C3: Data Driven Applications & Data Operations

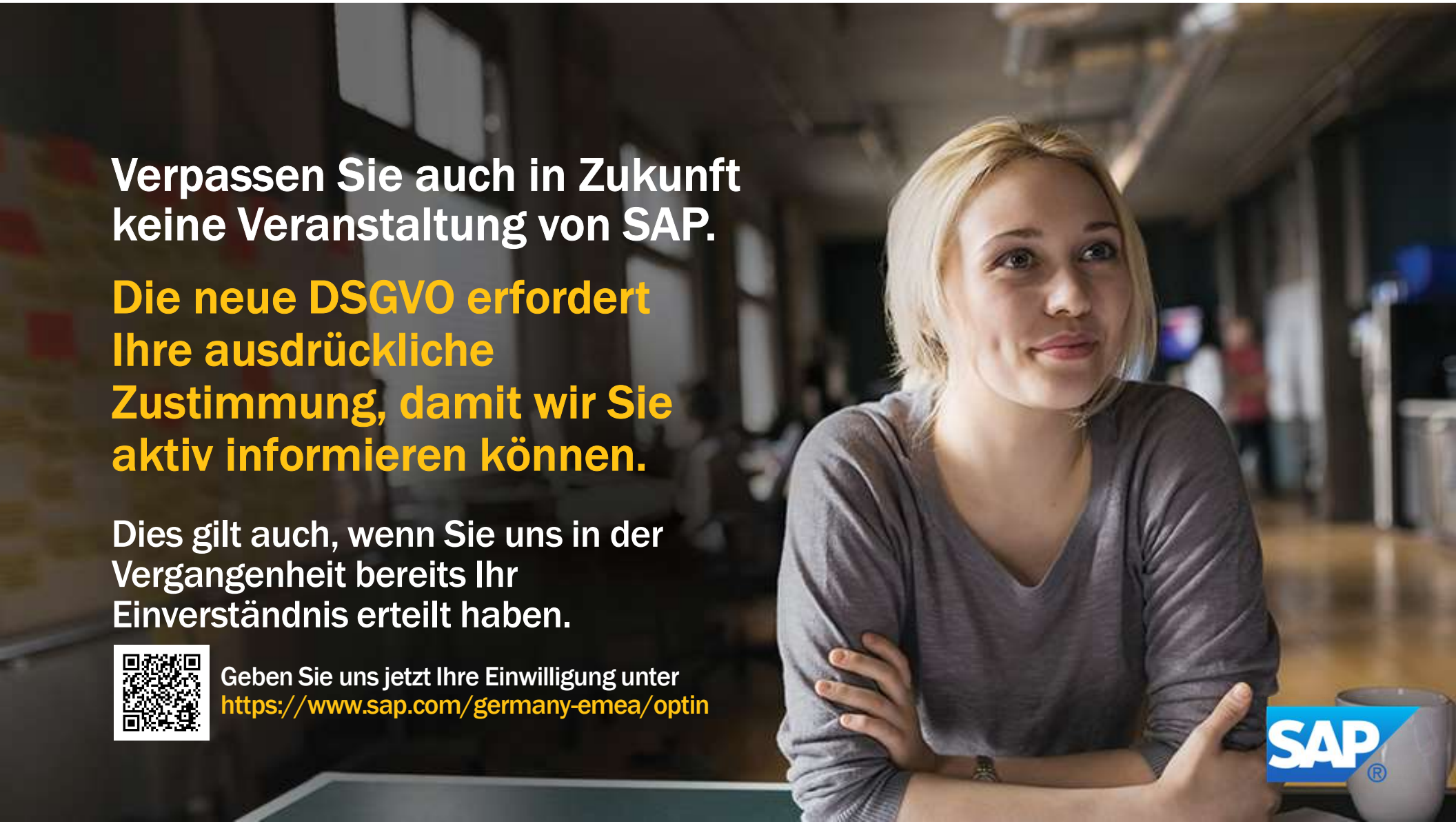
Michael Probst, Solution Advisor, SAP Switzerland

THE BEST RUN SAP

Legal disclaimer

The information in this presentation is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. This presentation is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP's strategy and possible future developments, products, and platforms, directions, and functionality are all subject to change and may be changed by SAP at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. This document is provided without a warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or noninfringement. This document is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this document, except if such damages were caused by SAP's willful misconduct or gross negligence.

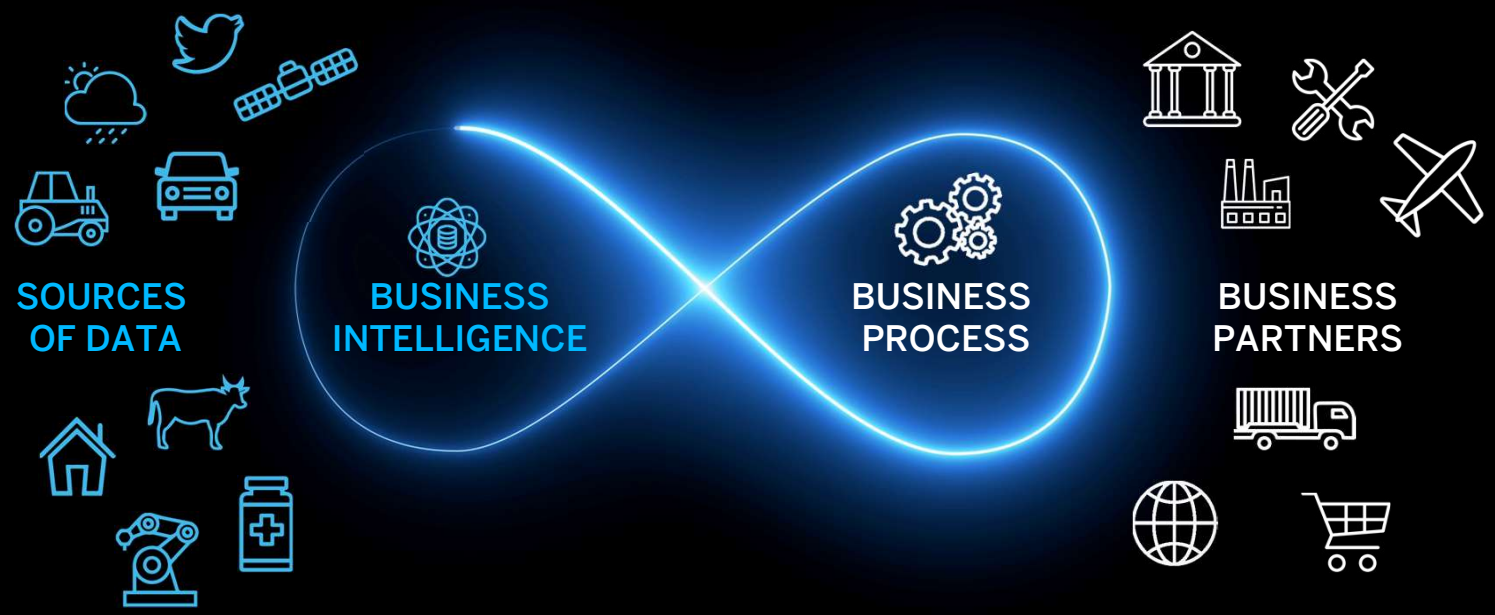
All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.


