Data Driven Innovation:

Modern Data Platform as Foundation for Transformation and New Services

Karsten Haldenwang, SAP

Public

September 2017

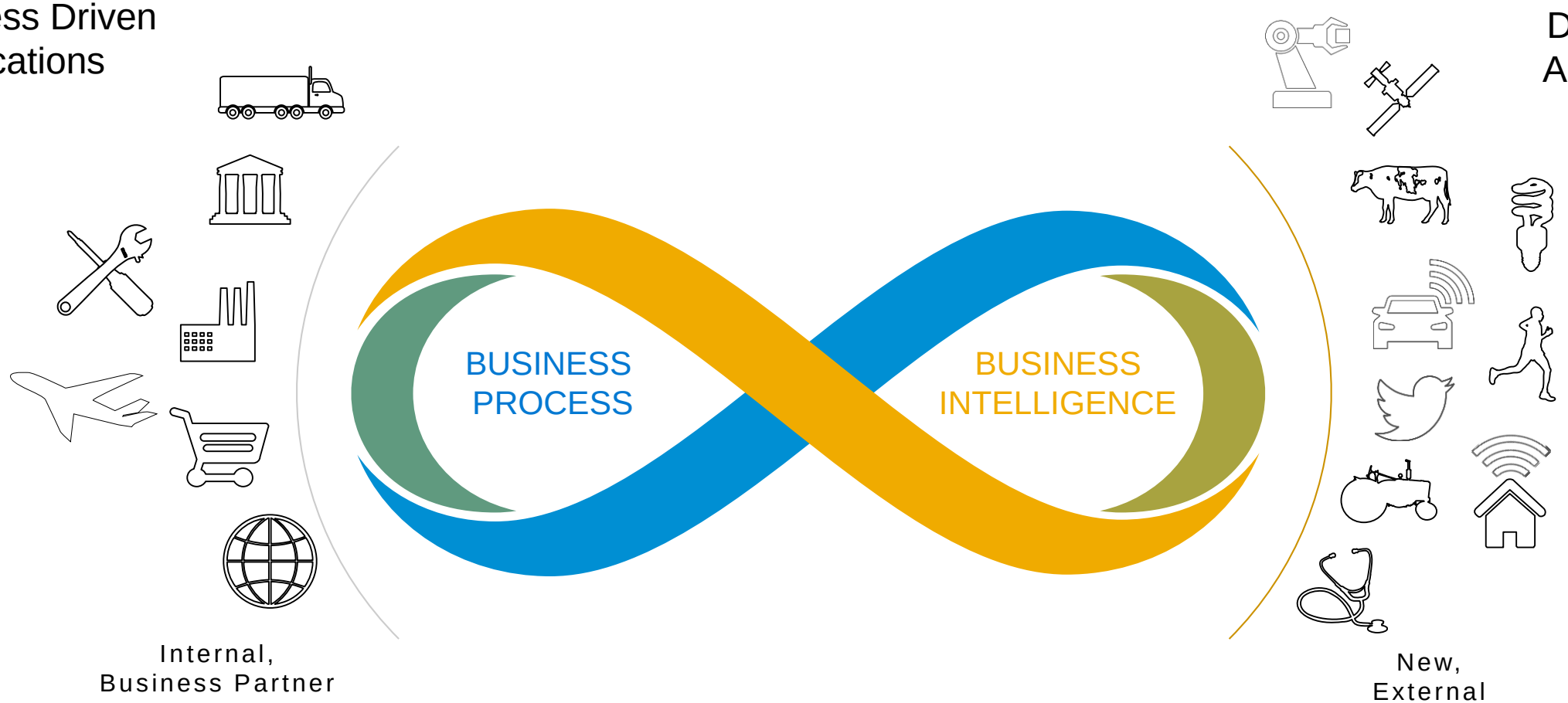
SAP Run Simple

DATA
Connected Data as a Strategic Asset

From Connected Data to Insight to Action

Process Driven Applications

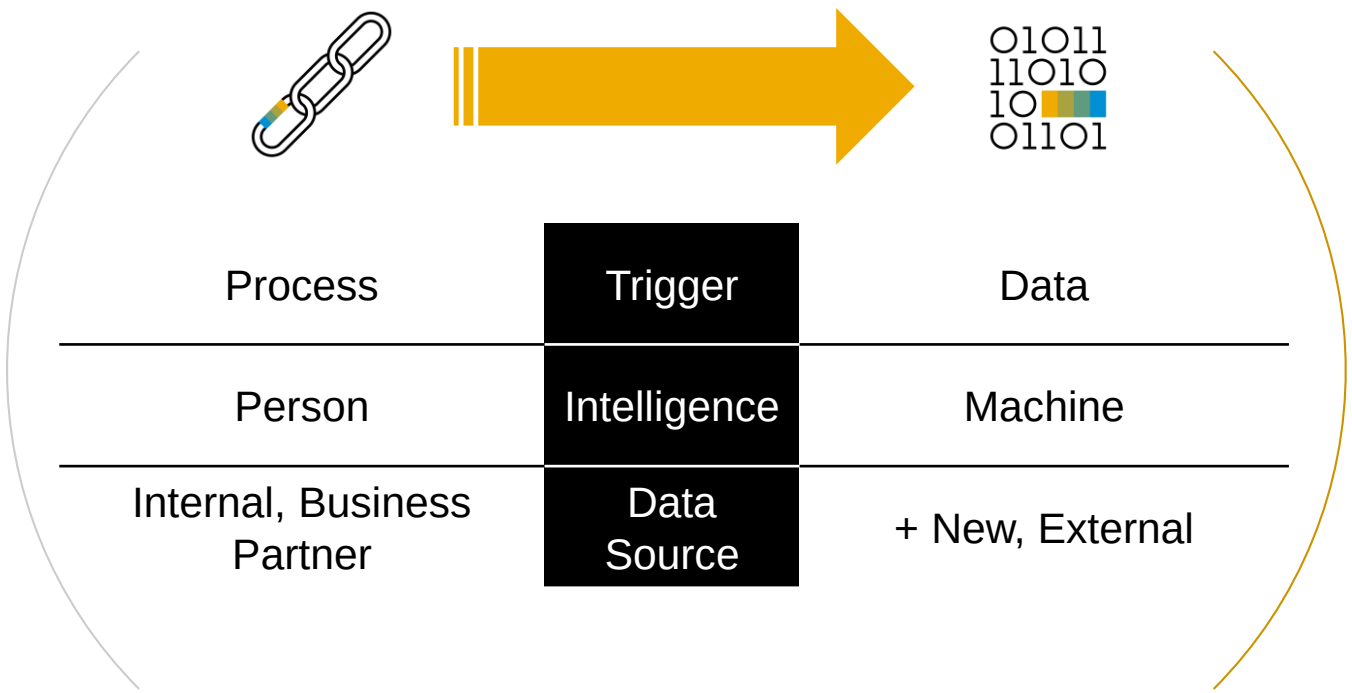
Data Driven Applications



© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

From Data to Data Driven Applications

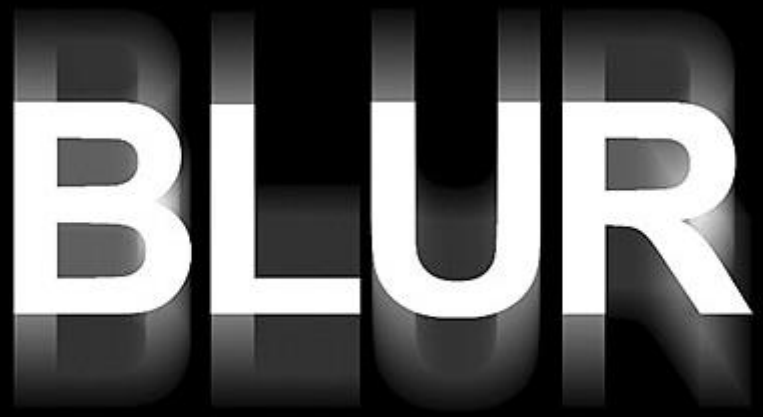
Process Driven Applications



Data Driven Applications

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

DIGITAL / PHYSICAL

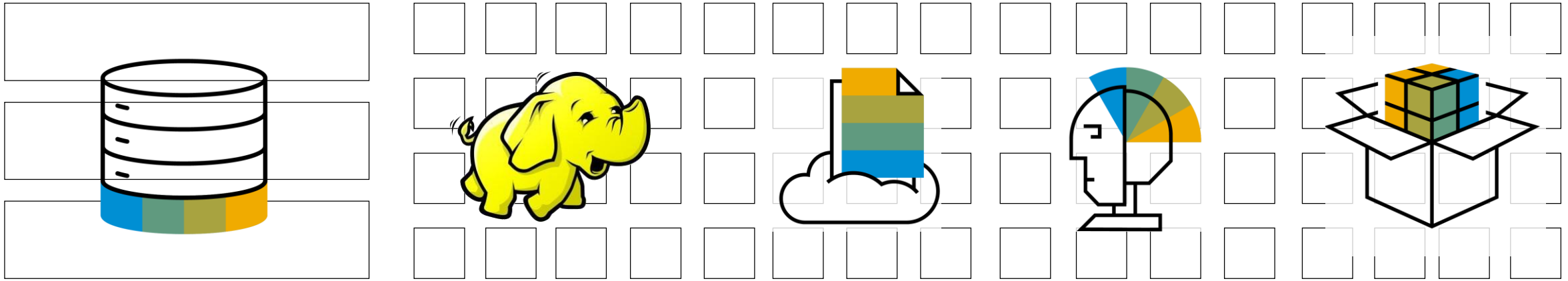


© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

New Challenges New Technologies New Challenges

New challanges require new technologies

Distributed systems in a distributed landscape



Existing Systems

Hadoop/Spark/
NoSQL

Cloud Storage
(i.e. AWS S3)

Machine
Learning
(Python, Spark,
Tensorflow)

Containers
(Kubernetes, Docker)

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

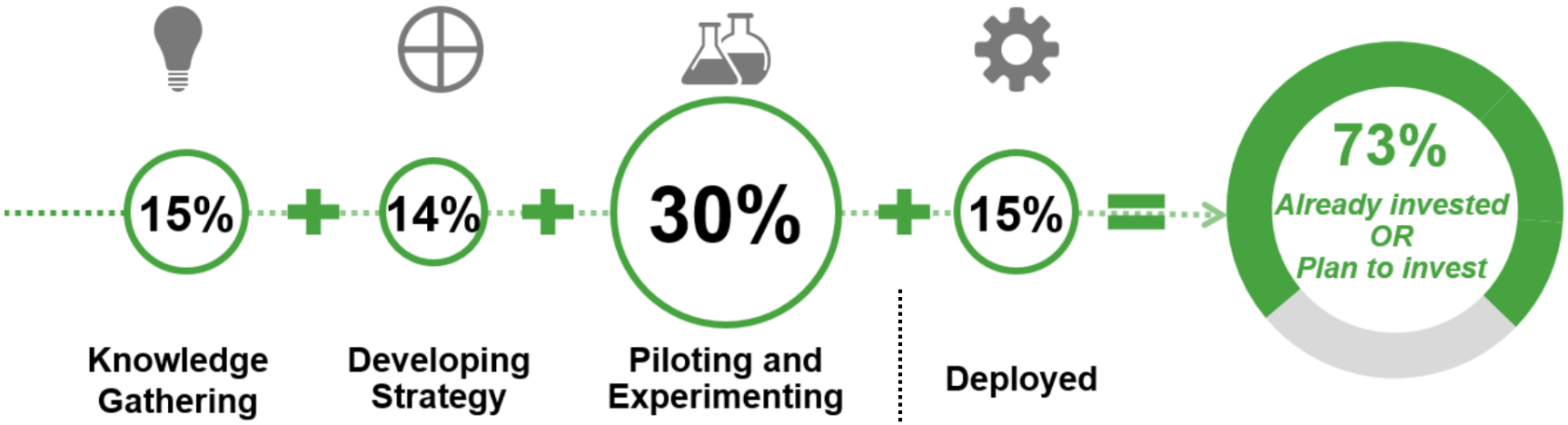
Data volume is NOT the problem anymore

> **Delivering connected data across on-premises and cloud sources is not trivial**, especially when it involves large data volumes, complex data models, and high speed of ingestion.”

