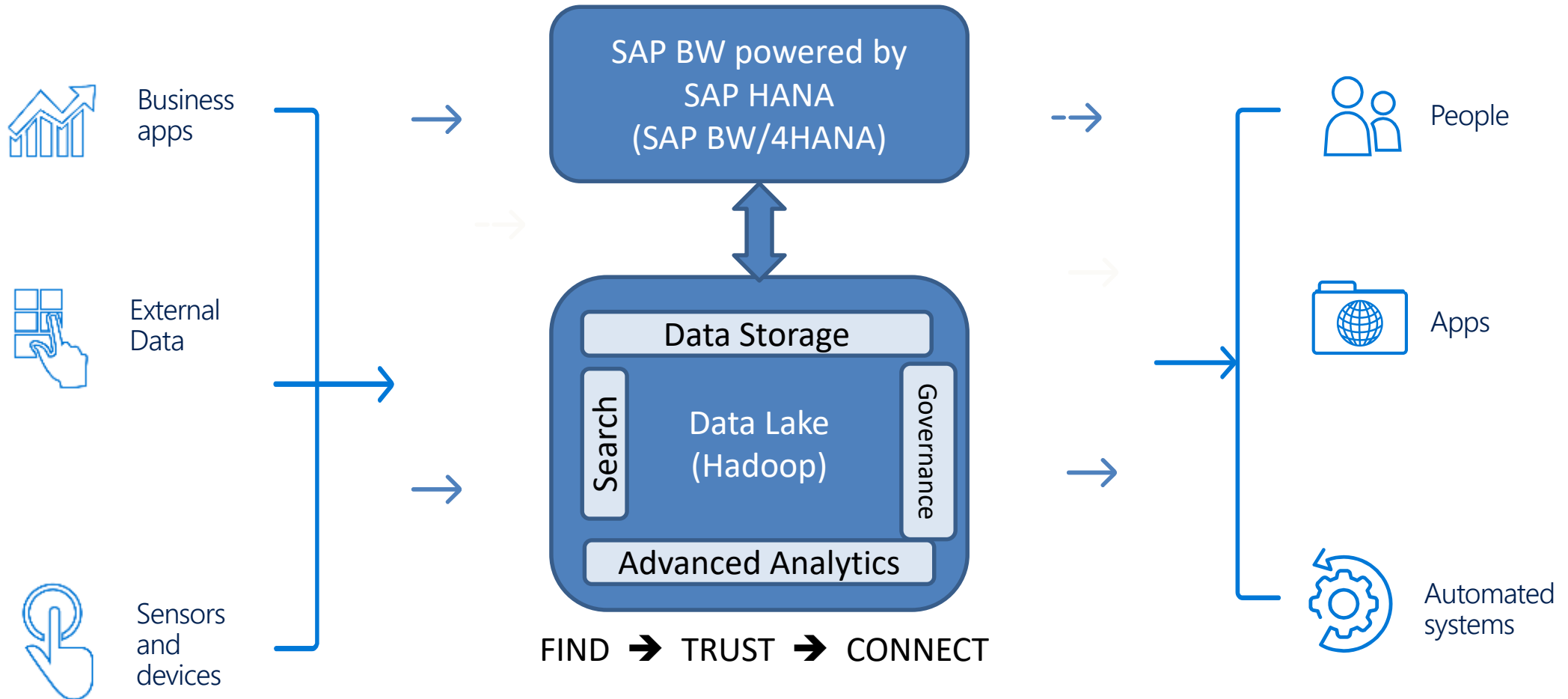
What's different?:
- Prioritized data is available to the lake, so don't need to move it when we need it
- Governance of data lake allows us to find what we need
- Analytic capabilities where the data resides
- Supports unstructured data

ASUG

Big, Connected Data Roadmap

We are here

F16
- **Data Lake Installed**
- **Data Catalog POV Completed**
- **Operational Procedures**
- **Pilot Project**

F17
- **Governance & Standards**
- **Implement Data Catalog**
- **Integrate Master Data**
- **Selected Data Lake projects**
- **Data Science Tools & Processes**
- **Plan DataMart Consolidation**

F18 - F19 - F20
- **Data Lake Open to All**
- **Consolidate & Catalog all DataMarts**
- **Prioritize & "Connect" Data**

Oracle

Datamart Consolidation

Hadoop

Enterprise HANA (Native HANA)

Datamart Consolidation

EIM/SDH/Vora

Analytics on Connected Data

SAP BW on HANA

BW Optimization

SAP BW/4HANA

Find → *Trust* → *Connect*

asuG

Data Lake Technology Landscape
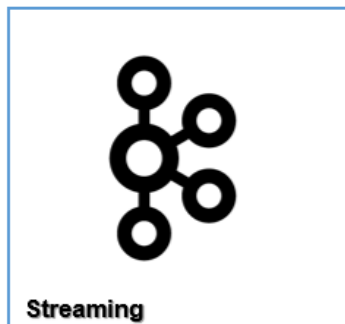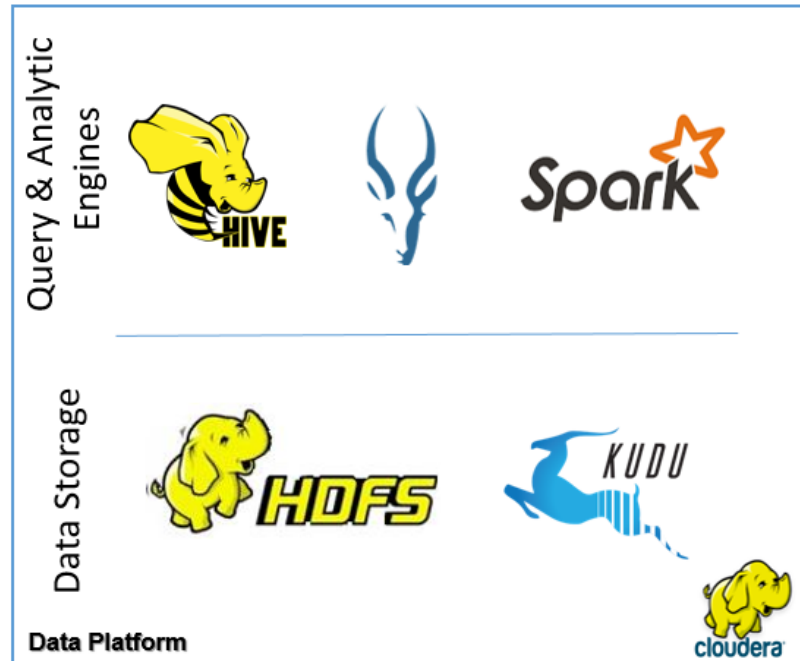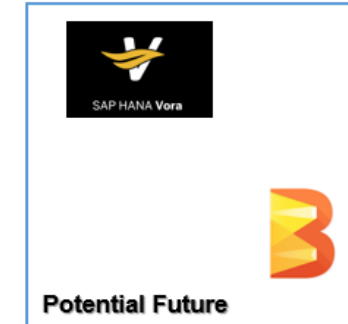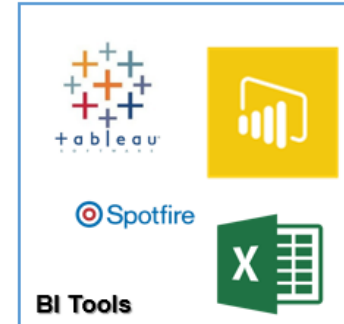
Why GMI needs a Data Catalog