Verpassen Sie auch in Zukunft keine Veranstaltung von SAP.
Die neue DSGVO erfordert Ihre ausdrückliche Zustimmung, damit wir Sie aktiv informieren können.
Dies gilt auch, wenn Sie uns in der Vergangenheit bereits Ihr Einverständnis erteilt haben.
Geben Sie uns jetzt Ihre Einwilligung unter
https://www.sap.com/germany-emea/optin
SAP

Please opt-in now

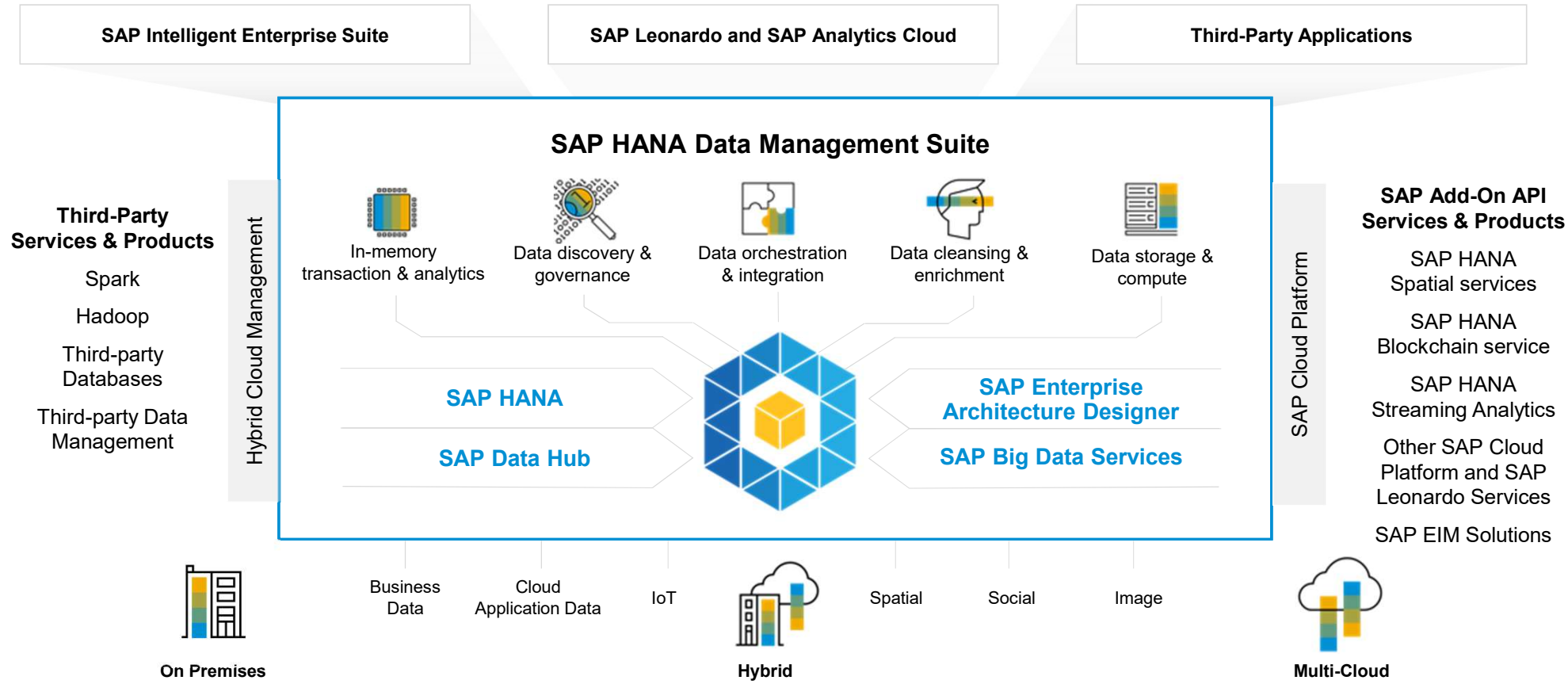
https://www.sap.com/germany-emea/optin

The Infinity Model



SAP HANA Data Management Suite

Trusted Data | Connected, Intelligent Data | Cloud Architecture Flexibility



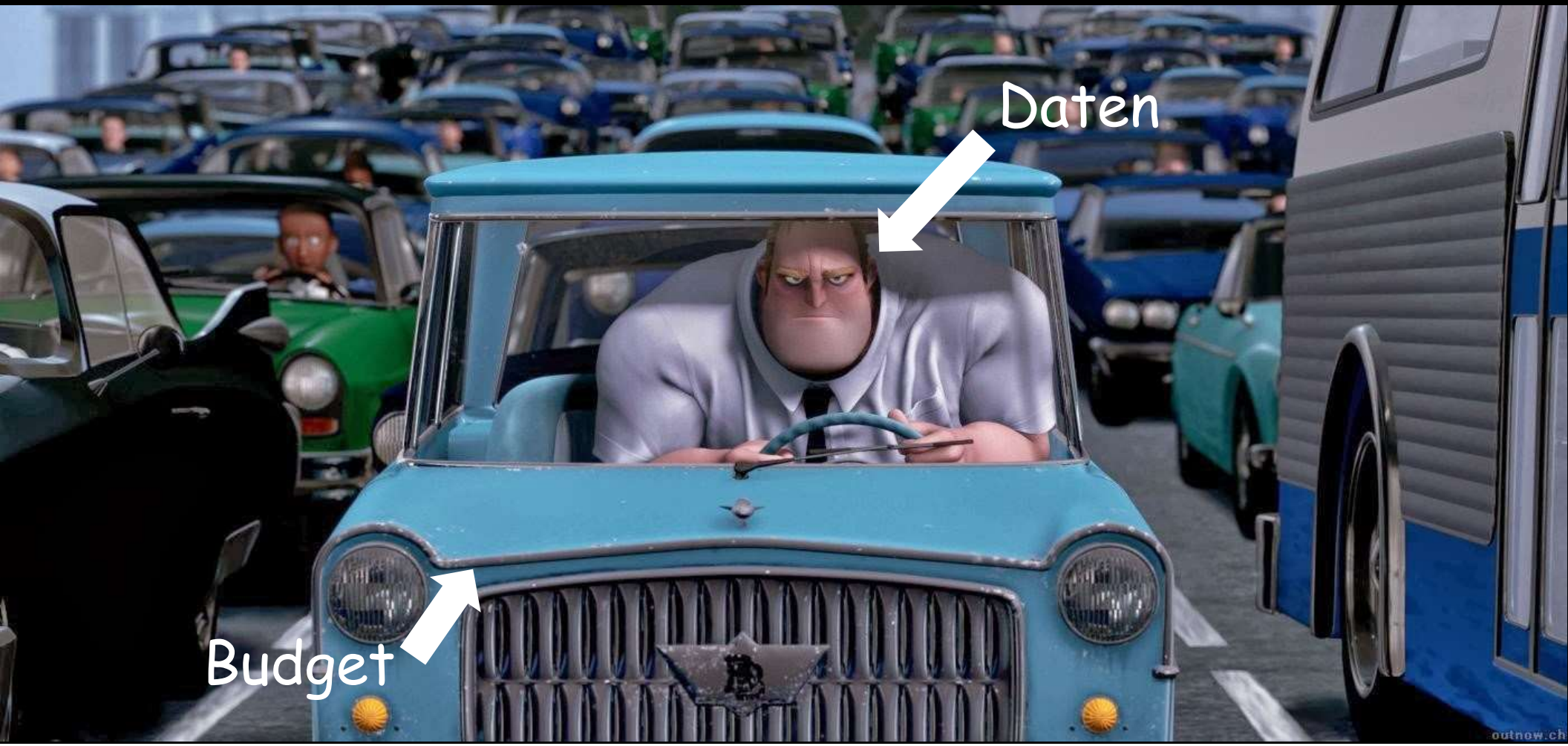
Agenda

Warum macht die SAP so etwas?

Wen interessiert das?