Noel Yuhanna,
Forrester

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Gartner: „Still challenging to get Hadoop projects into production stage“



(1) Skills

(2) Technology Integration Challenges

(3) Determining Value

Gartner, 2017 (ID: G00310700)

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Supporting Data Innovation Projects…

From Initial Pilot to Large Scale Success

PIONEER

Show possibilities

Inspire business

Get buy-in

AUTOMATE

Automate & orchestrate

Schedule & monitor

Document

SCALE

Cloud deployment

Governance

Security

USER ACCEPTANCE

Self-Service Data Preparation

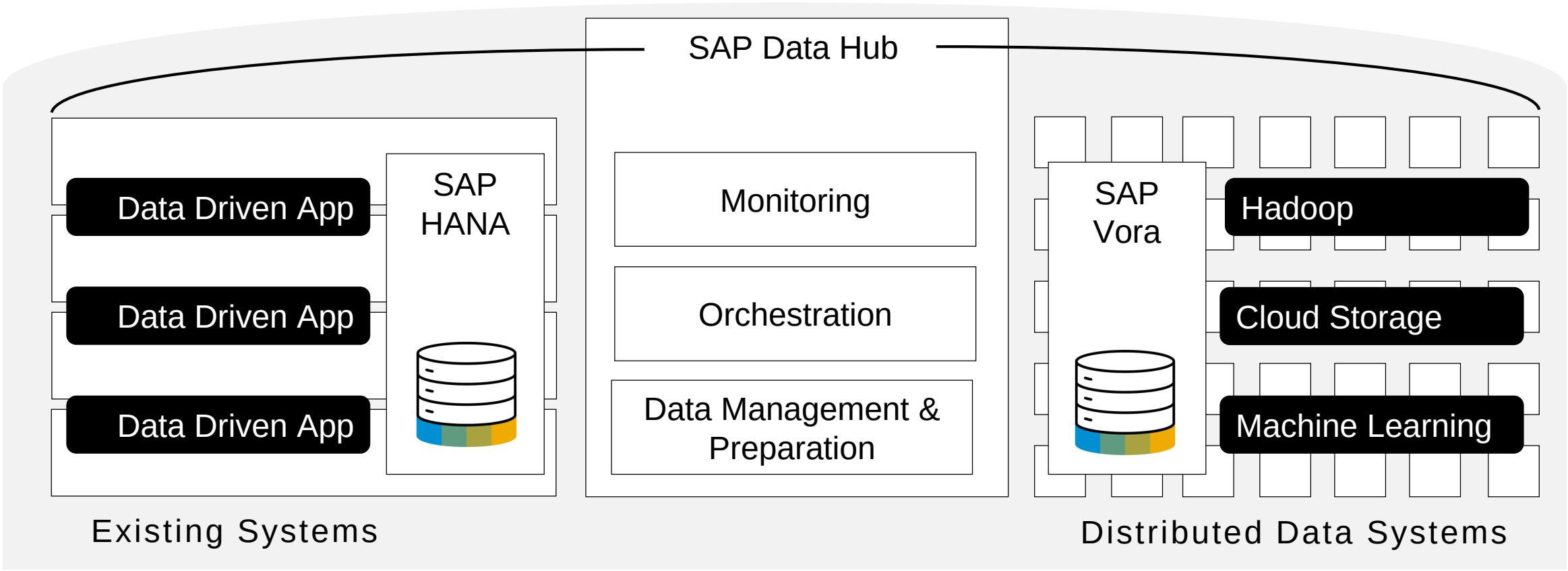
Data driven apps

Use your own tools & languages

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Platform for Data Driven Applications

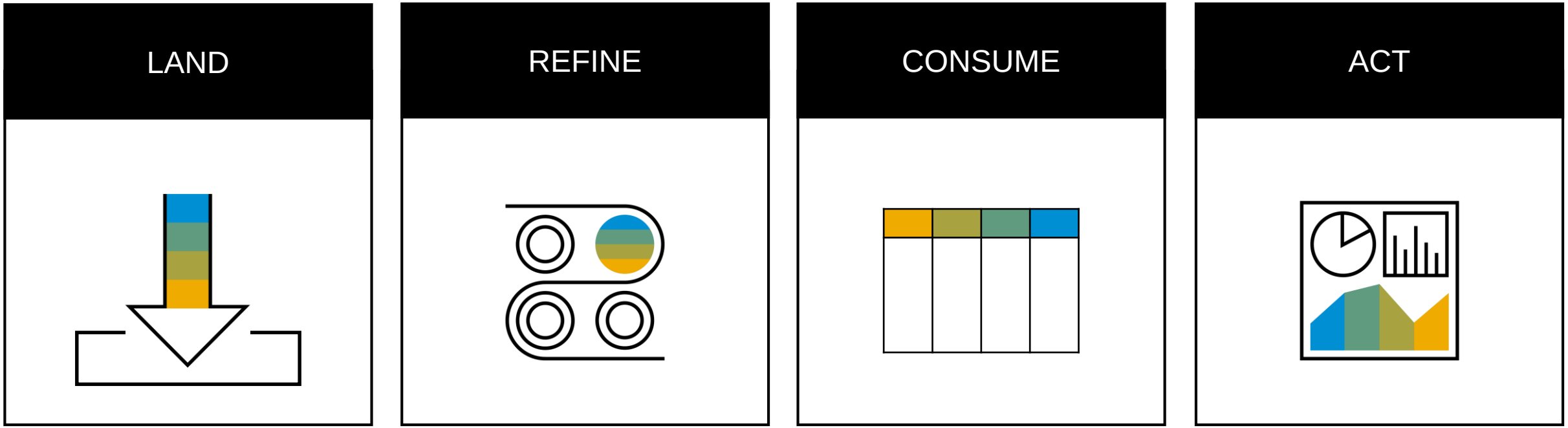
Simplify: Existing Skills, Technology Integration, Stable Operations



SAP Cloud Platform

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

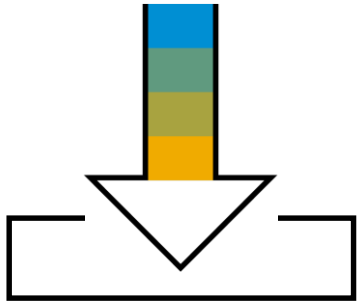
From Data to Data Driven Applications

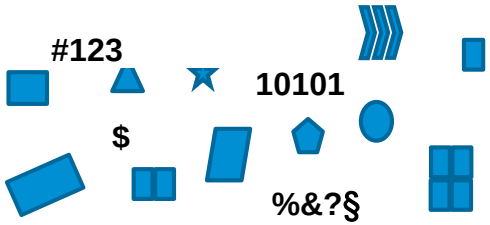


© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

From Data to Data Driven Applications

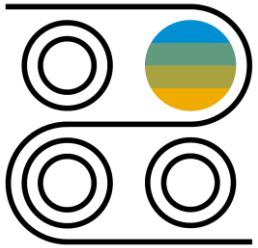
LAND



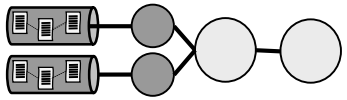


SAP DS | HDFS | S3

REFINE



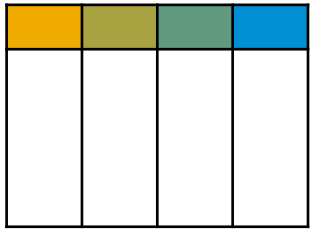
Join Cleanse Script Anonymize
Filter Look-up Mask Parse