Find
- What data do we have in the Data Lake?
- How do I know if this is the right dataset to use?
- Who owns it and who can I have a conversation with to learn more about it?
- Can the business find this information on their own?

Trust
- How good is the data and what validations do we have in place?
- Has this data been certified?
- Where is it being used and how it is being used?
- Can I contribute my knowledge to the documentation?

Connect
- How do I connect to it?
- Can I connect to it?
- What is the lineage of this data?
- Can this catalog tool connect to my system?
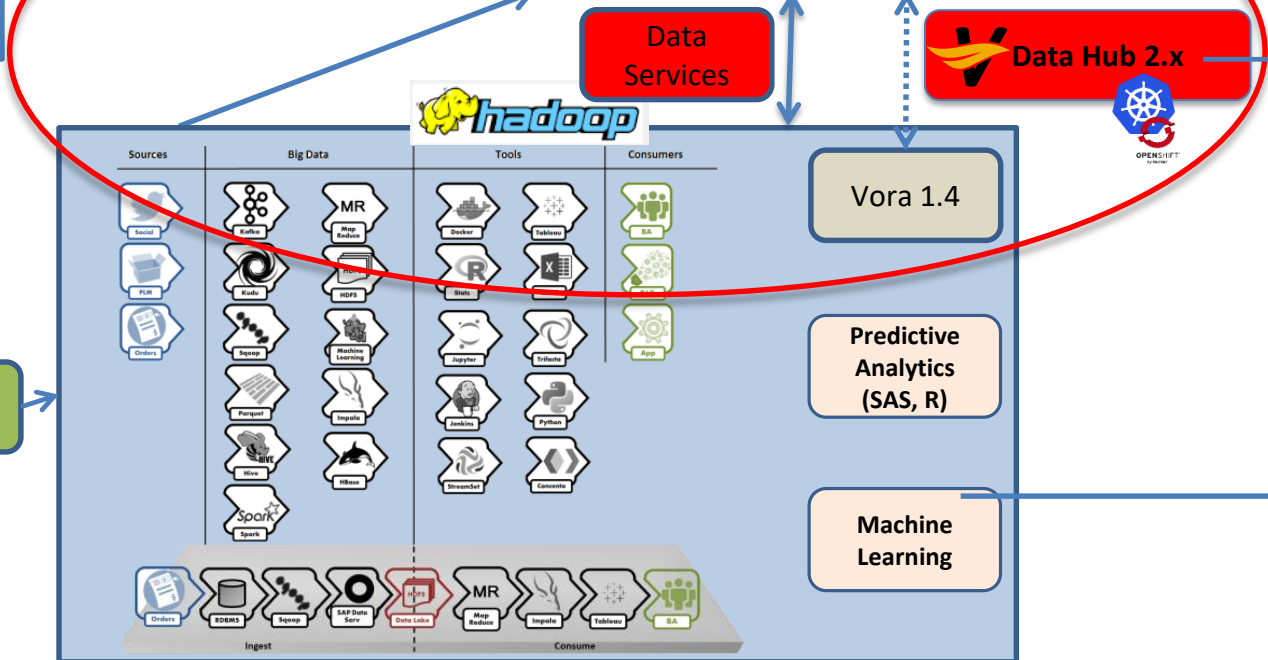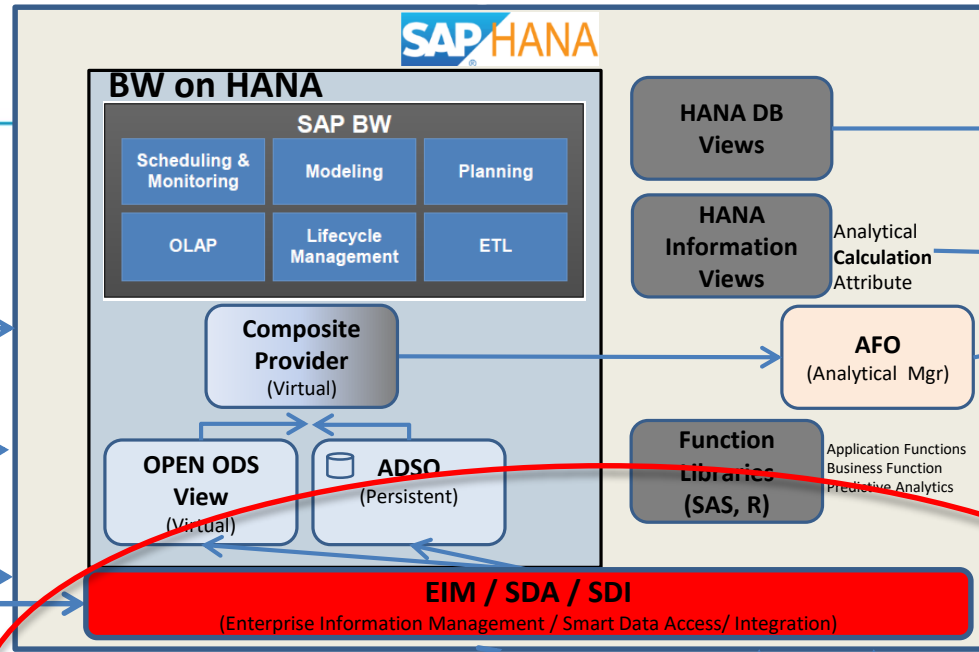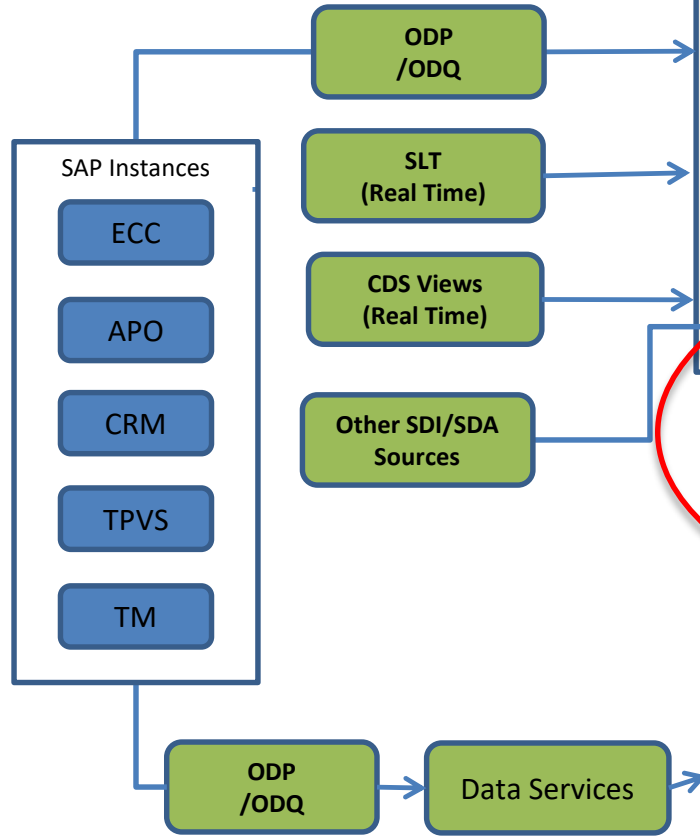
The Intelligent Data Warehouse