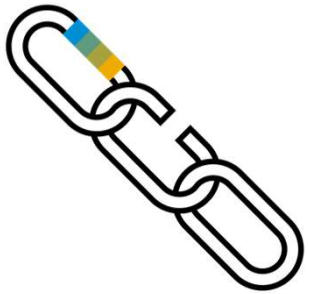
Wie funktioniert‘s? Und welches Produkt macht das?

Menge und Kosten…sind nicht der Grund…



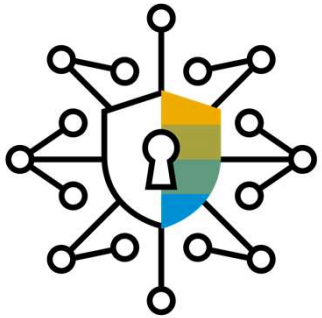
Challenges to uniting the landscape

Overcoming silos, complexity to drive better operations and insight



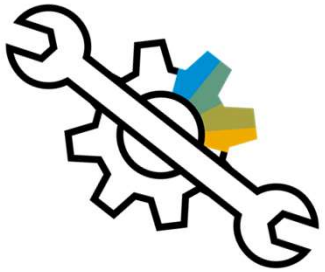
Missing link

between Big Data and Enterprise Data



Lack of enterprise-ready security and governance

across complex modern landscapes



Many limited tools = high Effort

to productize complex data scenarios across data landscape

> **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

SAP Data Management

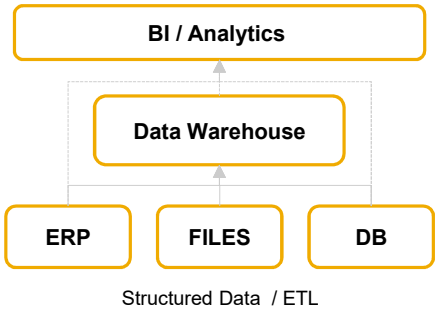
1 Enable Data Driven Applications

2 Drive Broad User Acceptance

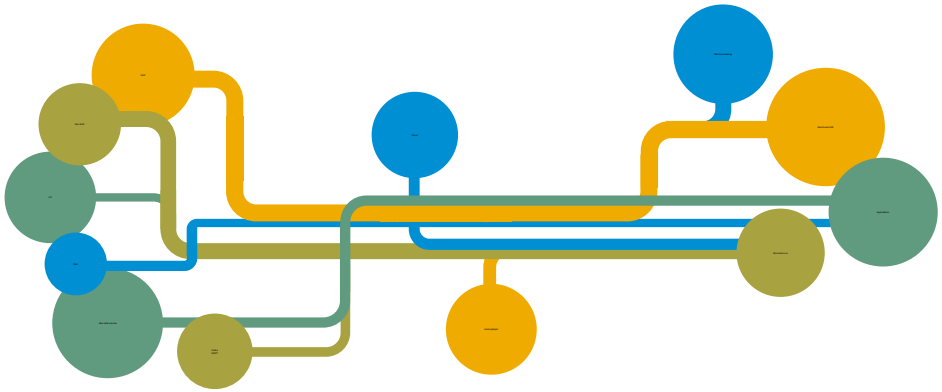
3 Grow from Pilot to Industrialization

SAP Data Management:
Enable Data Driven Applications

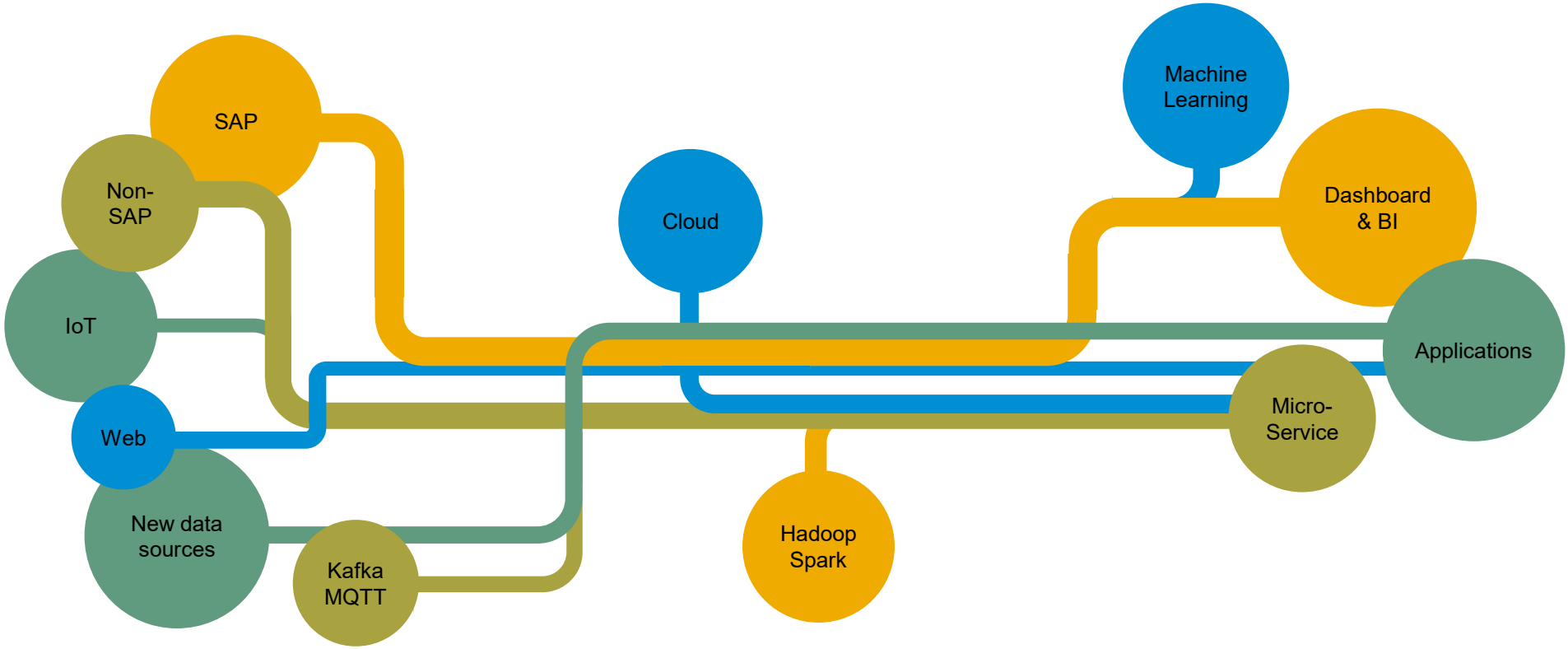
From a centralized, relational, on premise DWH approach…



