



# ML & Data Science with SAP Data Intelligence

## From isolated PoC to Intelligent Enterprise Application

Elisa Masetti, SAP Presales Data Scientist  
February 18, 2020

PUBLIC

## Challenge of Enterprise Data Science

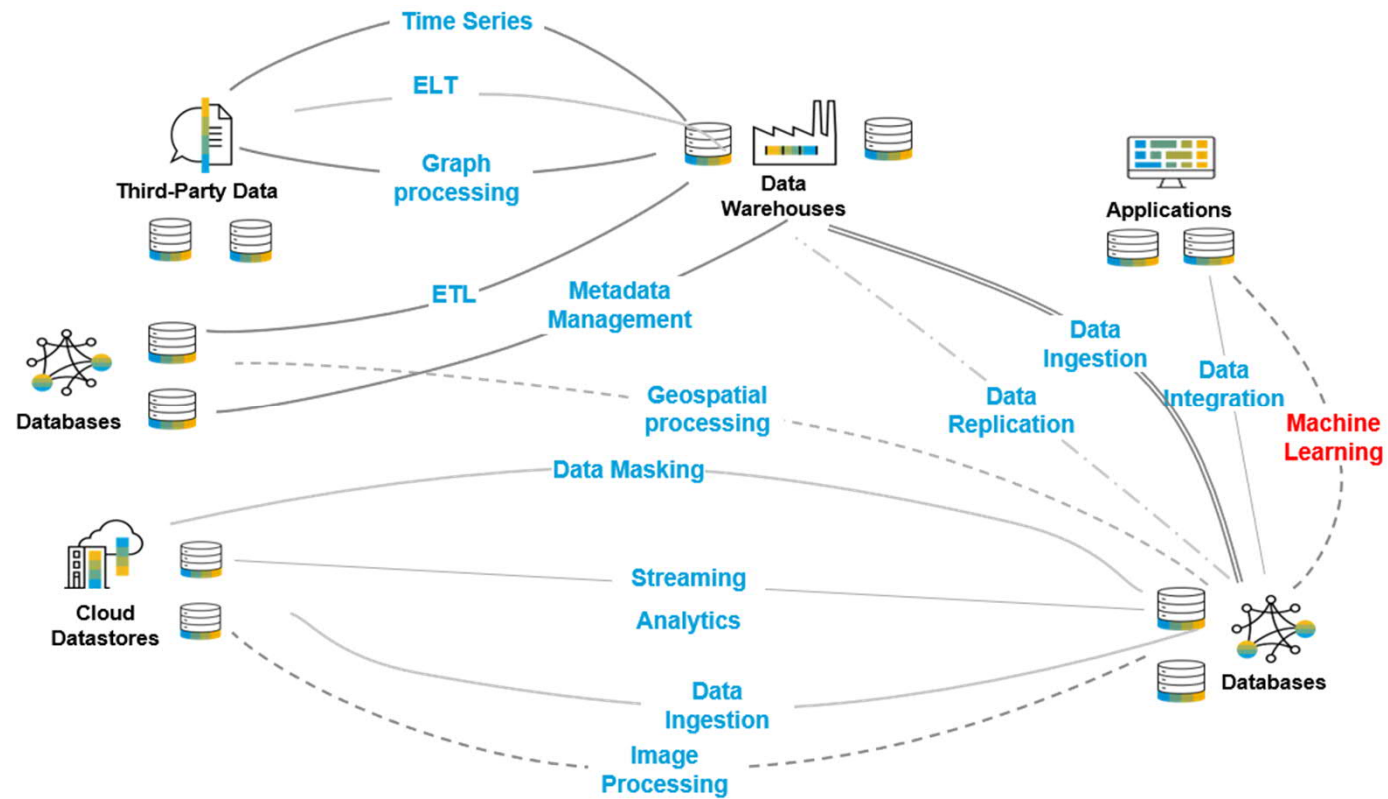
“Business managers surveyed by Gartner identified **time to value** as one of their biggest challenges [...] it takes an average of **52 business days** for their team **to build a predictive model**, and **longer to deploy into production.**”

Survey Analysis: Traditional Approaches Dominate Data and Analytics Initiatives, Gartner 2018