THE Data Management platform for the Intelligent Enterprise based on SAP HANA

- Close Integration with S/4HANA & Cloud Applications (SFSF, Ariba,..)

- Schema-less DWH with high automation and minimal Modeling

- Data Lake as primary high volume and computation persistency

- Scalable Storage and Data Processing capabilities in Cloud / On-Premise

- Data Processing beyond OLAP with ML / Predictive Analytics etc.

SAP Cloud Platform — *Intelligent Data Warehouse* — aws

SAP Analytics Cloud

| Self Services | Business Planning | Predictive Analytics |

SAP BW/4HANA | 3rd Party / Open APIs | **SAP Data Hub**

OLAP	Meta Data Governance
Data Modelling	
Process & Orchestration	Data Pipelines
Data Management	

| Massive Data Store | High Volume Compute | Advanced Data Proc. |

Data Lake / Object Stores

SAP HANA Data Management Suite

S4 HANA

asug

Data Science use cases

- SRM – Strategic Revenue Management

- eCommerce

- Sentiment analysis

- ML for demand planning

- ML for financial planning
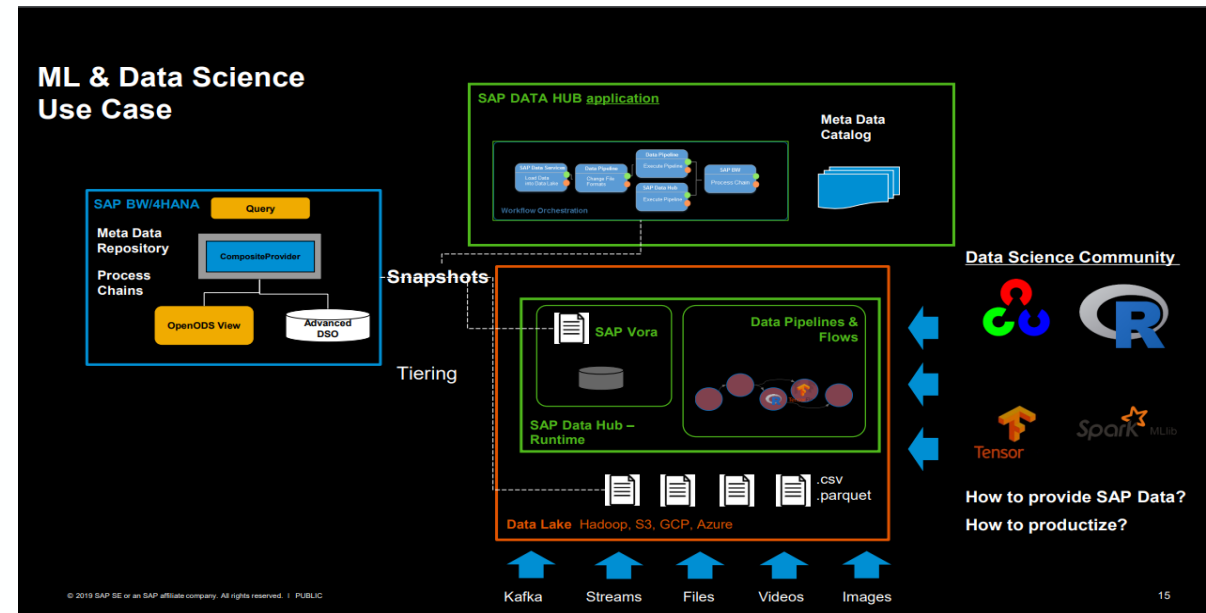
What is SAP Data Hub?
The Lego Analogy
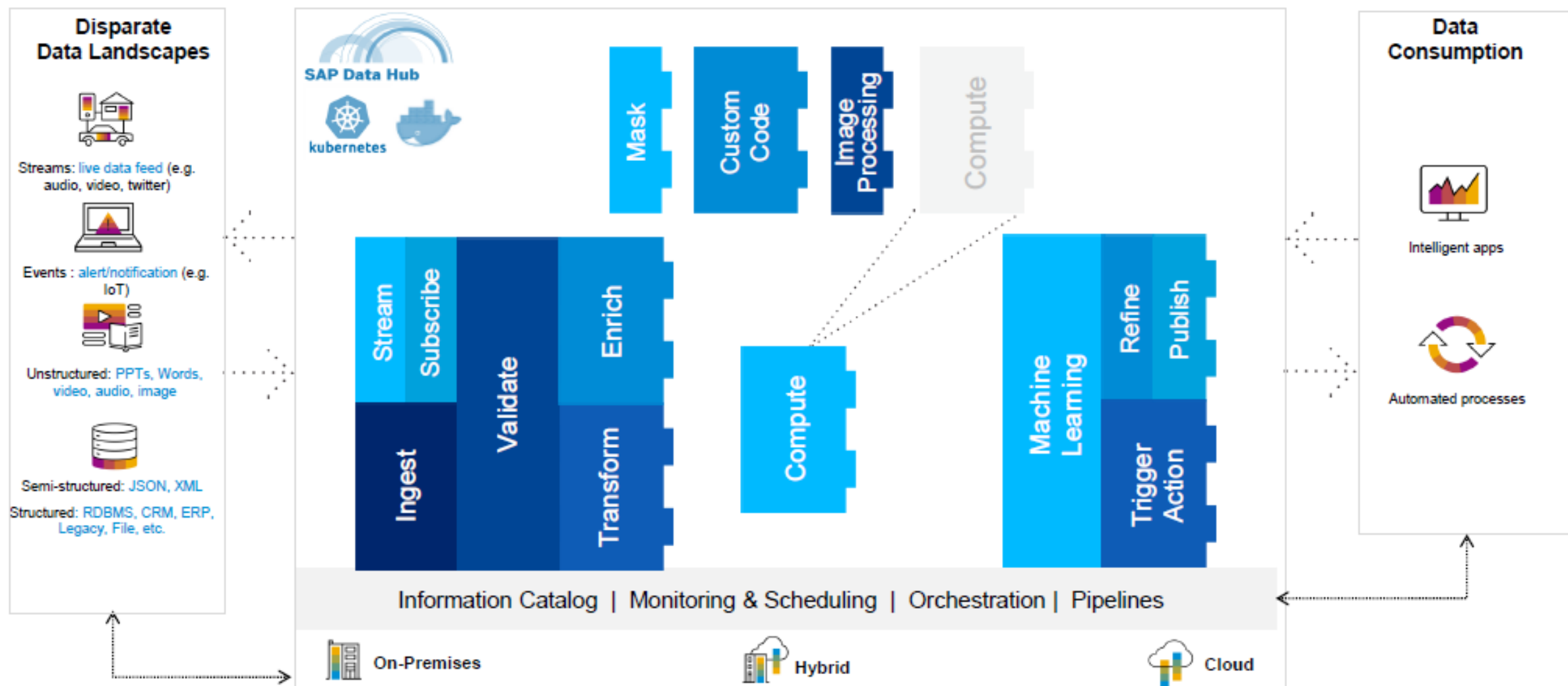
SAP Data Hub Capabilities