to orchestration of decentralized, distributed data in a hybrid environment …

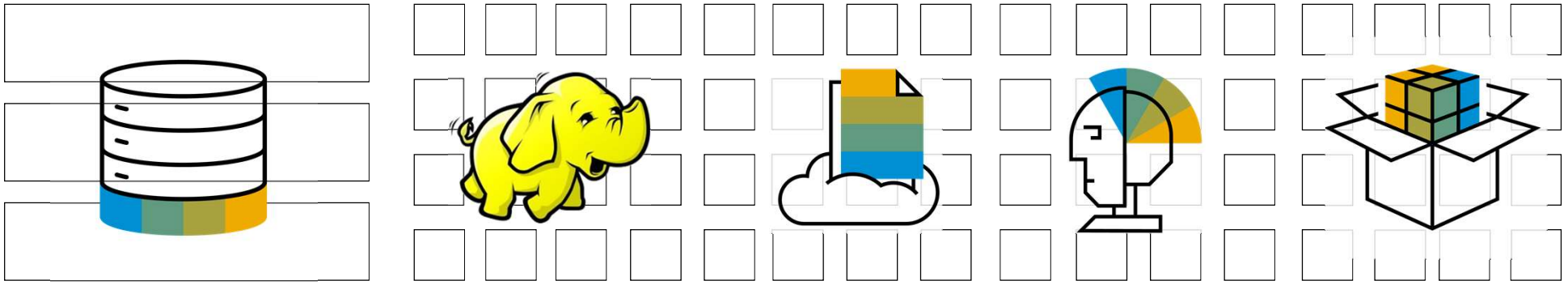


SAP Data Management:
Enable Data Driven Applications



New challenges require new technologies

Distributed systems in a distributed landscape



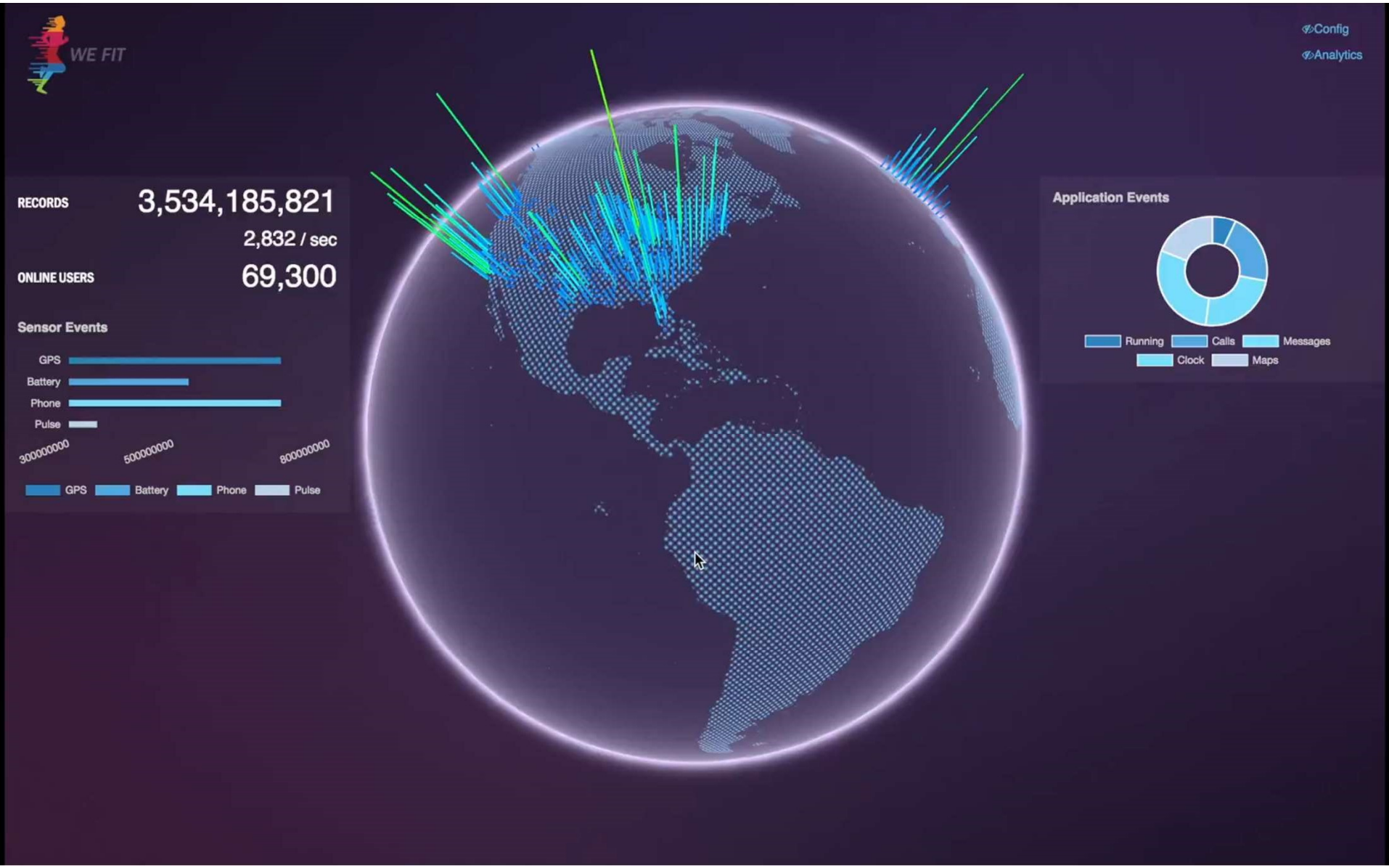
Existing Systems

Hadoop/Spark

Cloud Storage
(i.e. AWS S3)

Machine Learning
(Python, Spark, Tensorflow)

Containers
(Kubernetes, Docker)


WE FIT
Config
Analytics
RECORDS
3,534,185,821
2,832 / sec
ONLINE USERS
69,300
Sensor Events
GPS
Battery
Phone
Pulse
300000000
500000000
800000000
GPS
Battery
Phone
Pulse
Application Events
Running
Calls
Messages
Clock
Maps

Overview | Casual Jogger | Persistent Runner | Enthusiastic Sprinter | Professional Athlete

Casual Jogger

Number of Casual Joggers in Million

10.2

52%

Total Revenue with Casual Joggers in Million $

745

Average Revenue per Casual Jogger in $

73

3 Most Popular Shoes for Casual Joggers

Persistent Runner

Number of Persistent Runners in Million

5.1

26%

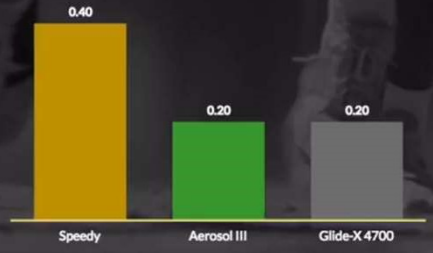
Total Revenue with Persistent Runners in Million $

1,739

Average Revenue per Persistent Runner in $

341

3 Most Popular Shoes for Peristent Runners



Enthusiastic Sprinter

Number of Enthusiastic Sprinters in Million

3.7

19%

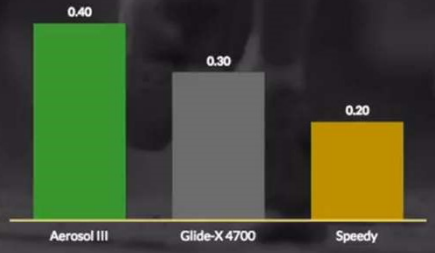
Total Revenue with Enthusiastic Sprinters in Million $

914

Average Revenue per Enthusiastic Sprinter in $

247

3 Most Popular Shoes for Enthuiastic Sprinters



Professional Athlete

Number of Professional Athletes in Million

0.6

3%

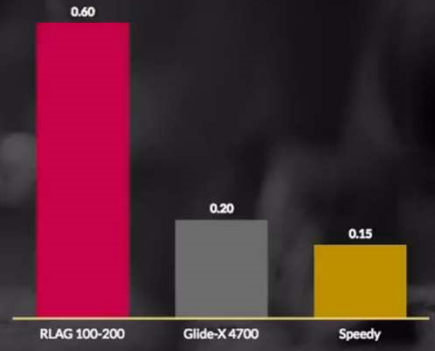
Total Revenue with Professional Athletes in Million $