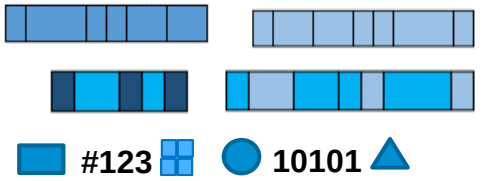


Spark Job Python Script

Data Hub

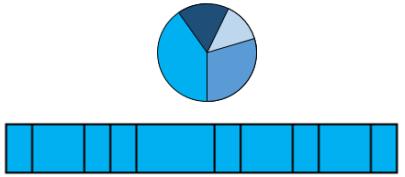
CONSUME





Vora

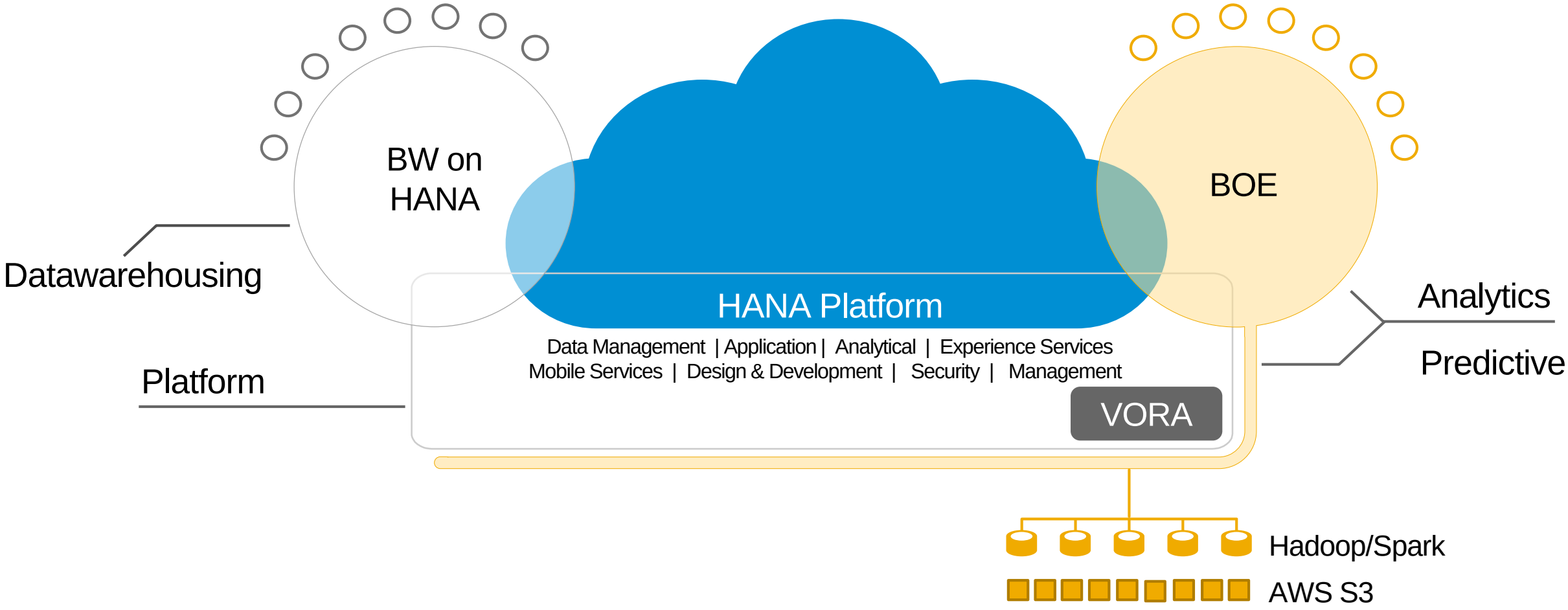
ACT



HANA | BW | Apps

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

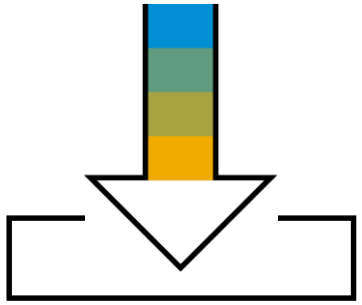
Kundenbeispiel
High-Level Architecture

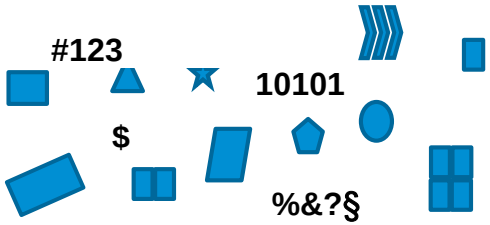


© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

From Data to Data Driven Applications

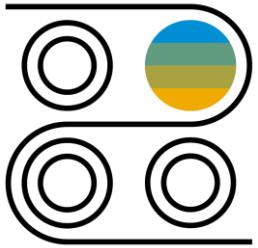
LAND



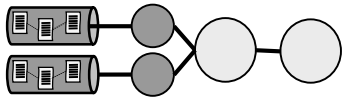


SAP DS | Kafka | HDFS | S3

REFINE



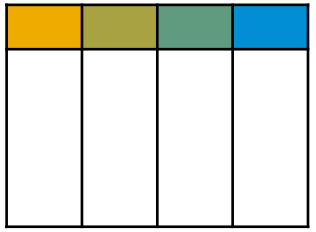
Join Cleanse Script Anonymize
Filter Look-up Mask Parse

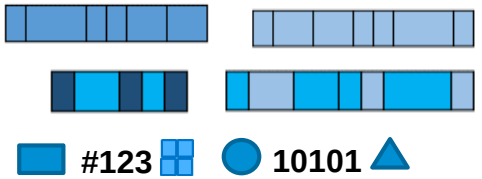


Spark Job Python Script

Data Hub

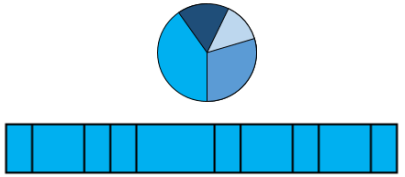
CONSUME





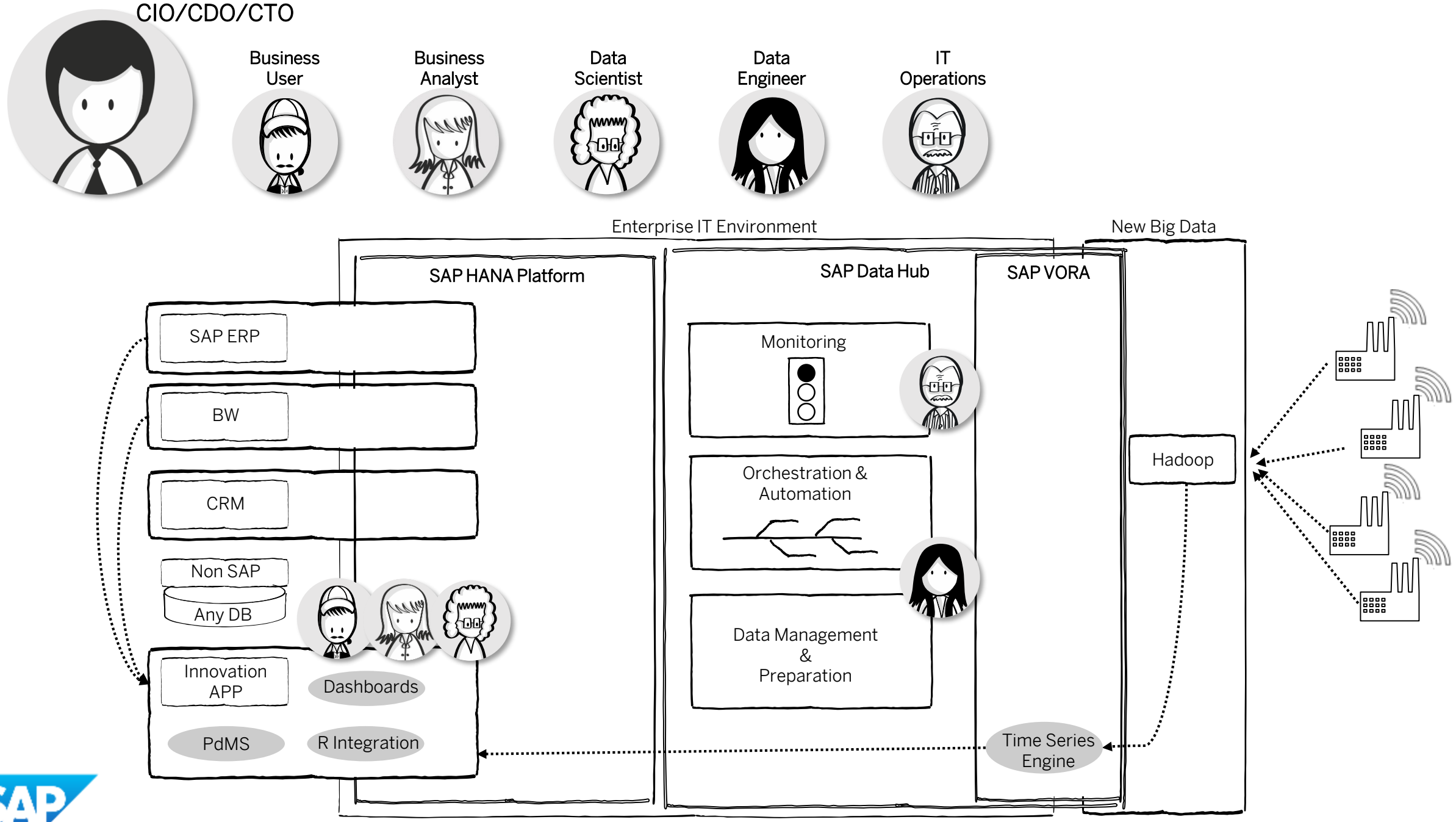
Vora

ACT

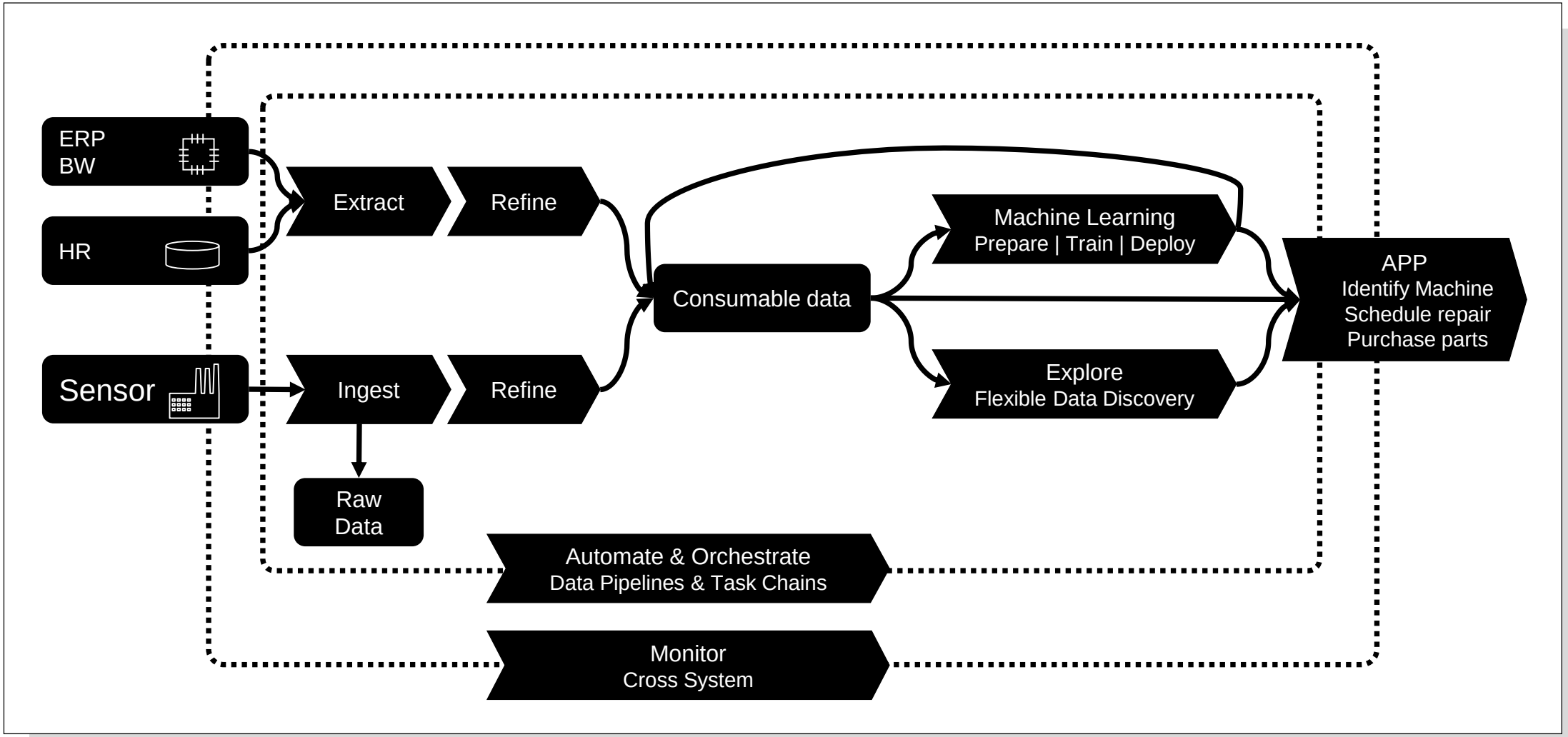


HANA | BW | Apps

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER


CIO/CDO/CTO
Business User
Business Analyst
Data Scientist
Data Engineer
IT Operations
Enterprise IT Environment
New Big Data
SAP HANA Platform
SAP Data Hub
SAP VORA
SAP ERP
BW
CRM
Non SAP
Any DB
Innovation APP
Dashboards
PdMS
R Integration
Monitoring
Orchestration & Automation
Data Management & Preparation
Hadoop
Time Series Engine

Task Chain for IoT Scenario



© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

SAP Data Hub

SAP Data Hub – Scope

SAP DATA HUB - Distributed Data Hub & Refinery

Refines enterprise big data across a heterogeneous data landscape to drive digital business decisions