SAP Data Hub

Connectivity

Discovery & Profiling

Orchestration & Processing

Scalable Storage

Workspace
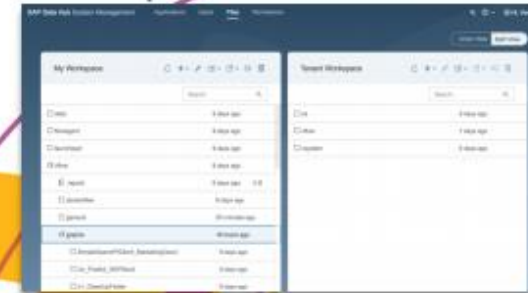
Metadata Governance
SAP Data Hub Metadata Explorer

A centralized location for

browse connections | monitoring | metadata catalog | search datasets | publications | labels

Data Integration & Processing
Replication via SAP LT Replication Server into SAP Vora

Enables direct data replication capability into SAP Vora via SAP LT Replication Server

SAP LT Replication Server is now able to write data out of SAP systems directly into SAP VORA. From there data can be moved to Hadoop or further processed by SAP Data Hub.

For details check note 2647941.
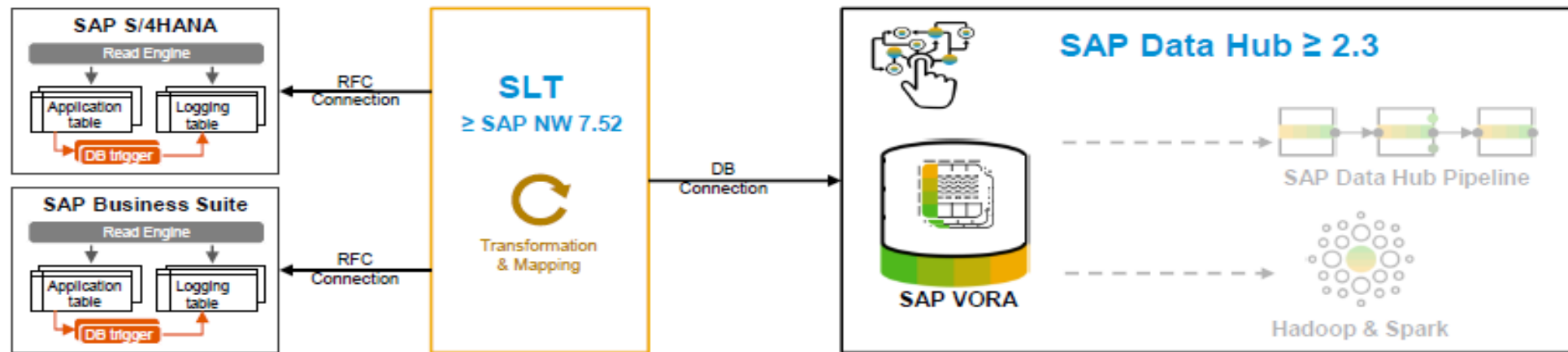
Click here to view the recorded demo