812

Average Revenue per Professional Athlete in $

1,354

3 Most Popular Shoes for Professional Athletes



Not Assigned 7.8

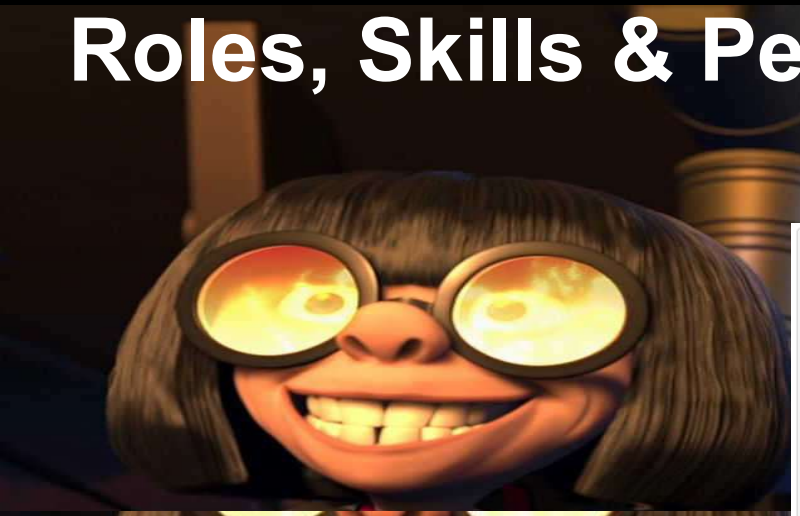
SAP Data Management

1 Enable Data Driven Applications

2 Drive Broad User Acceptance

3 Grow from Pilot to Industrialization

Roles, Skills & Perspective



```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")
```

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

ON-PREMISE | CLOUD | HYBRID

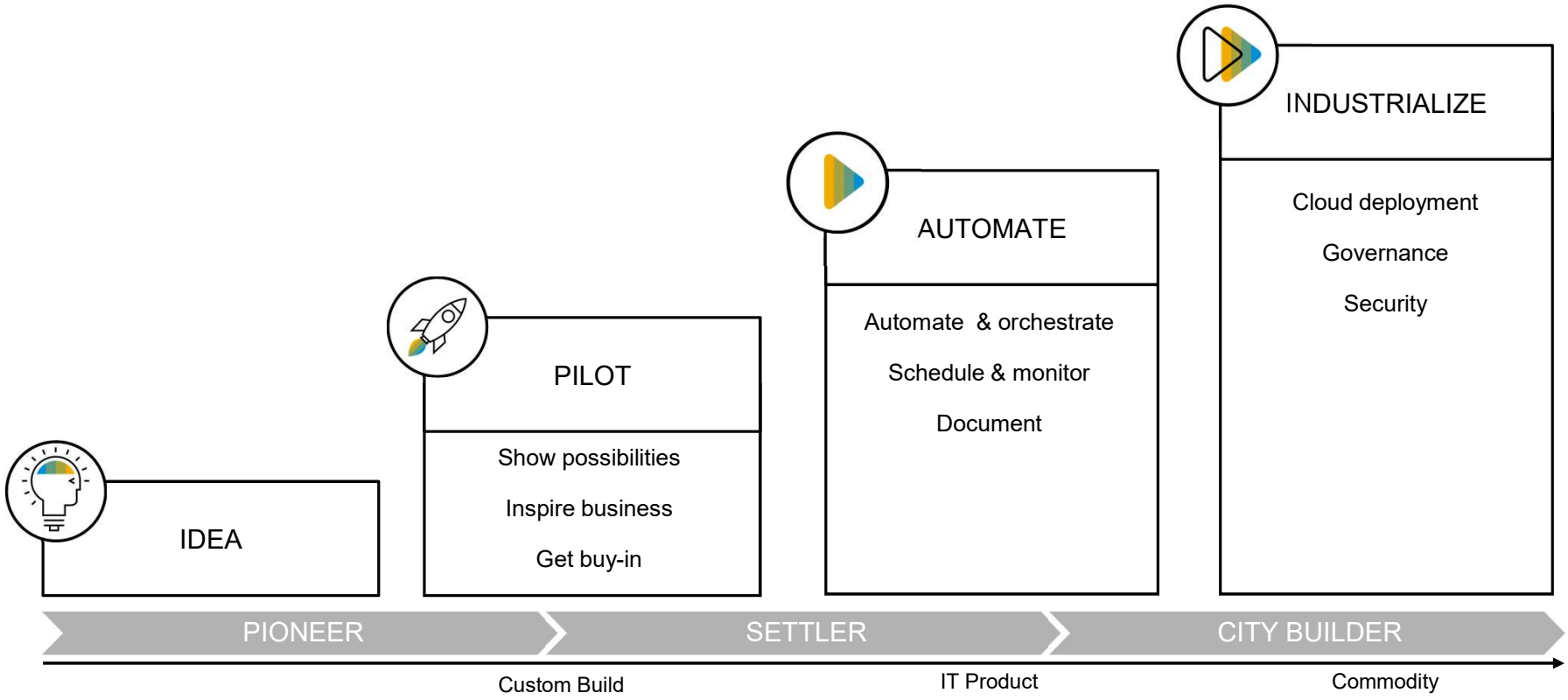
SAP Data Management

1 Enable Data Driven Applications

2 Drive Broad User Acceptance

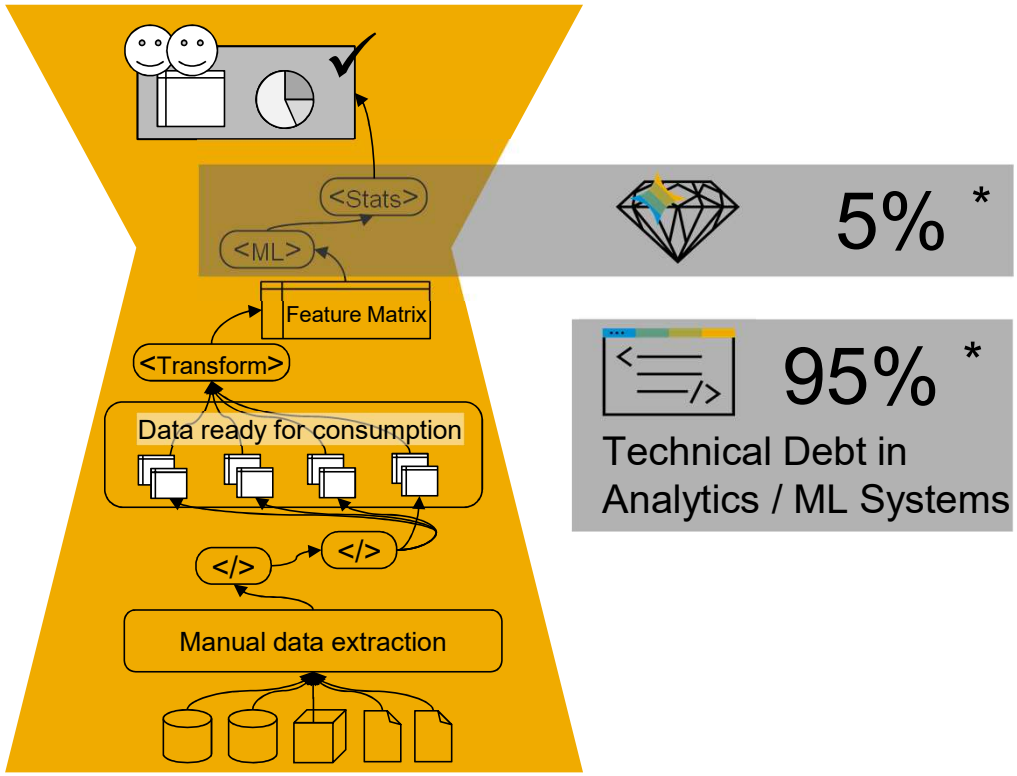
3 Grow from Pilot to Industrialization

SAP Data Management
From Pilot to Industrialization



Credits to Simon Wardley. Good start to read https://businessvalueexchange.com/blog/2017/04/19/designing-constant-evolution-pioneers-settlers-town-planners/