Data Cockpit

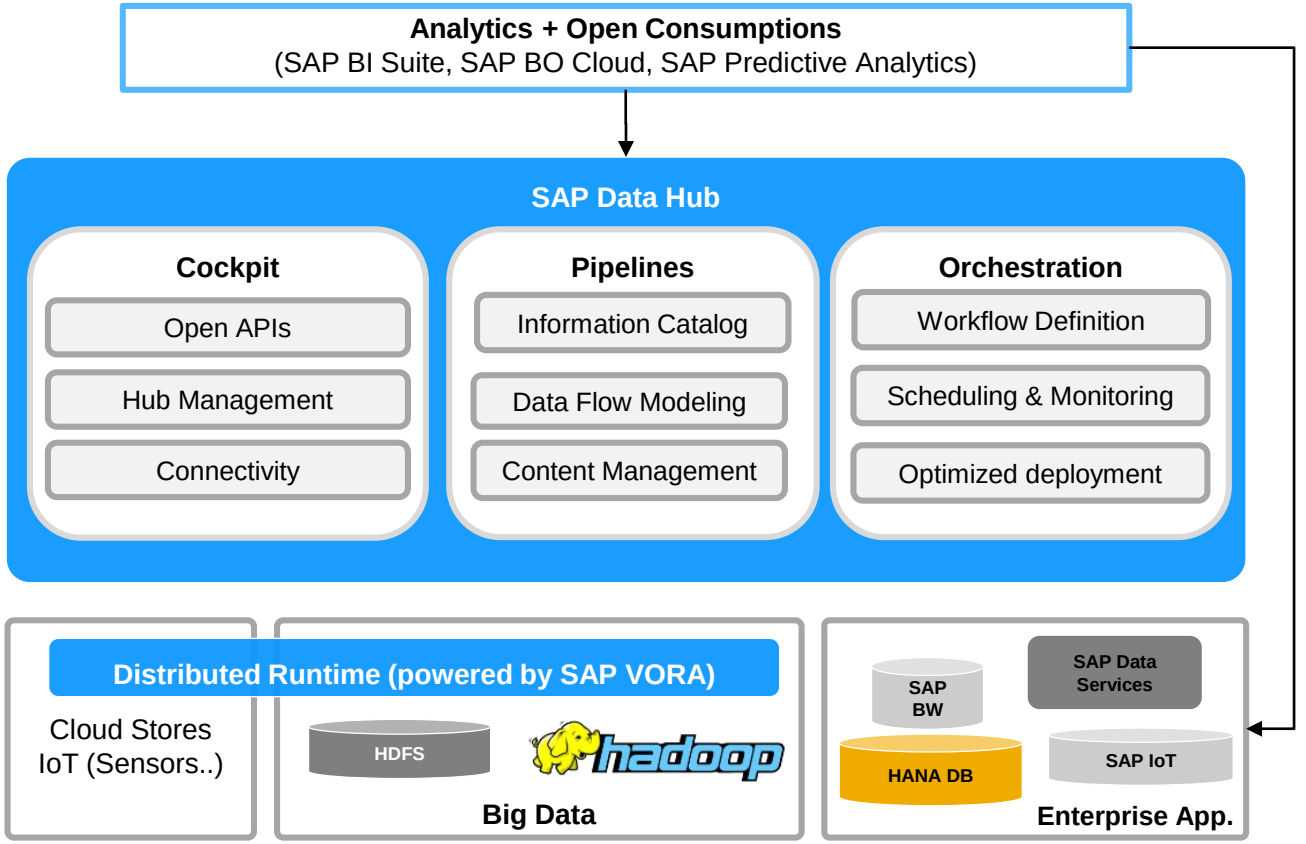
- Role-based Hub Experience with a modern UX
- Creation of logical hub structure e.g. zone management
- Connectivity to hybrid system landscapes
- Big Data Discovery & Security

Data Pipelines

- Data Ingestion (Connectors) and Integration
- Data Pipelines suited for Server less Computing
- Data Quality Flow with predefined operations
- Meta Data Services across stores with Lineage

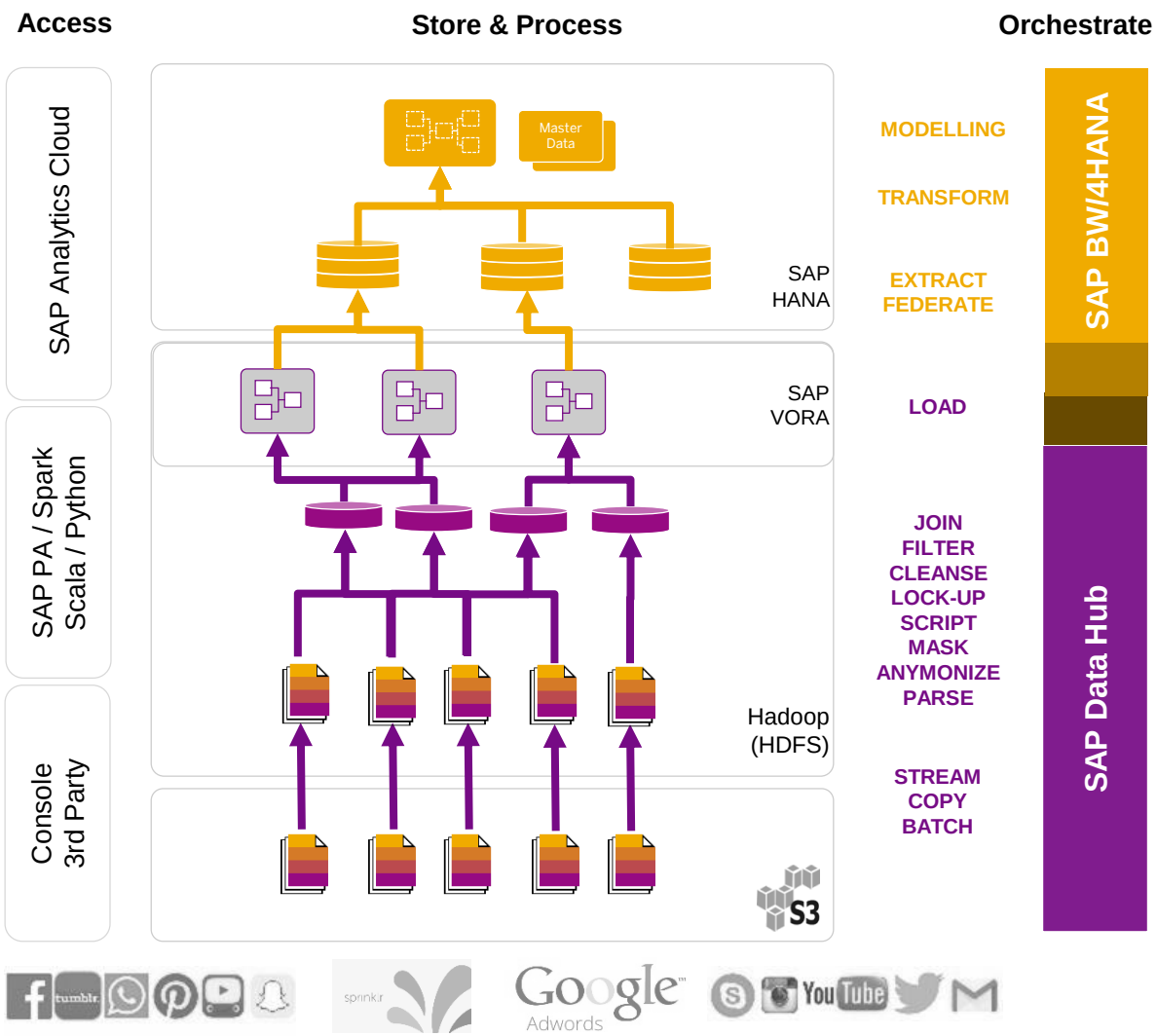
Orchestration

- Workflow scheduling
- Monitoring and alerting
- Pipeline deployment to SAP HANA and Vora



© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

SAP Data Hub – Analyzing Web Log Data with Data Warehouse



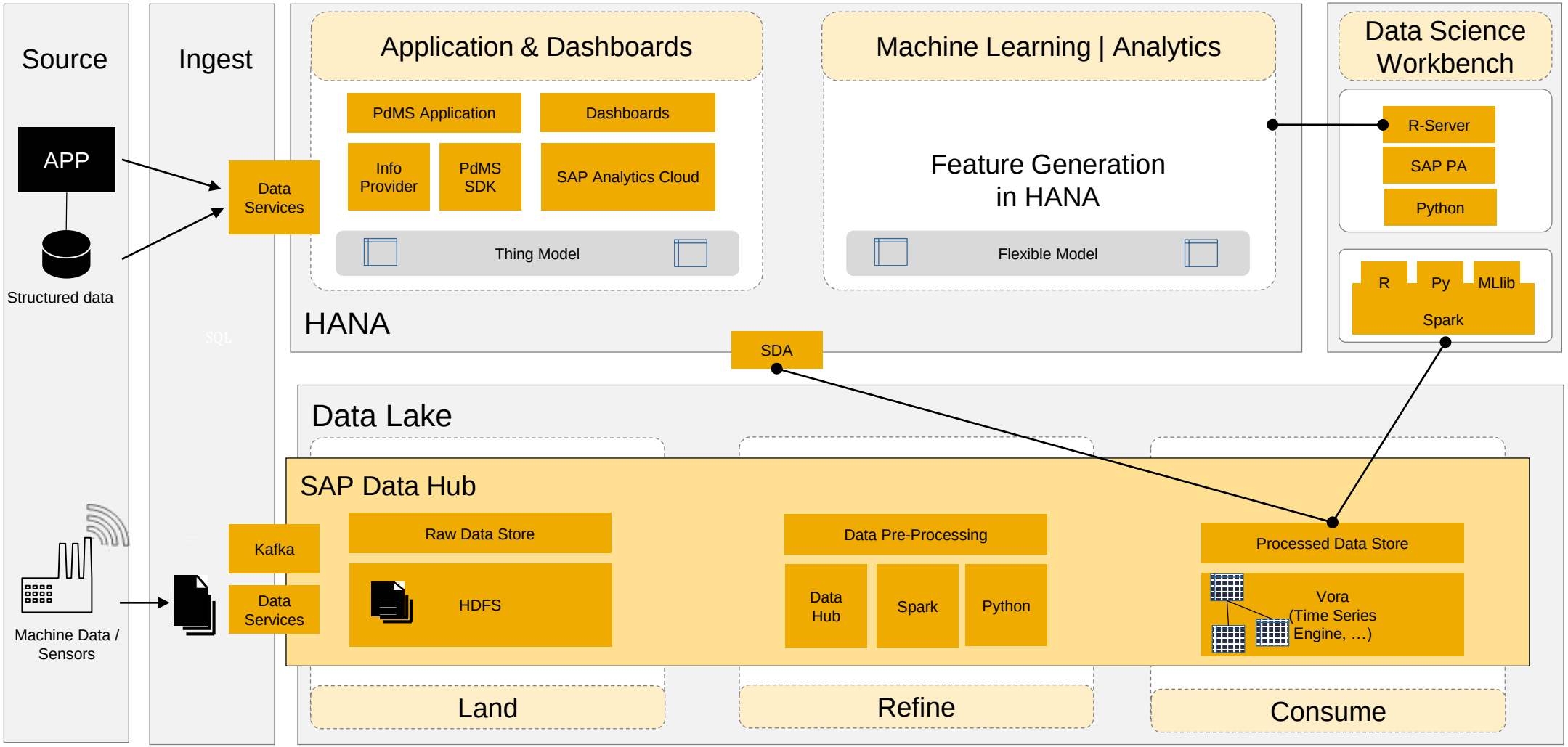
Example Scenario

- 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

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

IoT Architecture for Predictive Maintenance



© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

USER EXPERIENCE

Same thing – different perspective

A logistic regression takes me only a few lines of code.