ASUG

SDH: Change Data Capture to Kafka Topic
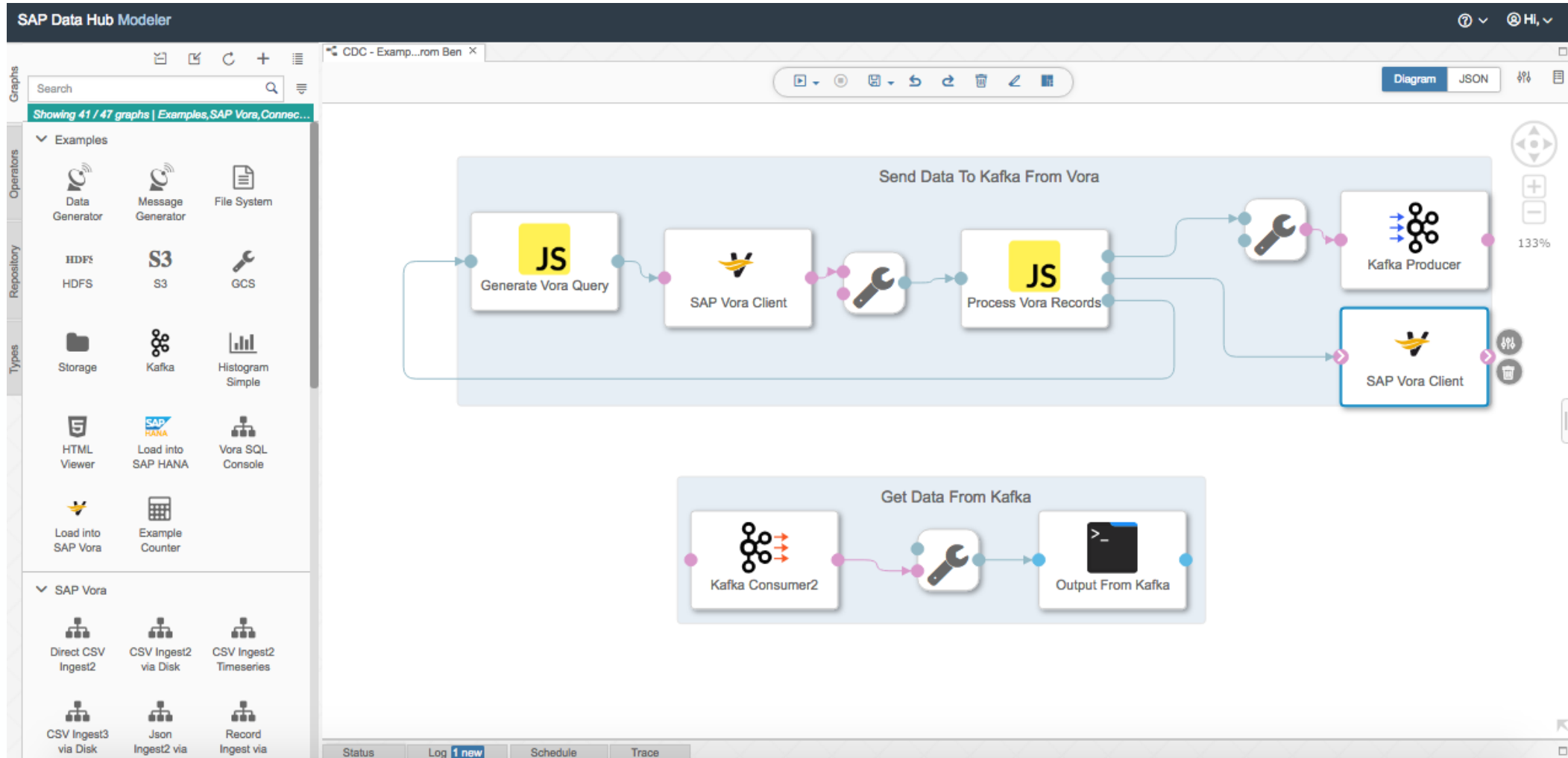
SDH Connection Management:
SLT, BW, HANA, S3, GCS, OData, Vora, etc.

Enterprise Data Capabilities Desired

- Realtime replication from SAP sources (to Kafka)
- OLAP on Hadoop with Excel integration (like BW)
  - Federated OLAP on both HANA and Hadoop (without replication/duplication). Vora?
- Data Catalog – automated, metadata and business-facing governance capabilities. Metadata Explorer?
- Data Science workbench for ML/AI use cases
- Enterprise Scheduling integration
- HA HDFS namenode for failover

ɑsuG

Key Issues for GMI

- Kerberized HDFS – not supported in SDH 2.3

- Metadata Explorer (Data Catalog) in SDH 2.3 primarily for developers, not business users

- Replication from ECC to Kafka via SLT SDK better solution than adding a "hop" in Vora within SDH

- Cost / Value for individual projects vs enterprise perspective

asUG

Where do we go from here?

- **Finish BW on HANA Optimization for BW/4HANA conversion**
    - Hardware refresh in 2019
    - SDI: Redevelop Open Hubs
    - Real-time data enabling capabilities

- **Evaluate SAP Data Hub 2.5+ capabilities**
    - Data Discovery & Governance for Business (Profiling, Catalog, Lineage, Impact Analysis, ADP)
    - Federation and enterprise OLAP use cases (Intelligent Data Warehouse)
    - Data Science capabilities

- **Determine Cloud Strategy**
    - Provider capabilities, delivery and deployment via containers, object stores, S/4 and BW/4, Hadoop, etc.

- **Continue to communicate** with and influence SAP Data Hub Product Management team

Connecting customers with products and innovations from SAP

SAP Customer Engagement Initiative (CEI)
Discuss planned functionality with customers and partners to gain valuable input during development phase

SAP Beta Testing
Customers can experience new SAP software before official release for early prototyping with customer specific data and processes

SAP Customer Connection
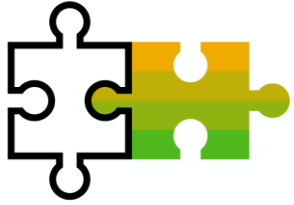Focused projects to decide on customer improvement requests for maintenance products

SAP Continuous Influence
Continuous collection and delivery of improvements for high-growth products
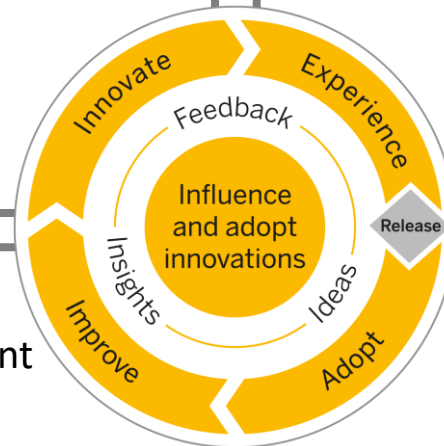
SAP Early Adopter Care
Engage in customer implementation projects to minimize project risk and support successful deployments of new SAP products

Innovate · Experience · Adopt · Improve · Insights · Ideas · Feedback · Release

Influence and adopt innovations

Visit influence.sap.com for all Influencing Opportunities

Key enhancements in SAP Data Hub 2.5

Enterprise Application Integration

- Standardized interface to integrate SAP Cloud solutions
- Integration model to consume and interact with SAP S/4HANA & SAP Business Suite systems
- Orchestration of SAP Cloud Platform integration to interact with processes

Meta Data & Data Excellence

- Meta data extraction for SAP S/4HANA & SAP Business Suite systems
- Embedded self-service for data preparation & quality assurance
- Extend anonymization capabilities included in pipeline development model
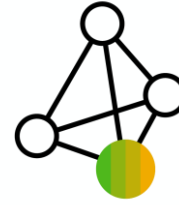
Connectivity & Processing

- SQL processing for files
- Node.js as executable environment embedded
- Visualization concept to support pipeline and application development
- Leveraging SAP Cloud Platform Open Connectors to consume external sources

Deployment & Operations

- Simplified installation process
- Connectivity for High Availability (HA) setups
- Kerberos support for Hadoop services

Enterprise Application Integration
Standardized interface to integrate SAP Cloud solutions

One SAP Cloud Data Integration (CDI) for integration into all SAP Cloud solutions