Example:
Typical Machine Learning System



5% *

95% *

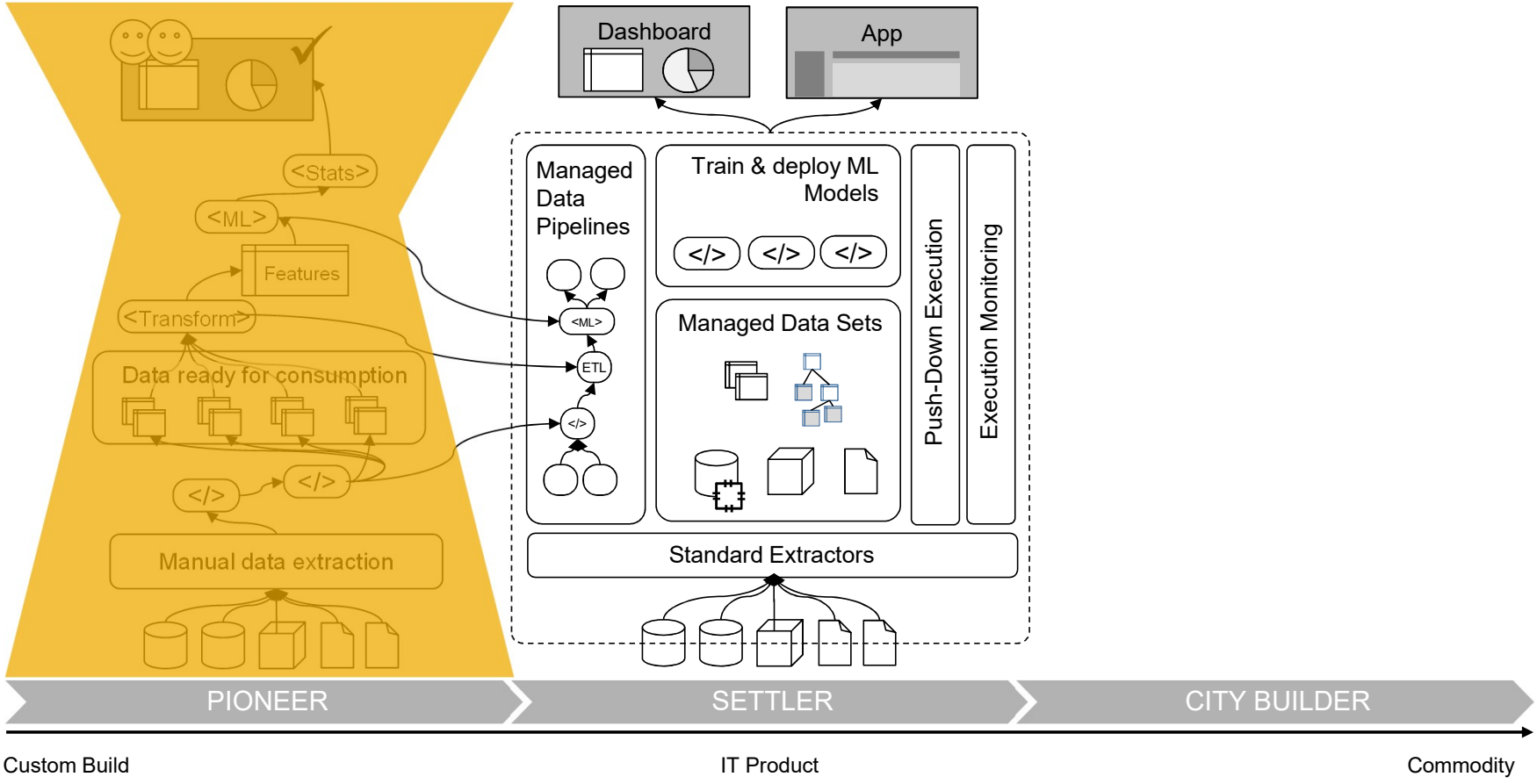
Technical Debt in Analytics / ML Systems