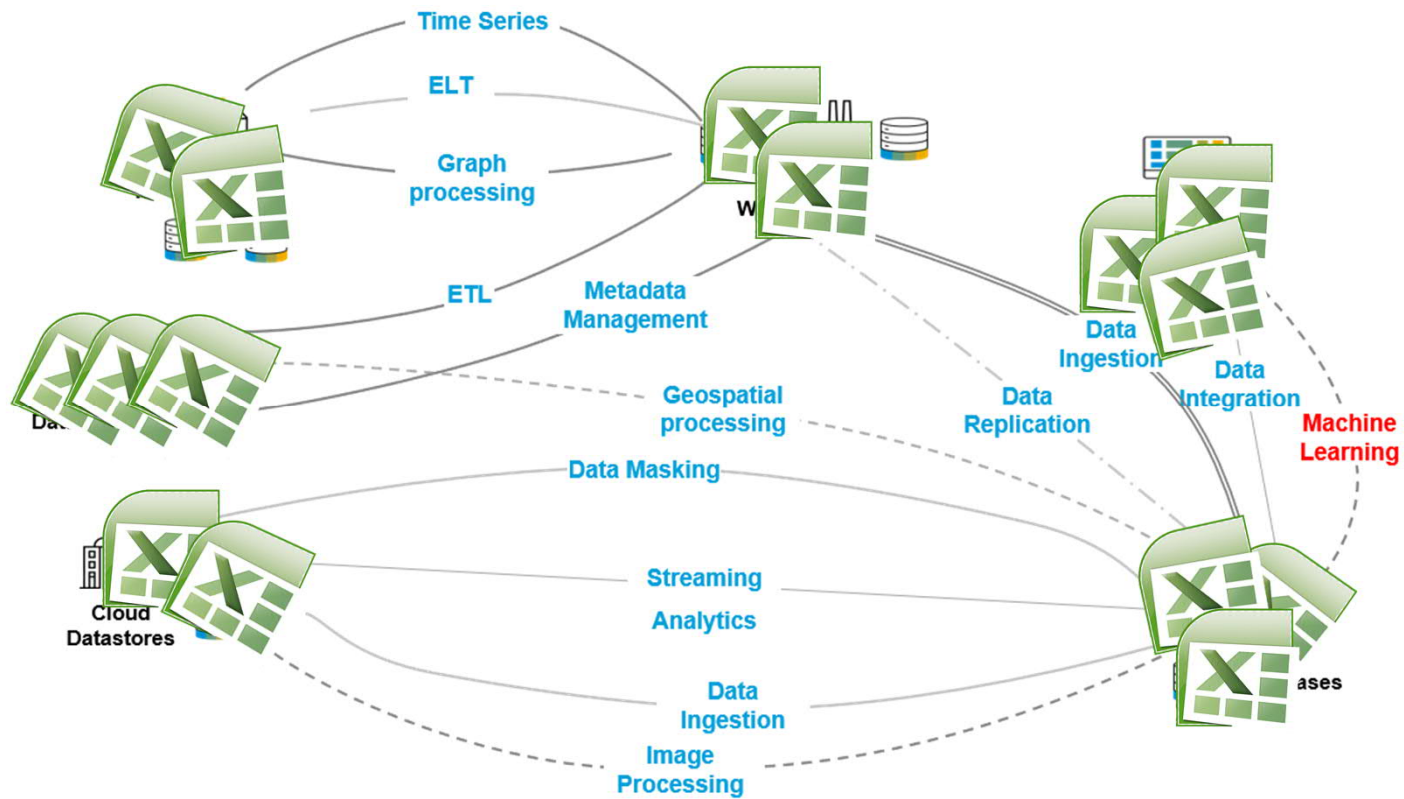
“ Over 60% of models developed to operationalize them were **never actually operationalized.**”