SAP Analytics Cloud

SAP BW/4HANA	SAP Data Hub

Cloud Data Integration API

Intelligent Suite

SAP Fieldglass
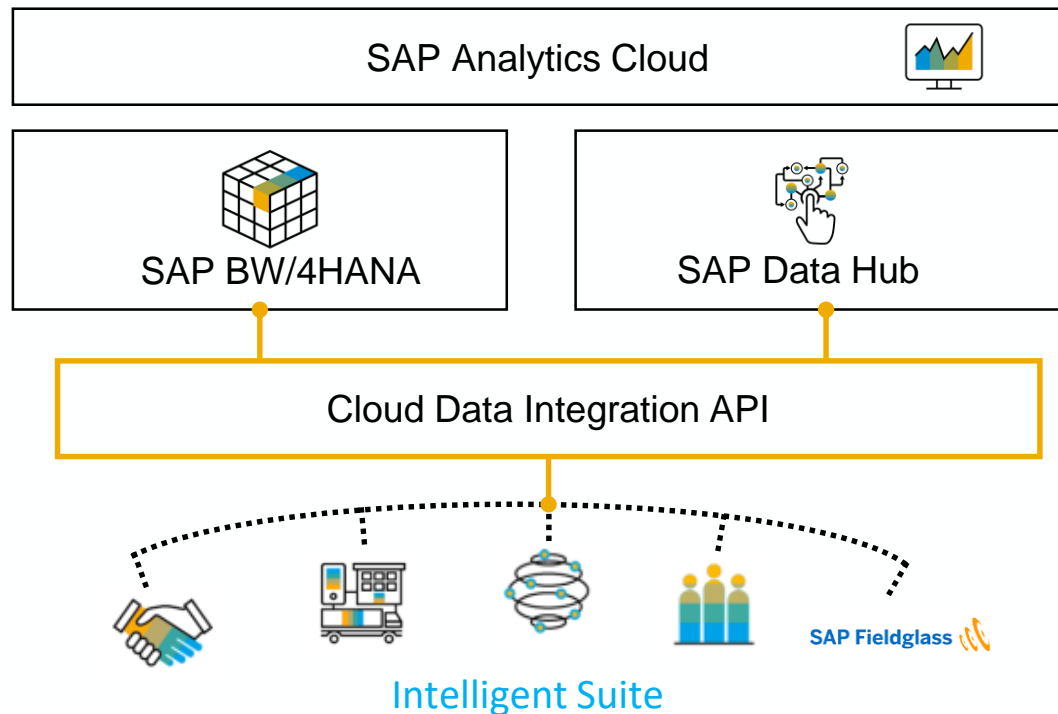
Main Use Cases

- Holistic Data Management with the SAP Data Hub
- Building Data Warehouse Analytics with SAP BW/4HANA
- Advanced Analytics and Planning with SAP Analytics Cloud

Capabilities

- Support both full and delta requests
- Seamless integration of data and metadata
- OData v4 as communication protocol

SAP Cloud Data Integration (CDI) is a
core concept that all SAP solutions in the cloud using
ONE API based on open standards

Currently state SAP Fieldglass implemented CDI, other solutions will follow

asUG

Enterprise Application Integration
Integration model to consume and interact with SAP S/4HANA and SAP Business Suite systems

One model to consolidates all interaction scenarios between SAP Data Hub and an ABAP-based SAP systems directional and bi-directional
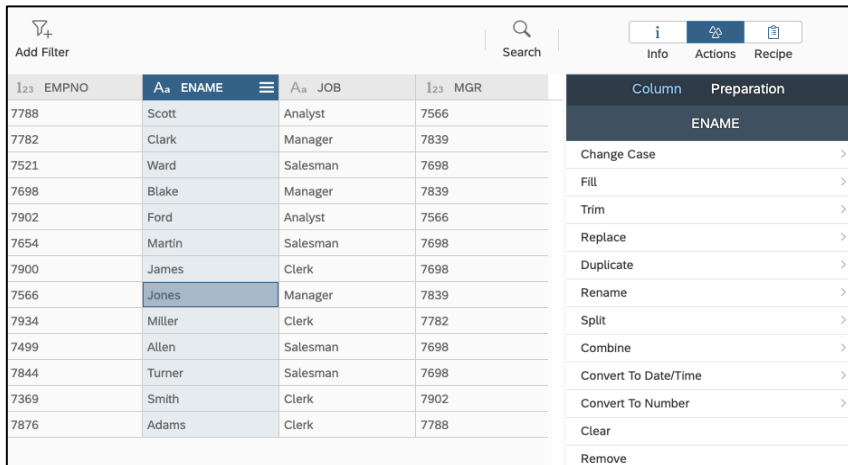
SAP Business Warehouse

SAP S/4 HANA

SAP Business Suite

SAP S/4 HANA

SAP Data Hub

Capabilities

Provide **ABAP meta data** to the SAP Data Hub Meta Data Explorer

ABAP functional execution that is triggerable as a SAP Data Hub Operator

ABAP data provisioning to transfer data into SAP Data Hub
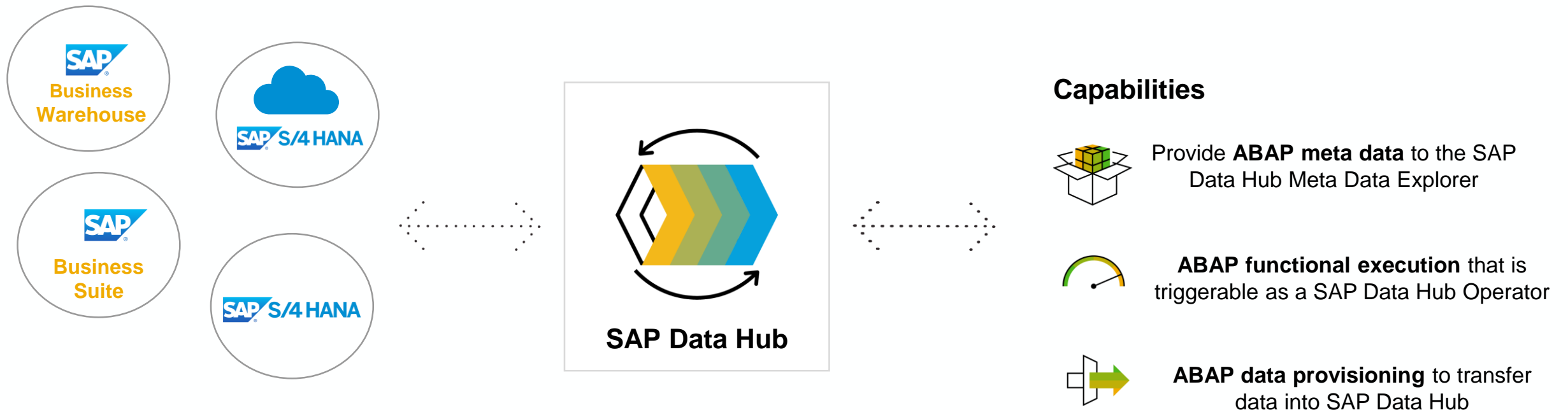
Integration requires certain system level, planned at least SAP S/4HANA 1909, SAP S/4HANA cloud 1908, SAP NetWeaver 7.00 with DMIS 2011/2018 Q4/2019 version. Certain functionality can only be made available for certain release levels.