PIONEER | SETTLER | CITY BUILDER

Custom Build | IT Product | Commodity

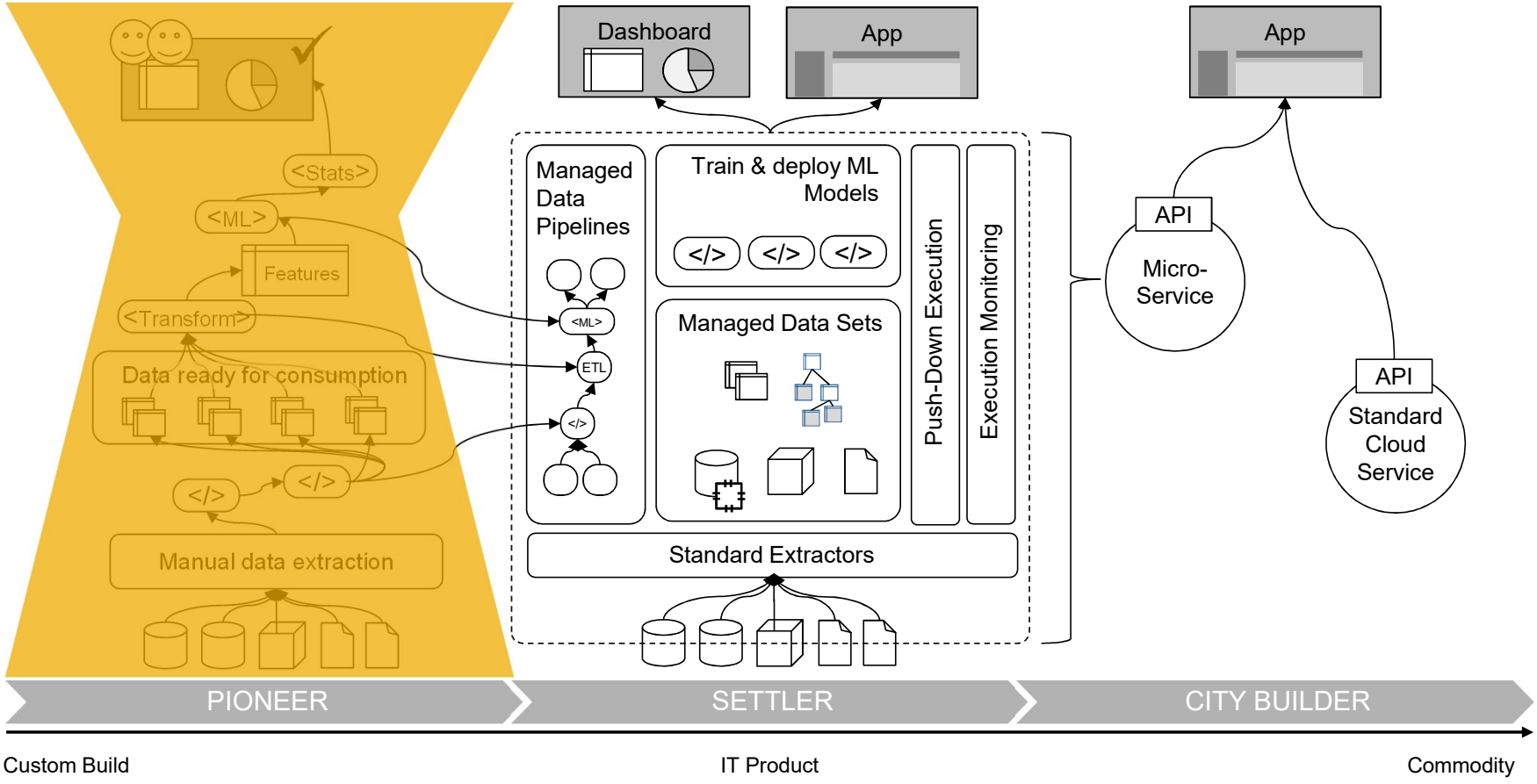
* Google Research Paper on ML Systems

Example:
Typical Machine Learning System



* Google Research Paper on ML Systems

Example:
Typical Machine Learning System

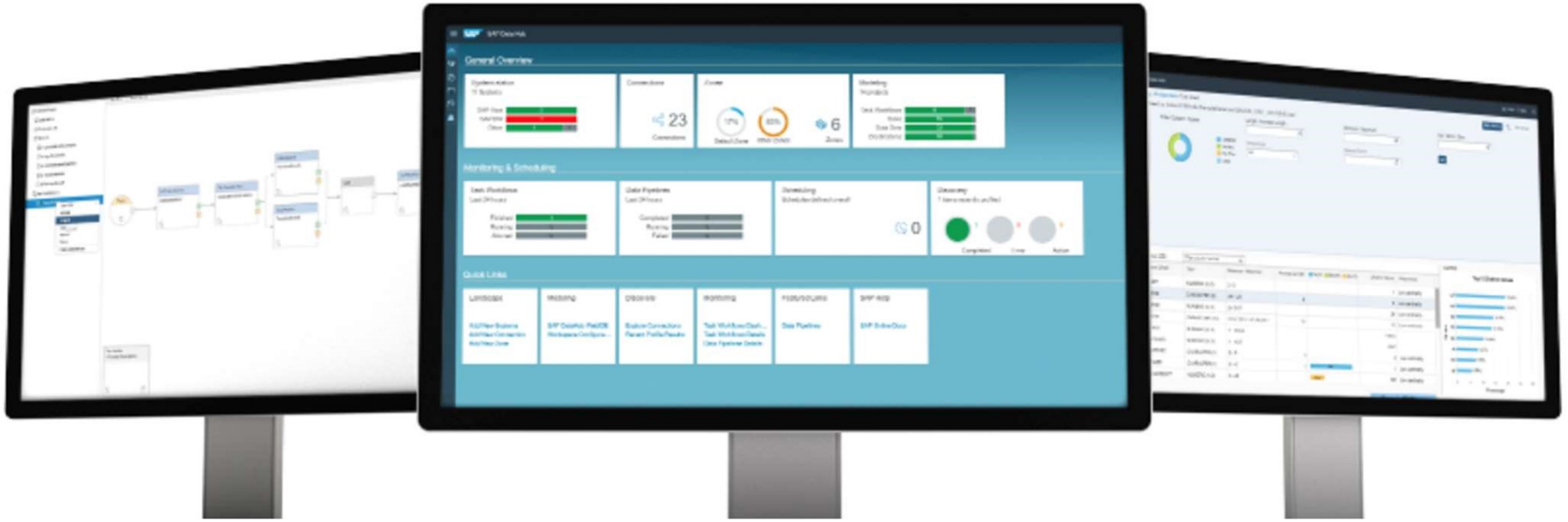


* Google Research Paper on ML Systems

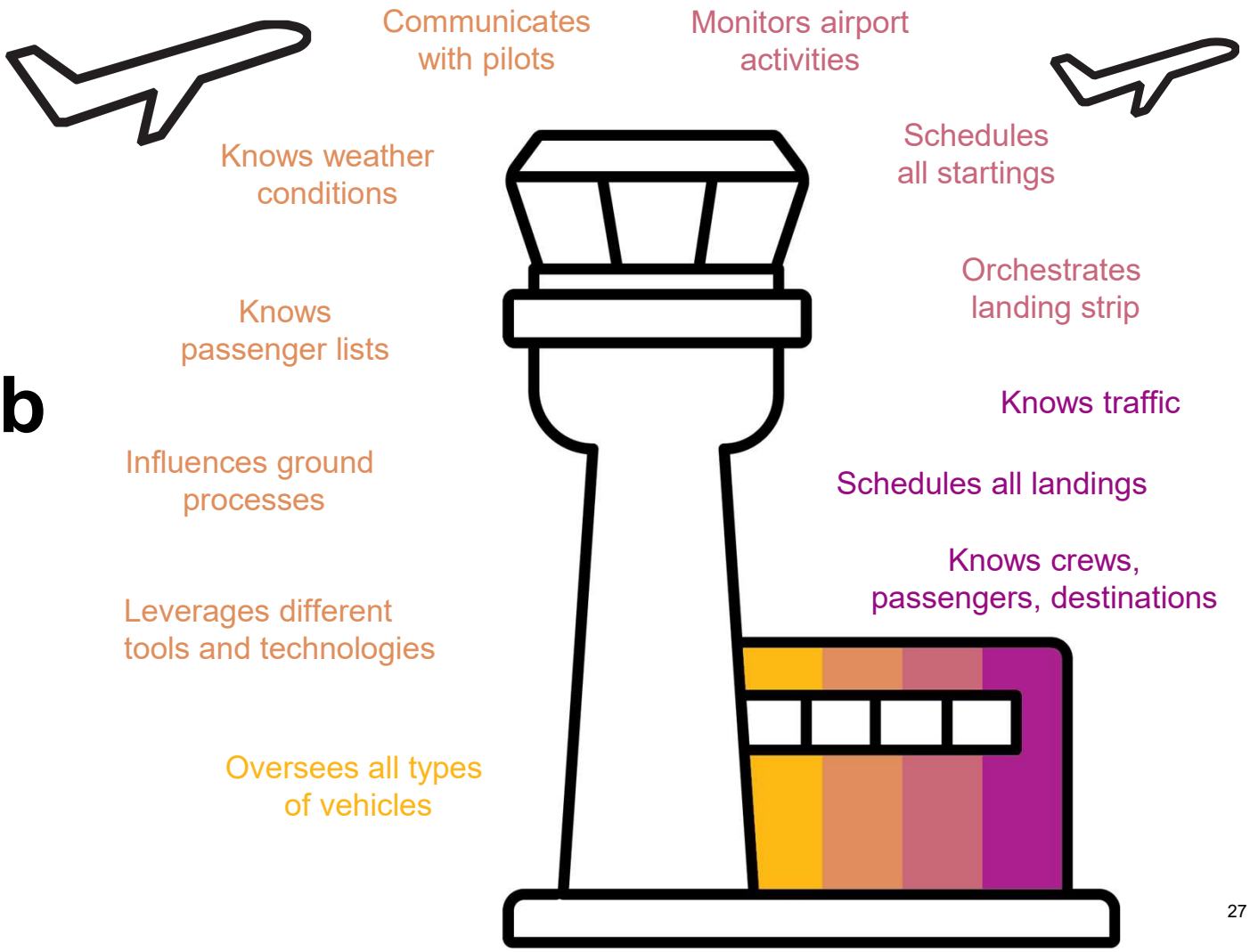
Plattform für Data Driven Applications

We can help you tackle all these challenges

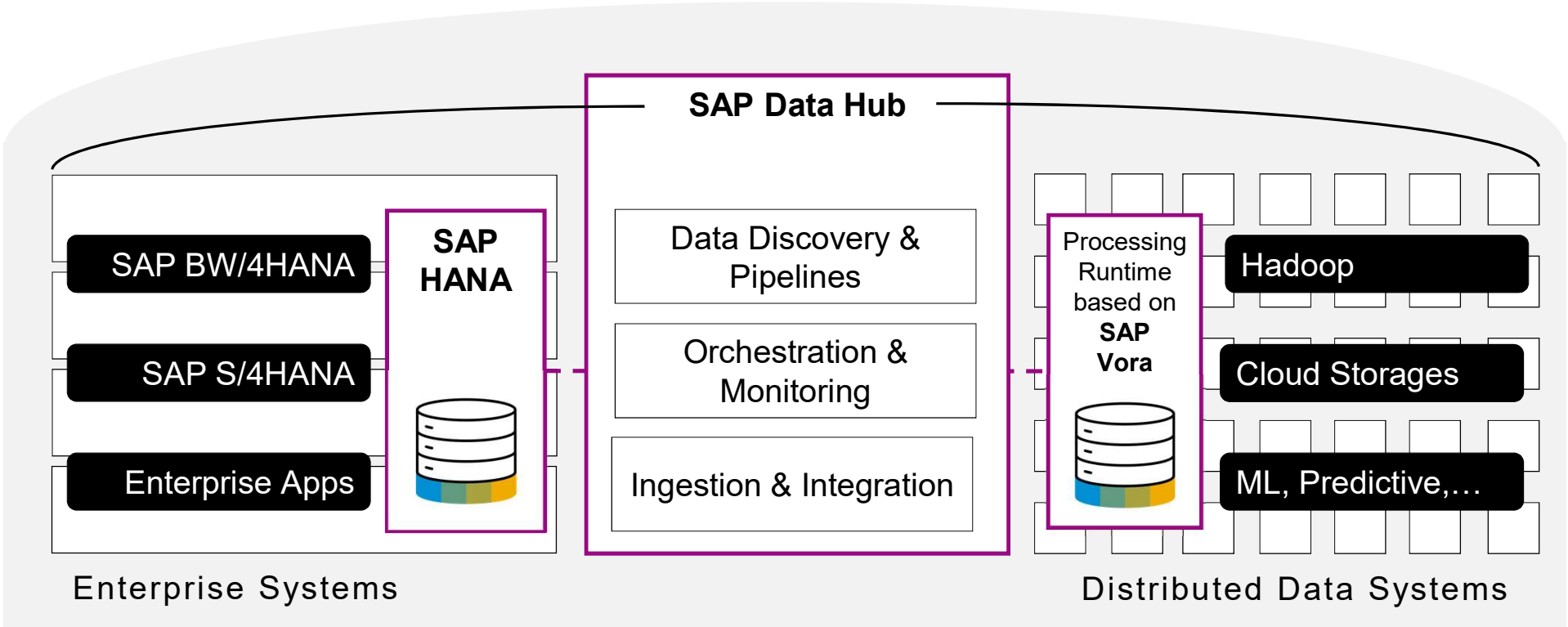
SAP Data Hub



SAP Data Hub



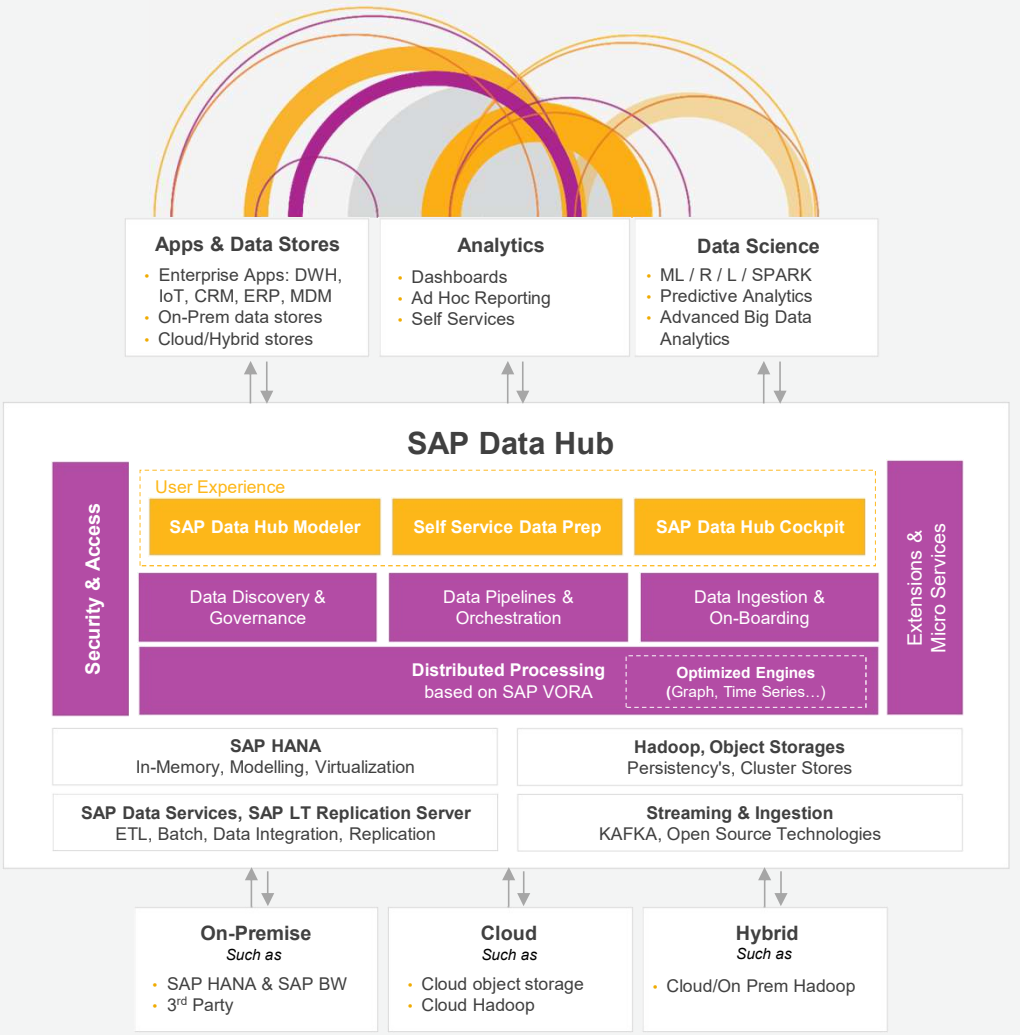
SAP Data Hub
Unifying Data Silos



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 Hub – consumption and enablement of data-driven apps

- Perform data discovery and profiling on your data from different data sources and systems
- Refine and enrich your data within the creation of data pipelines
- Schedule and execute data workflows and automate your DataOps actions
- Ensure proper governance and security management

SAP Data Hub is a data operations (DataOps) management solution that enables agile management of data in a diverse landscape across the organization.

This enterprise-ready solution provides governance and orchestration for data refinement and enrichment, using pipelining of many complex data processing operations, like machine learning (ML).

Combine data sources and find hidden patterns

Define, execute, and schedule your data pipeline visually

Process your landscape-wide data flows from end to end

How to use SAP Data Hub: Major use case scenarios

Rapidly integrate and use new data sources