Data Scientist

```scala
import org.apache.spark.mllib.classification.{SVMModel, SVMWithSGD}
import org.apache.spark.mllib.evaluation.BinaryClassificationMetrics
import org.apache.spark.mllib.util.MLUtils

// Load training data in LIBSVM format.
val data = MLUtils.loadLibSVMFile(sc, "data/mllib/sample_libsvm_data.txt")

// Split data into training (60%) and test (40%).
val splits = data.randomSplit(Array(0.6, 0.4), seed = 11L)
val training = splits(0).cache()
val test = splits(1)

// Run training algorithm to build the model
val numIterations = 100
val model = SVMWithSGD.train(training, numIterations)

// Clear the default threshold.
model.clearThreshold()

// Compute raw scores on the test set.
val scoreAndLabels = test.map { point =>
  val score = model.predict(point.features)
  (score, point.label)
}

// Get evaluation metrics.
val metrics = new BinaryClassificationMetrics(scoreAndLabels)
val auROC = metrics.areaUnderROC()

println("Area under ROC = " + auROC)

// Save and load model
model.save(sc, "target/tmp/scalaSVMWithSGDModel")
val sameModel = SVMModel.load(sc, "target/tmp/scalaSVMWithSGDModel")
```

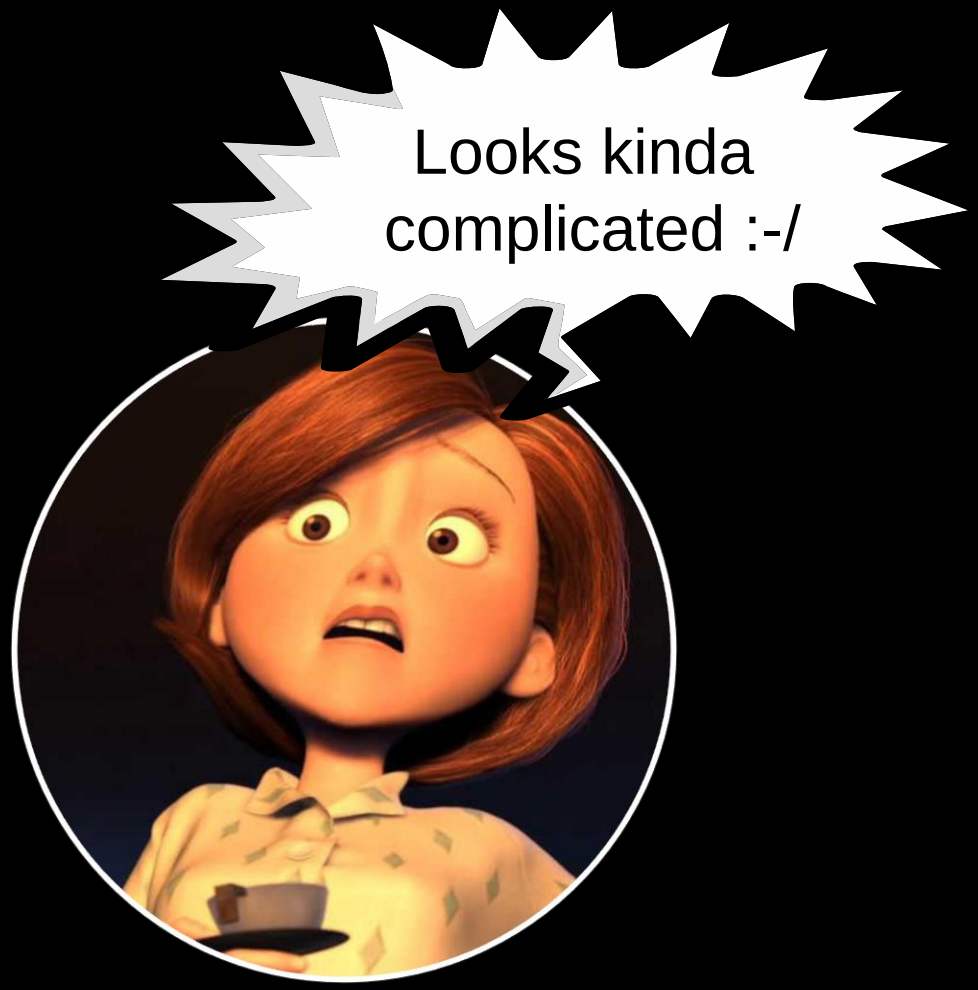
Split data

Train one model

Apply the model on new data

Evaluate model quality

Source: http://spark.apache.org/docs/latest/mllib-linear-methods.html



Business Analyst

```scala
import org.apache.spark.mllib.classification.{SVMModel, SVMWithSGD}
import org.apache.spark.mllib.evaluation.BinaryClassificationMetrics
import org.apache.spark.mllib.util.MLUtils

// Load training data in LIBSVM format.
val data = MLUtils.loadLibSVMFile(sc, "data/mllib/sample_libsvm_data.txt")

// Split data into training (60%) and test (40%).
val splits = data.randomSplit(Array(0.6, 0.4), seed = 11L)
val training = splits(0).cache()
val test = splits(1)

// Run training algorithm to build the model
val numIterations = 100
val model = SVMWithSGD.train(training, numIterations)

// Clear the default threshold.
model.clearThreshold()

// Compute raw scores on the test set.
val scoreAndLabels = test.map { point =>
  val score = model.predict(point.features)
  (score, point.label)
}

// Get evaluation metrics.
val metrics = new BinaryClassificationMetrics(scoreAndLabels)
val auROC = metrics.areaUnderROC()

println("Area under ROC = " + auROC)

// Save and load model
model.save(sc, "target/tmp/scalaSVMWithSGDModel")
val sameModel = SVMModel.load(sc, "target/tmp/scalaSVMWithSGDModel")
```