aSUG

Meta Data & Data Excellence
Embedded self-service for data preparation & quality assurance

Self-service and data-driven data preparation for business users

Main Use Cases

- End-to-end self-service data preparation

- Improve data quality to achieve data excellence

- Create new data sets based for scenario and project requirements

Capabilities

- Access a dataset to prepare the data based on a sample dataset

- Transform, shape, harmonize, curate, enrich the data by a simple click

- Profile, assess, transform, shape, and enrich the data

- View, present and report the outcome immediately

Connectivity & Processing
SQL processing for files

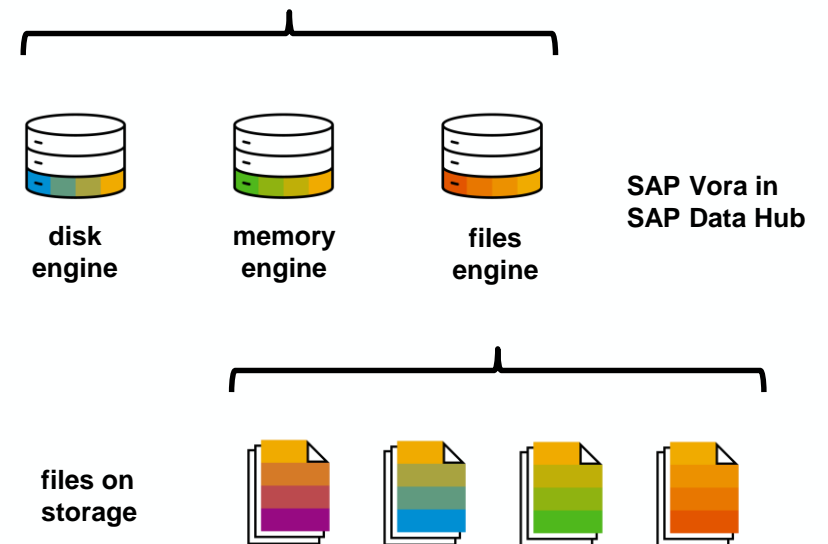## Easy development of combined query execution on different source in SAP Vora

Main Use Cases

- Data mostly stored in data lakes and object stores
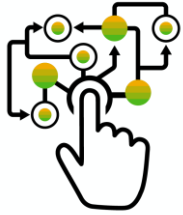- Avoid uneconomic loads of all data into database tables

Capabilities

- Loading of the relevant parts of files during query execution
- Combined queries on disk, memory and files possible
- Seamless embedded in SAP Vora Tools editor

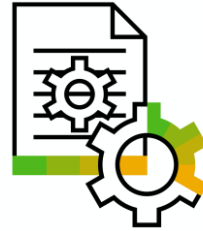SQL query on disk, memory, files or a combination

disk engine   **memory engine**   **files engine**

SAP Vora in SAP Data Hub

files on storage

asuG

Deployment & Operations

Simplified installation process

- Enabling customers with the ability to install and run SAP Data Hub offline (without internet access)
- Improvements in performance, reliability and security

Connectivity for High Availability (HA) setups

- Added HA configuration option in Connection Management (HDFS and HANA)
- HA support for HDFS checkpoint store
- Enabled Spark code generation to utilize HDFS HA config
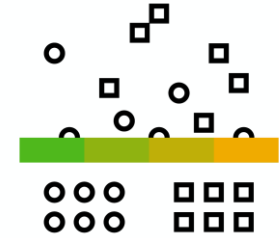
Kerberos support for Hadoop services

- All Data Hub components will support Kerberos authentication when accessing to an external Hadoop Services or HDFS
- For example full support for SAP Big Data Services clusters with Kerberos enabled

SAP Data Hub
Product road map overview – Key innovations

Recent innovations	2019 – Planned innovations[1]	2020 – Product direction[1]	2021 – Product vision[1]

SAP Data Hub in the cloud
- Deployments on Amazon Web Services, Microsoft Azure, Google Cloud Platform
- Supporting Big Data services from SAP as storage

Metadata governance
- Metadata catalog and search
- Visual data lineage for catalog objects
- Data profiles with business rules
- Semantical data extraction for SAP systems (such as SAP S/4HANA, SAP ERP)

Data pipelining (distributed runtime)
- Embedded ML: Tensor flow, Spark ML, Python, R, integration with SAP Leonardo Machine Learning Foundation
- Data snapshots for SAP BW/4HANA and SAP HANA
- Predefined anonymization, data masking, and data quality operations
- Integrated diagnostic framework

Application integration and content
- Release of first SAP Data Hub–based industry applications – such as total workforce insights
- Enhanced connectivity such as DB2, MS SQL Server, MySQL, Google Big Query

SAP Data Hub in the cloud
- SAP Data Hub as managed service on SAP Cloud Platform
- Further certifications for Huawei, Alibaba Cloud

Metadata governance
- Extract SAP semantics (such as business objects)
- Business terms and glossary
- Integration with SAP Information Steward
- Embedded data preparation capabilities

Data pipelining (distributed runtime)
- Agnostic multi-cloud processing
- SQL processing for data streams
- Create visual data pipeline applications

Application integration and content
- Data and meta data extraction for all SAP cloud solutions (such as SAP Fieldglass and SAP Concur solutions)
- Model, deploy, and push down processing logic to ABAP-based systems
- CDC for core data services views right at the source

Foundation for data science and ML
- Pipelines to prepare, training, interfere, and validate with built-in support for standard SAP and non-SAP data sources and algorithms
- Notebook integration out of the box

SAP Data Hub in the cloud
- Further data center and cloud provider availability

Metadata governance
- Team collaboration with social mechanisms (votes, likes, shares, and more)
- Information policy management compliance dashboard
- Self-learning metadata management
- Semantical data extraction for SAP systems (such as SAP S/4HANA, SAP ERP)

Data pipelining (distributed runtime)
- Embed ML-based operations and processes (automated-curation, generate new data, missing value suggestion, and more)
- Suggest complementary dataset to the ones currently considered by users
- Proactive tuning and self-correcting