Big Data warehouse use case

- Acquire new data sources with previously siloed data from traditional data warehouses, data marts, enterprise applications, and Big Data stores
- Combine all types of sources, including structured and unstructured data
- Smoothly move selected data sets across landscapes and rapidly take advantage of new data sources

Understand real-world performance
Internet of Things use case

- Tackle the challenge of deriving insight from vast quantities of raw data lacking semantic value and business context
- Unite data from messaging systems, cloud storage, and SAP's data management solutions and enterprise applications
- Use event-based pipelines scaling to execution of many pipelines in parallel at any time

Ease the cost and effort of data compliance
Enterprise information management (EIM) use case

- Move to centralized orchestration, data refining, scheduling, and monitoring to a holistic approach to data management
- Automate and audit data aggregation
- Enrich data sets in SAP Data Hub with metadata information by using the SAP Agile Data Preparation application

Use machine learning and predictive analytics
Data science use case

- Apply machine learning and predictive algorithms to any data set
- Operationalize and automate machine learning processes
- Insert machine learning and predictive processing into any scenario

And now what … ?

(here)

Vora GitHub Code Snipets (here)

HANA Academy on YouTube

Open SAP Course (here)

Self Learning

Workshop

SAP Tech Academy

Thank you.

Contact information:

Michael Probst
Solution Advisor
m.probst@sap.com
+41792777243