Split data

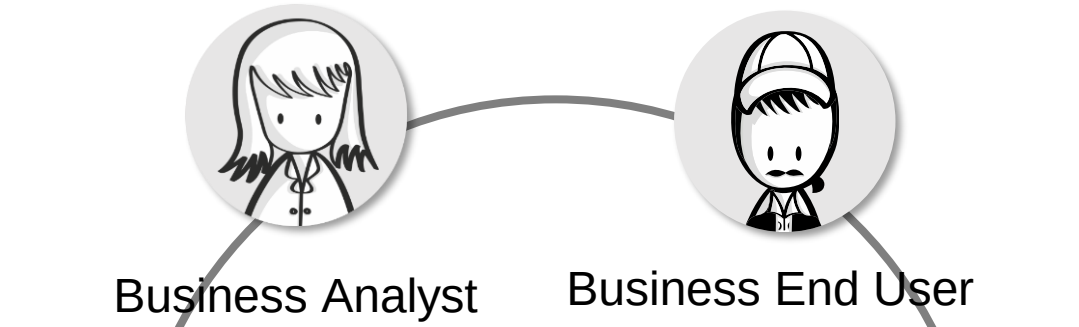
Train one model

Apply the model on new data

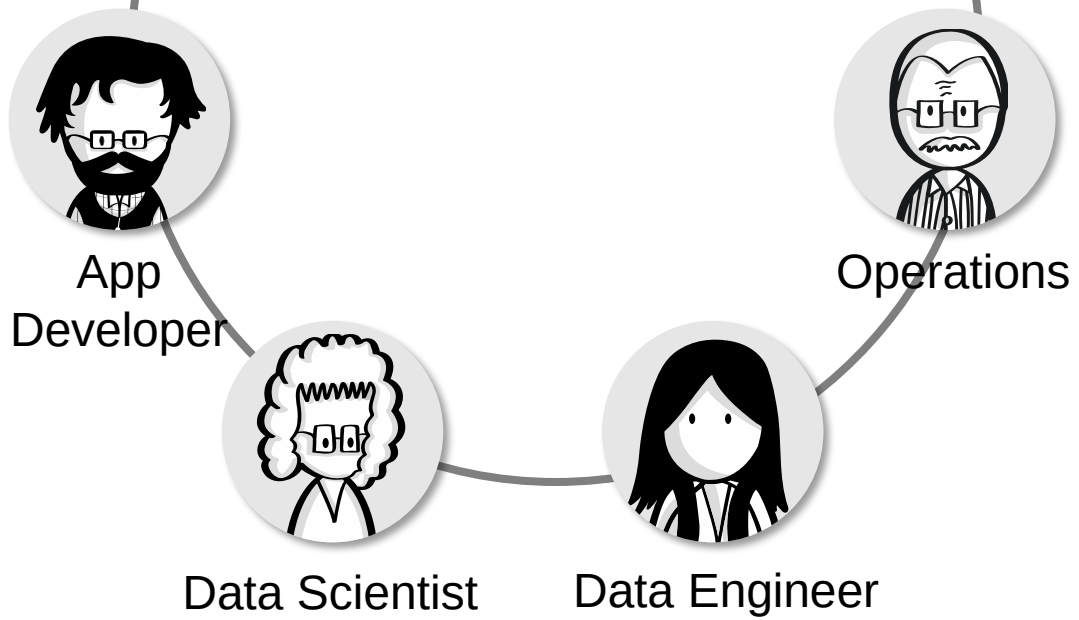
Evaluate model quality

Source: http://spark.apache.org/docs/latest/mllib-linear-methods.html

Data Innovation is a Team Effort



- Ease of use
- Big Data Self Service
- Use existing skills

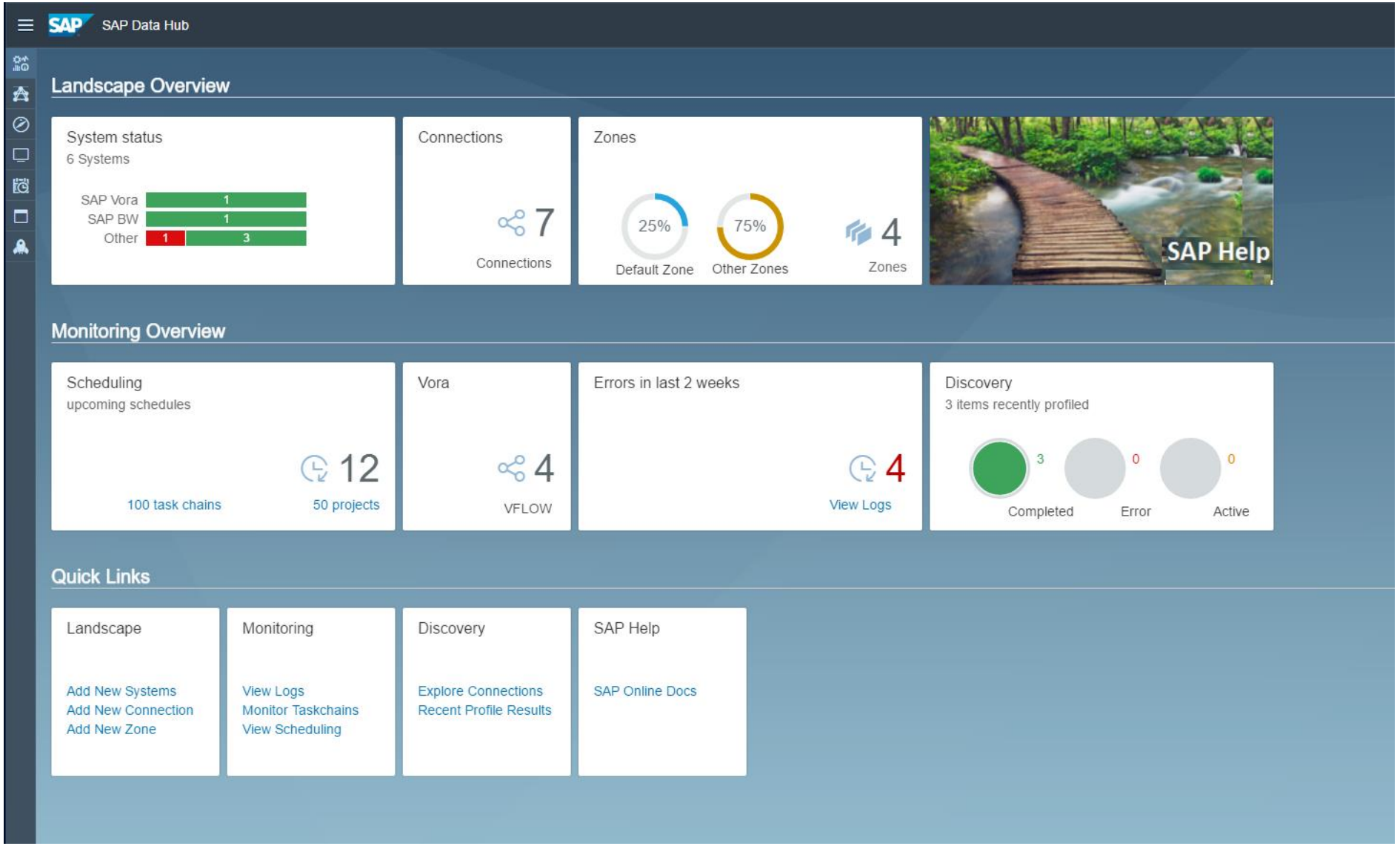


- Easy data access and collaboration
- Automation, orchestration & documentation
- Monitoring across systems

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Cross System Monitoring & Management

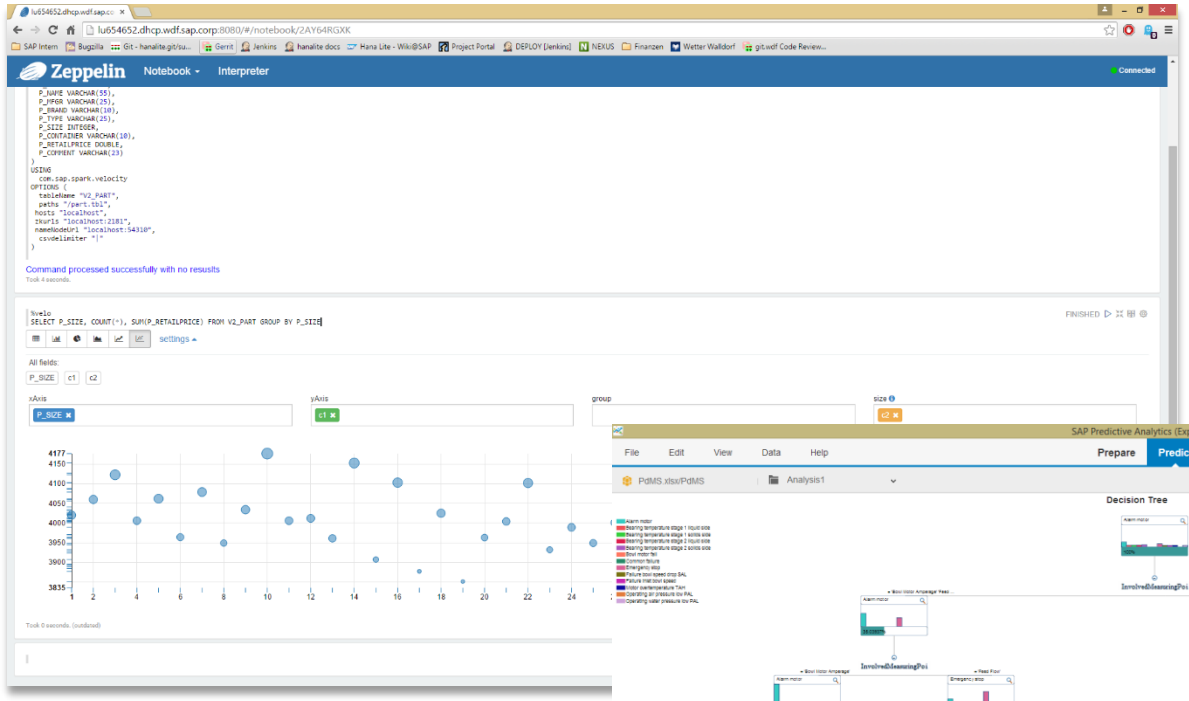
IT Operations

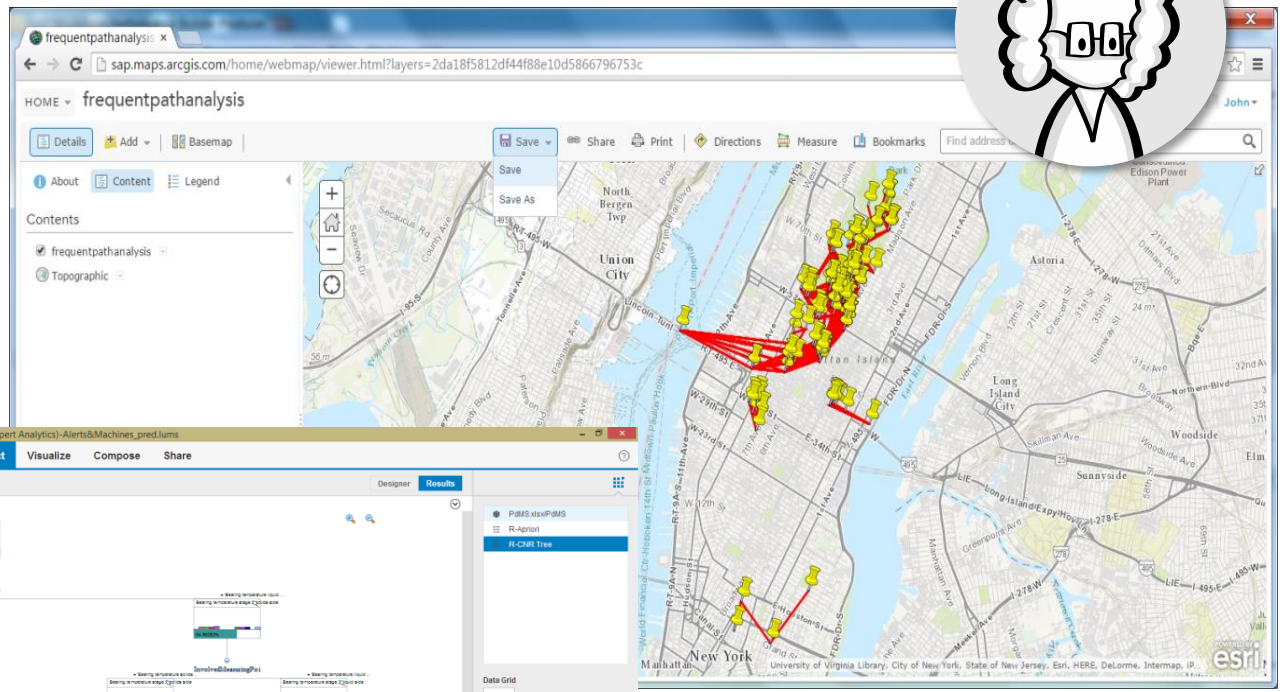


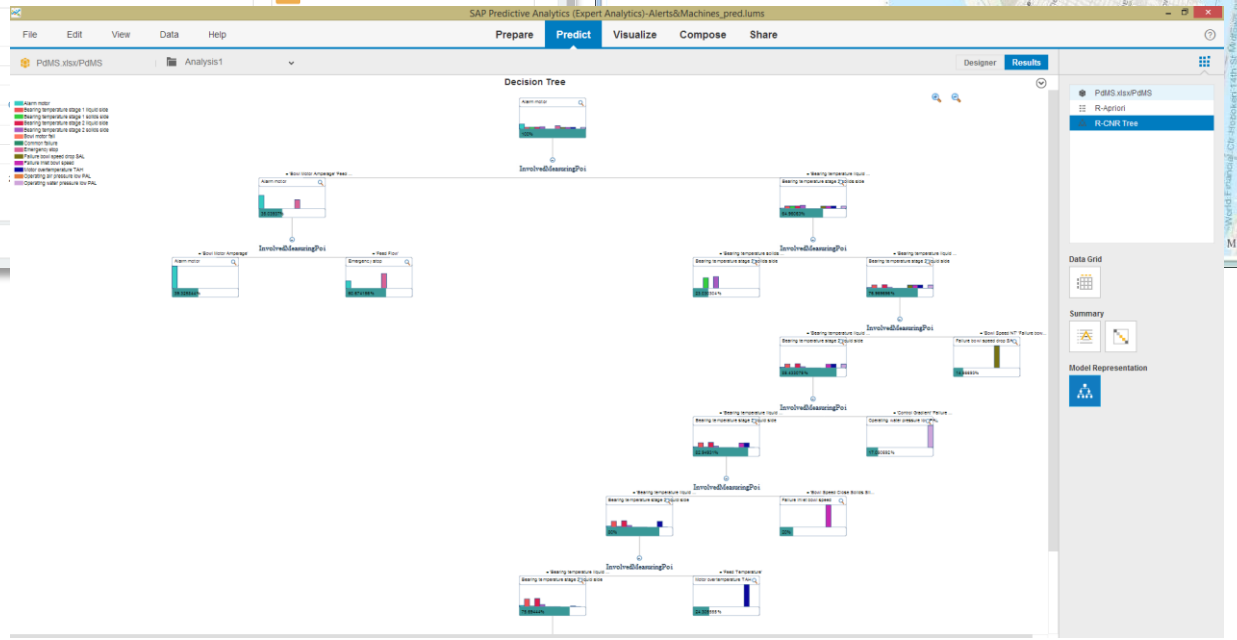
© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Data Scientist – Apache Zeppelin & SAP Predictive Analytics