Application integration and content
- Expand native connectivity driven by market
- Provide templates and predefined/extendable content for on-premise and cloud industry models and applications
- Predefined partner content delivery

Enable the data-driven enterprise
- Enable data-driven and completely automated (Big) Data enterprise applications
- Support new application paradigms
- Enable a simple, holistic data management view

Evolution of enterprise information management
- Unify existing capabilities
- Simplify data integration portfolio
- Comprehensive landscape management

End-to-end business application and processes
- Delivery of applications for business scenarios and industry use cases

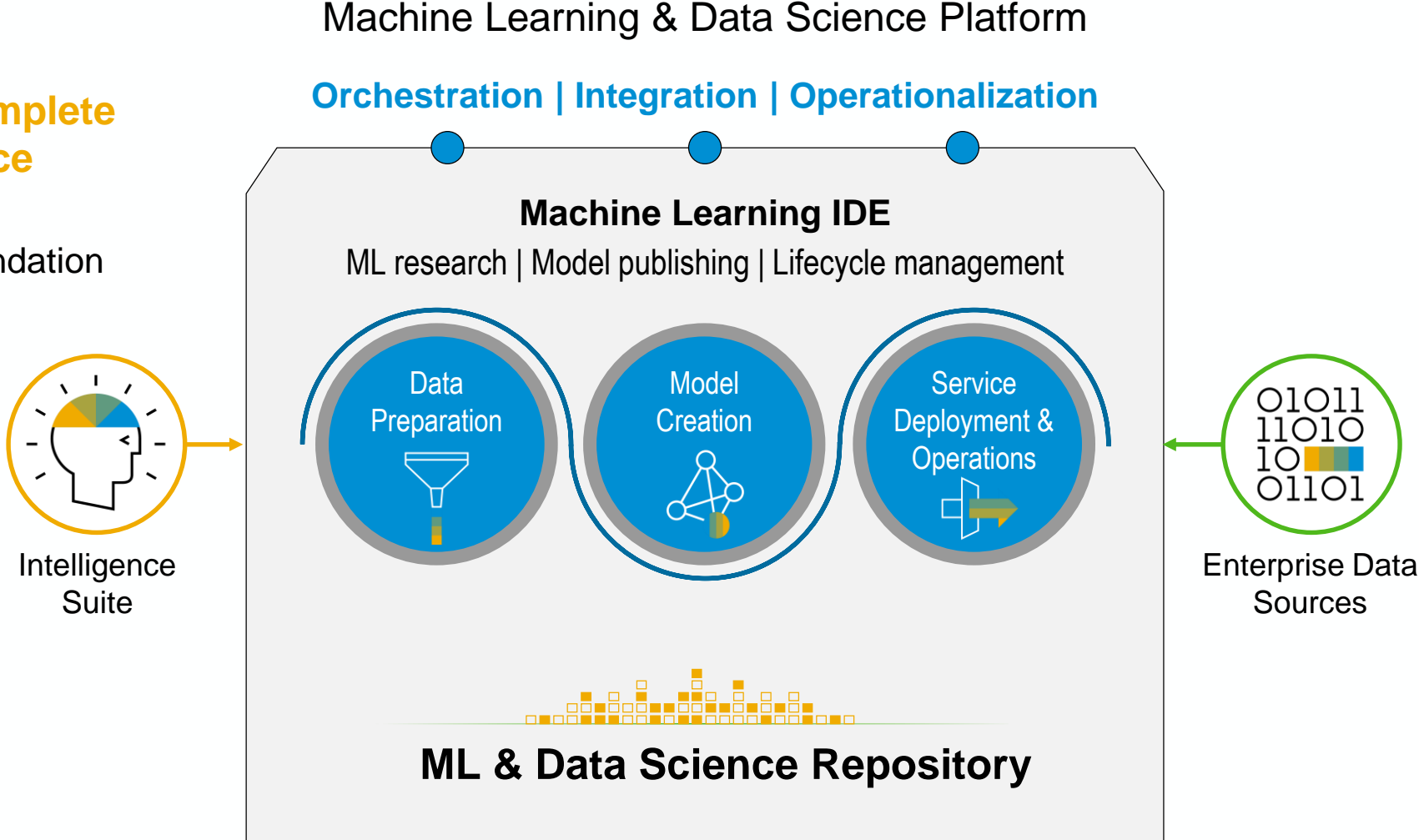# SAP Data Intelligence – Data Science Platform & SAP Data Hub aaS

Data science, machine learning, and data orchestration

Currently in Beta

Goal / Vision

Machine Learning & Data Science Platform

Building an open, scalable, complete machine learning & data science platform

- SAP Data Hub as a Service as foundation and flexible execution environment
- Tight integration into existing machine learning services
- Manage thousands of models in production
- Automate retraining, maintenance, and retirement
- Embed into SAP applications
- Stay compliant and auditable

Orchestration | Integration | Operationalization

Machine Learning IDE

ML research | Model publishing | Lifecycle management

Data Preparation

Model Creation

Service Deployment & Operations

Intelligence Suite

Enterprise Data Sources

ML & Data Science Repository

asug

Questions

Contact Information

Doug Maltby

Douglas.Maltby@genmills.com

Prasanthi Thatavarthy

Prasanthi.Thatavarthy@sap.com

ASUG

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
Douglas.Maltby@genmills.com & Prasanthi.Thatavarthy@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

SAP Data Hub – Additional Resources

- Data Hub 2.4 References
  - SDH Landing page: https://www.sap.com/products/data-hub.html
  - Help: https://help.sap.com/viewer/product/SAP_DATA_HUB/2.4.latest/en-US
  - Product Availability Matrix (PAM)
  - Data Hub 2.4 Central Release Note
  - Data Hub 2.3 Metadata Explorer Demo
- Tutorials, Blogs and Courses
  - What's New in SAP Data Hub 2.4
  - Tutorials: https://developers.sap.com/topics/data-hub.html#tutorials
  - OpenSAP:
    - Freedom of Data with SAP Data Hub: https://open.sap.com/courses/hub1
    - Modern Data Warehousing with SAP BW/4HANA: https://open.sap.com/courses/bw4h2
  - HANA Academy: Data Hub 2.3 Playlist
  - Cloud Appliance Library (CAL) SDH 2.4 Trial Edition: Now Available
  - Developer Edition: https://blogs.sap.com/2017/12/06/sap-data-hub-developer-edition/
  - TechEd 2018 Sessions
    - DAT361 – Model Data Ingestion Pipelines with SAP Data Hub
    - DAT261 – Discover Data, Prepare Data, and Manage Metadata with SAP Data Hub
    - DAT263 – Orchestrate Big Data Landscapes with SAP Data Hub

ASUG