Data Scientist



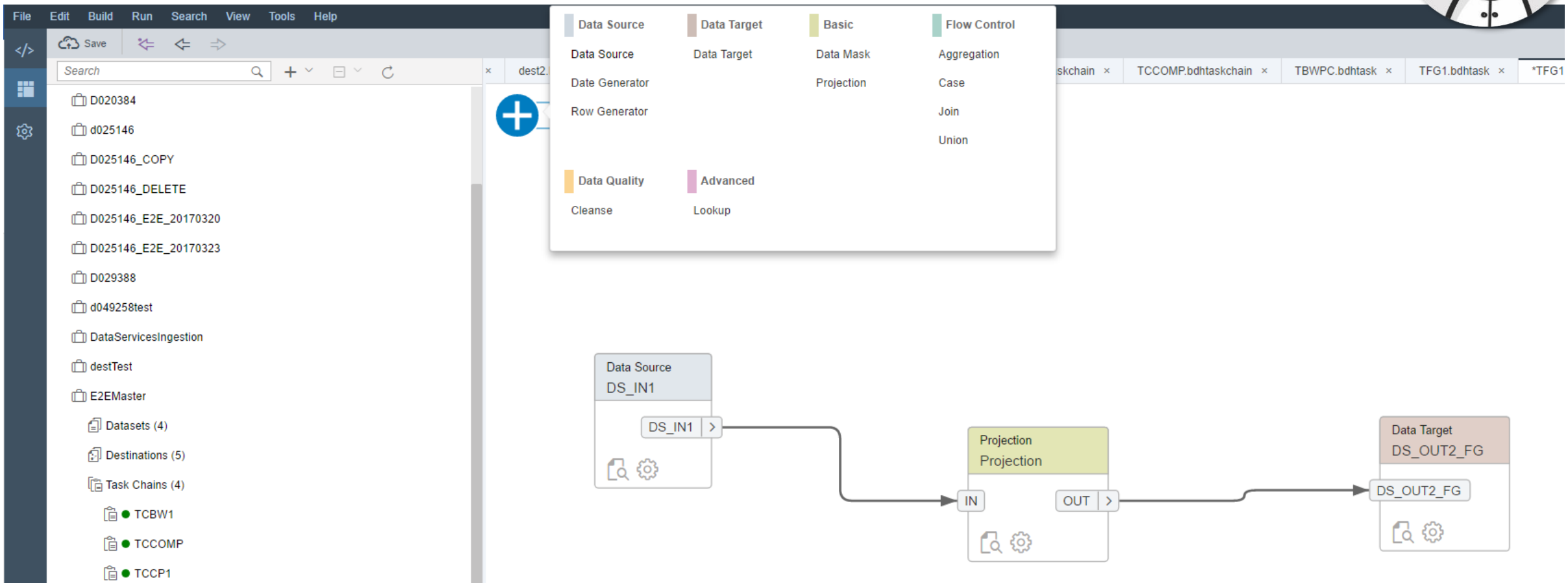




© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Big Data Self-Service Data Preparation

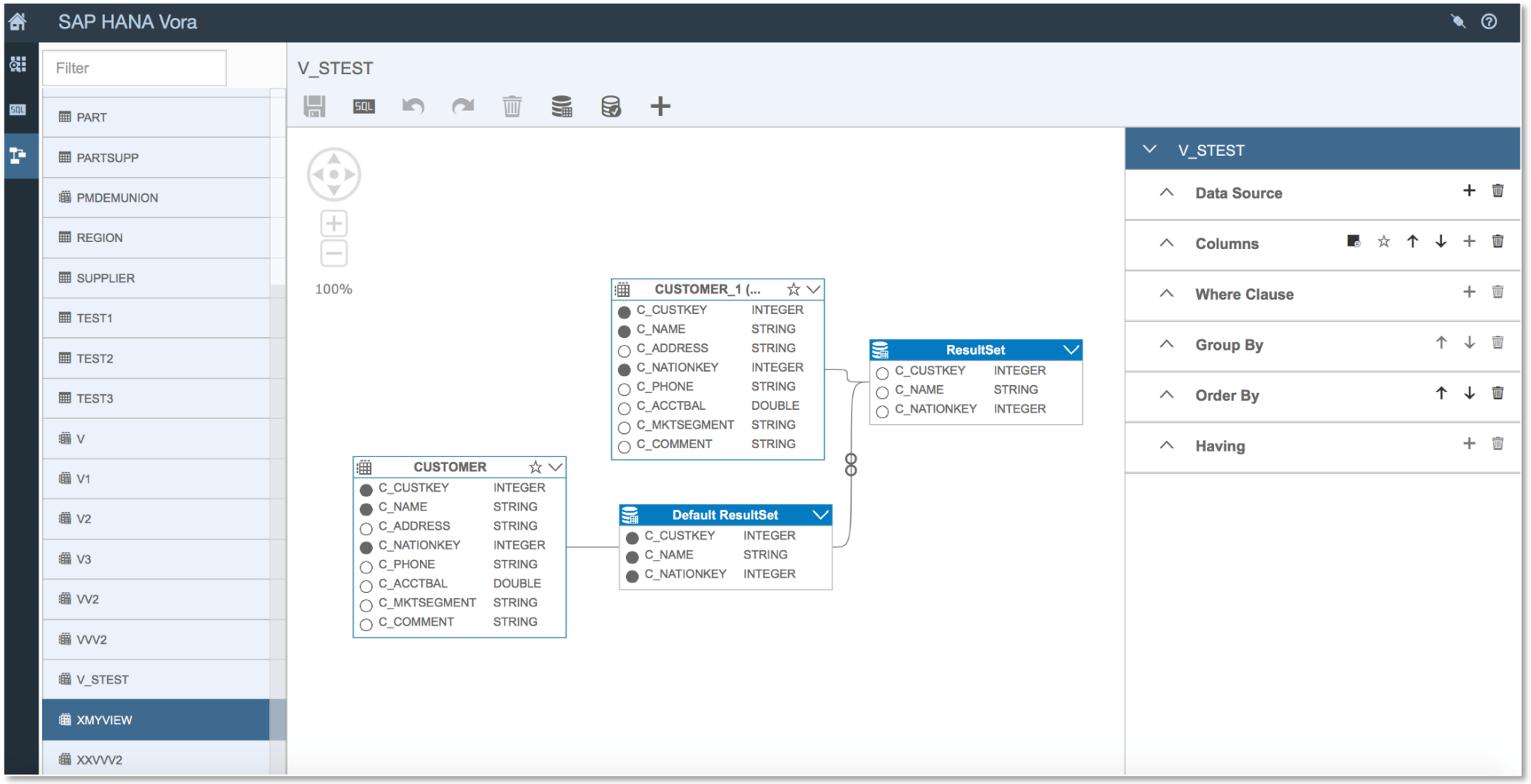
Business Analyst



© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Big Data Self-Service Data Modelling

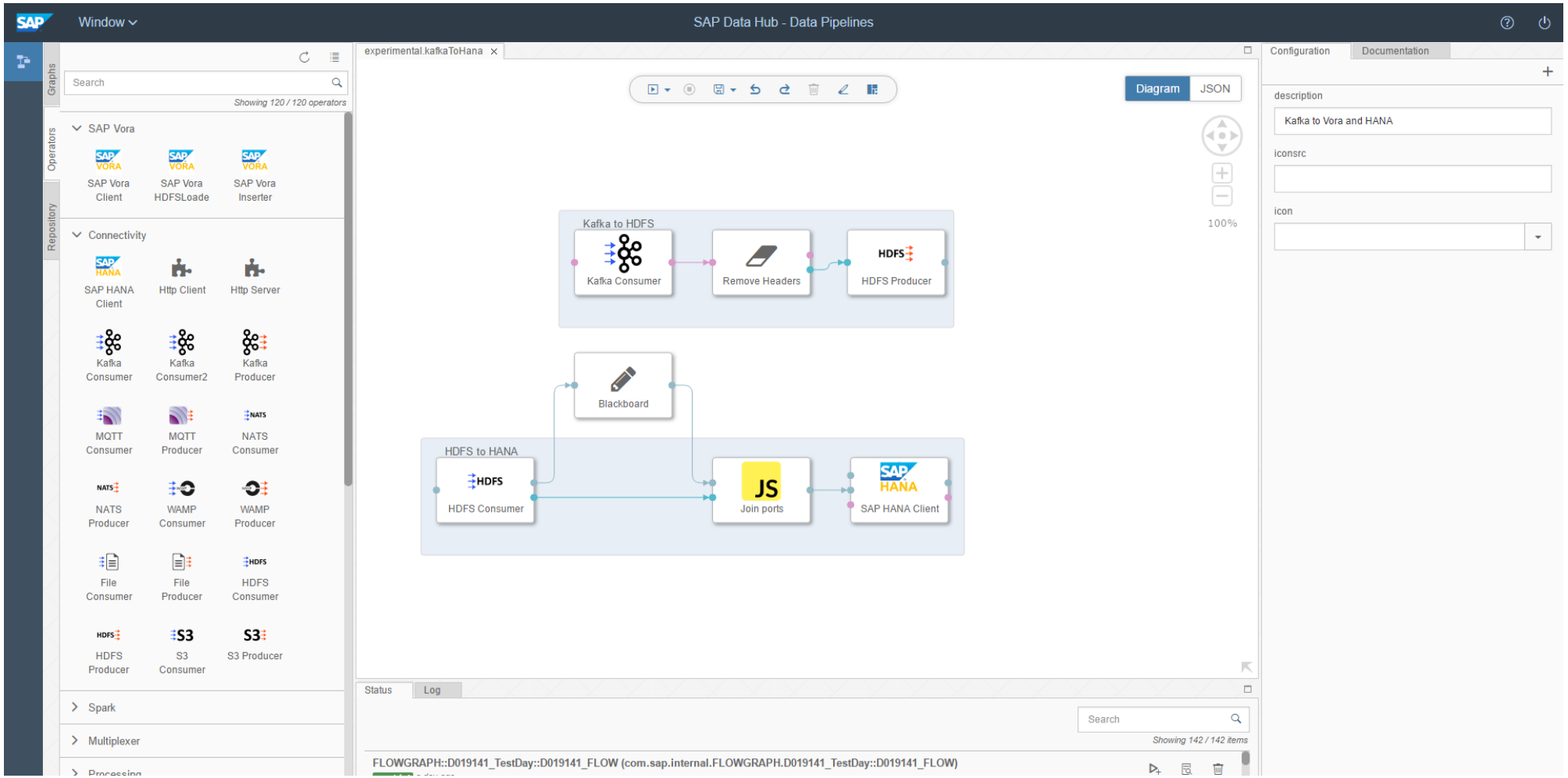
Business Analyst



© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Design Cross-System Task Chains

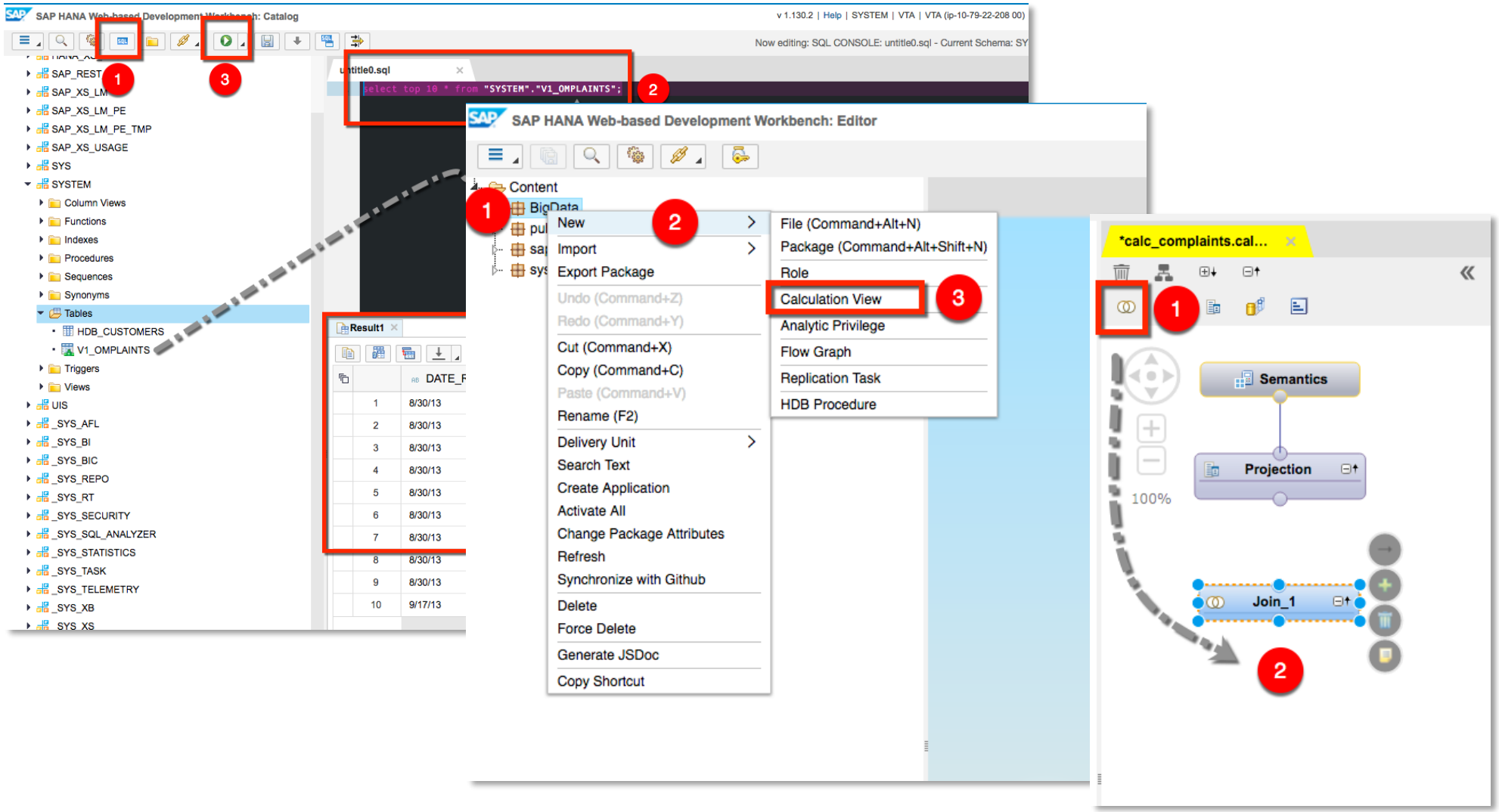
Data Engineer



© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

HANA Developer

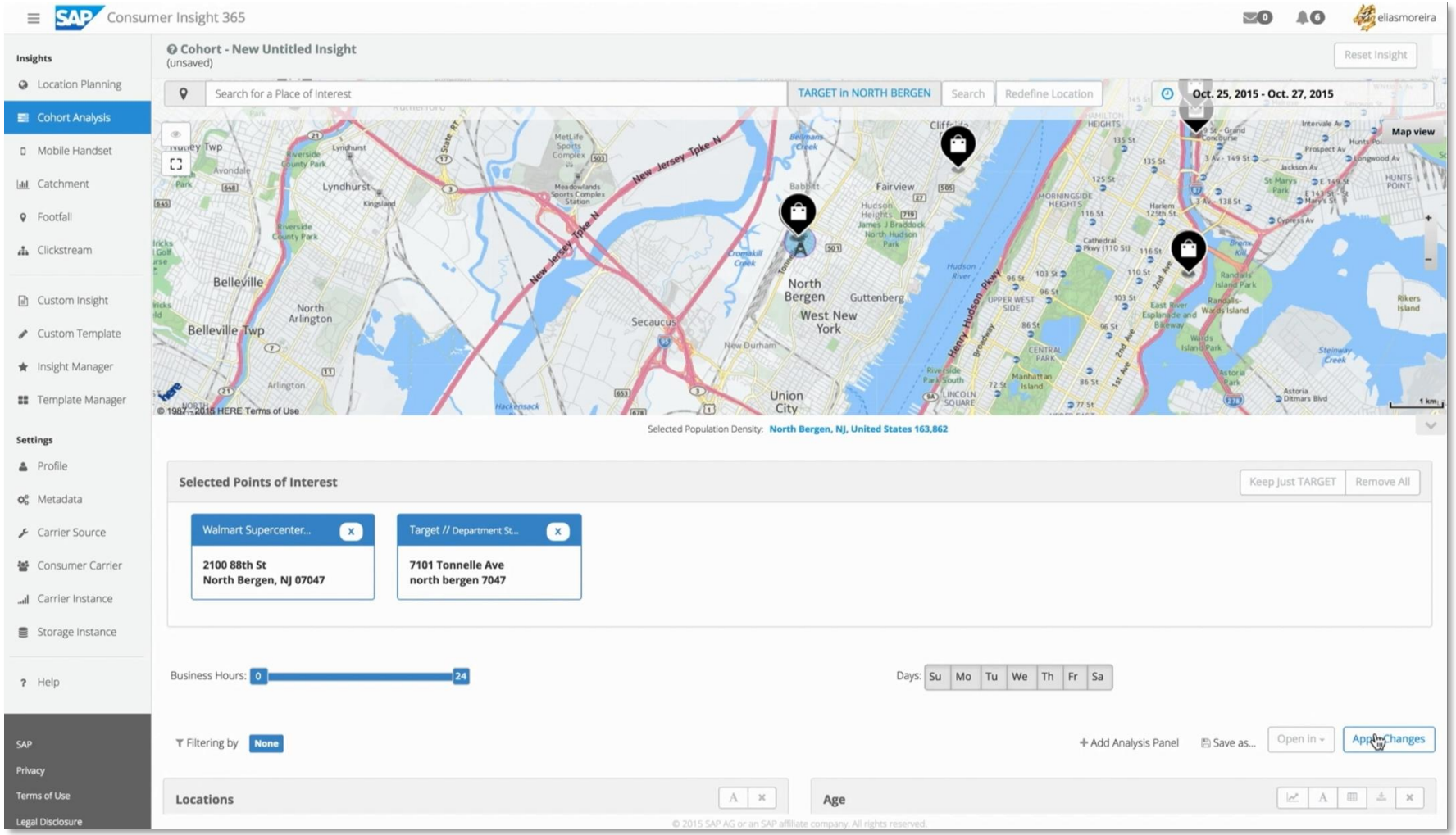
Data Engineer



© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Consume Data Driven Application

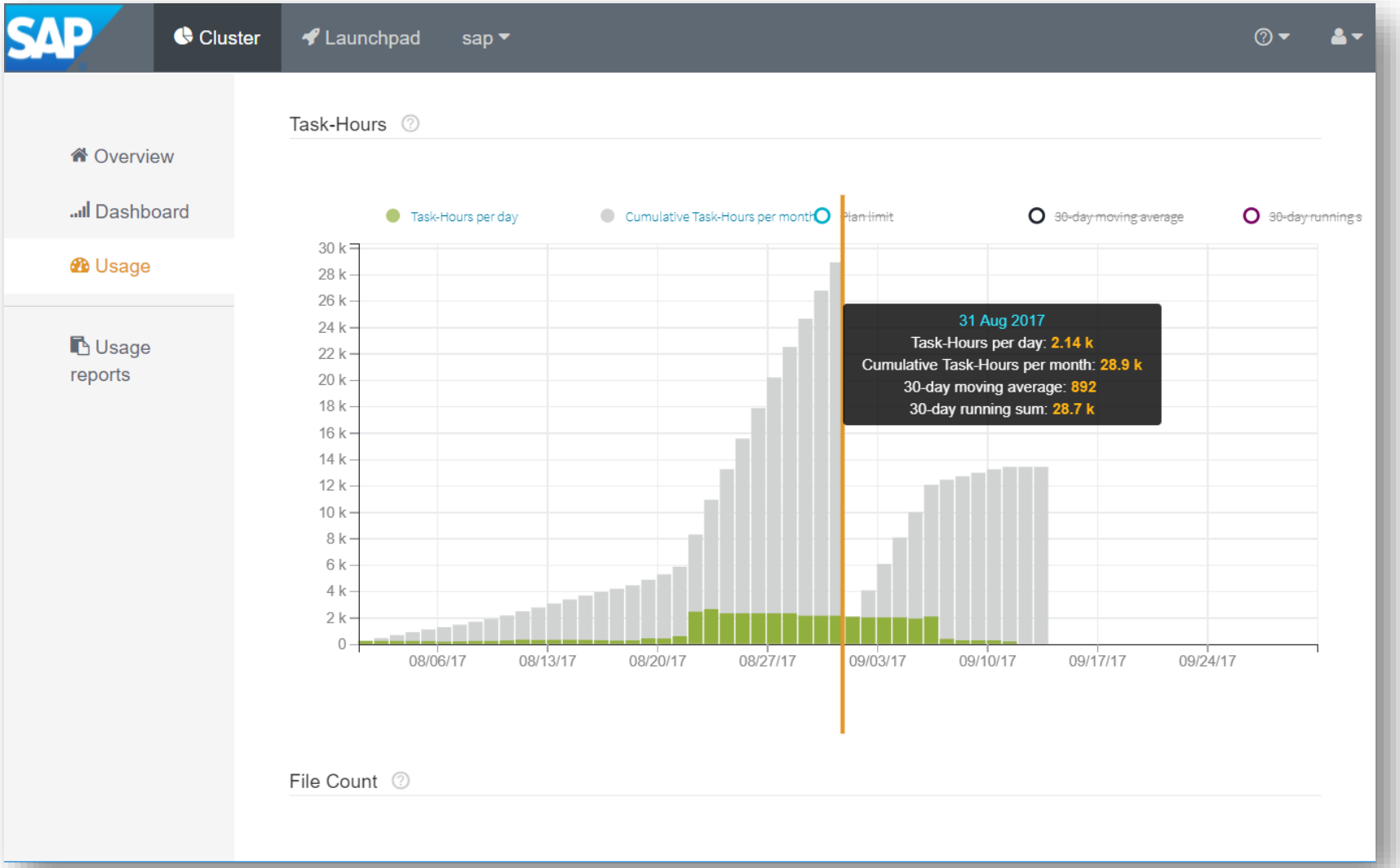
Business User



© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Stable.Fast.Predictable.

SAP Cloud Platform Big data Service



IT Operations

Stable & Fast:

- Jobs execute and
- Execute faster

Autoscaling

- Always ON
- Compute Burst

Cost & Ease

- Predictable & Transparent
- Consumption NOT Hardware
- No sizing needed

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

And now what … ?

~~H~~ANA Express (here)

~~V~~ora Developer Edition (here)

~~V~~ora GitHub Code Snipets (here)

~~H~~ANA Academy on YouTube

~~O~~pen SAP Course (here)

Self Learning

Design Thinking

SAP Tech Academy

© 2017 SAP SE or an SAP affiliate company. All rights reserved. I CUSTOMER

Thank you.

Contact information:

Karsten Haldenwang
Center of Excellence Database and Data Management
Middle & Eastern Europe

Solution Advisor HANA, Vora and Hadoop

karsten.Haldenwang@sap.com
+49-151-53858452

SAP Run Simple