Survey Analysis: Data Science Team , Gartner 2018

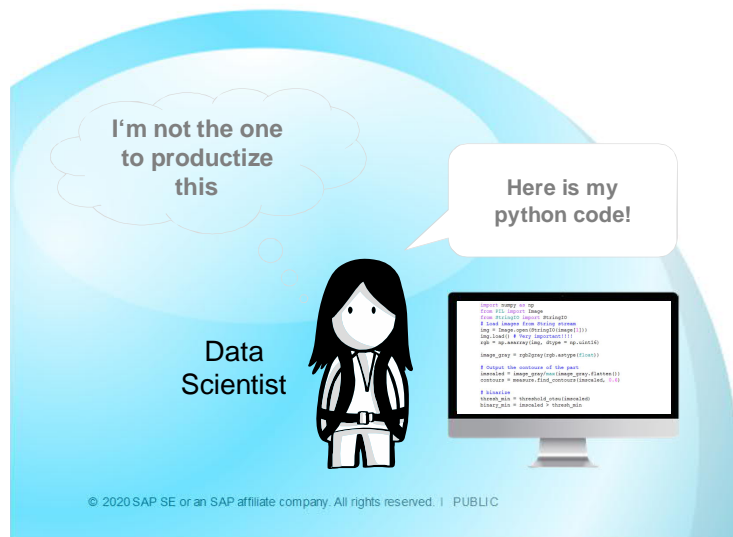
## WHY #1: Data Jam



## WHY #1: Data Jam

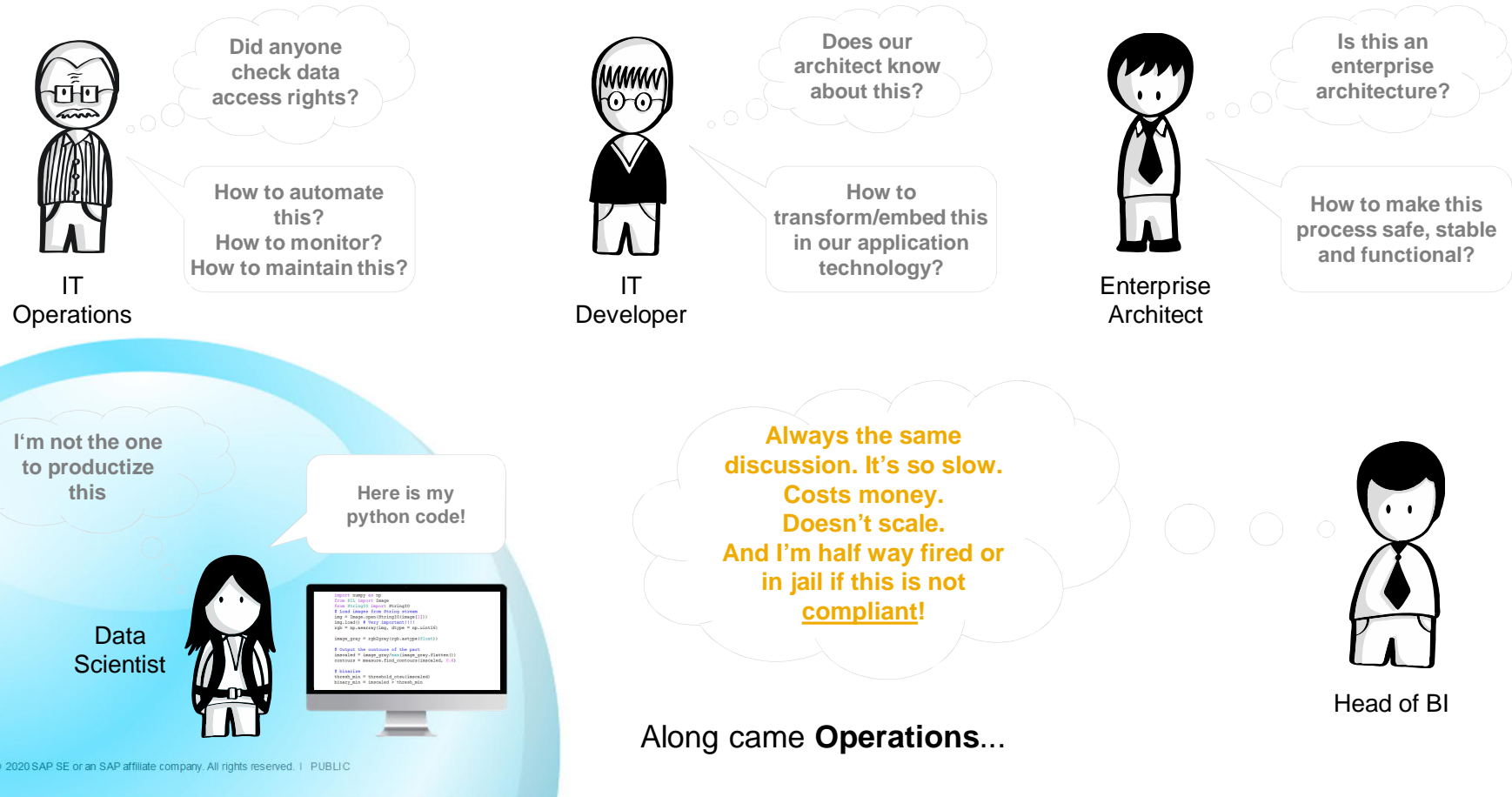


## WHY #2: Data Science lives in a “test tube”



Along came **Operations...**

## WHY #2: Data Science lives in a “test tube”

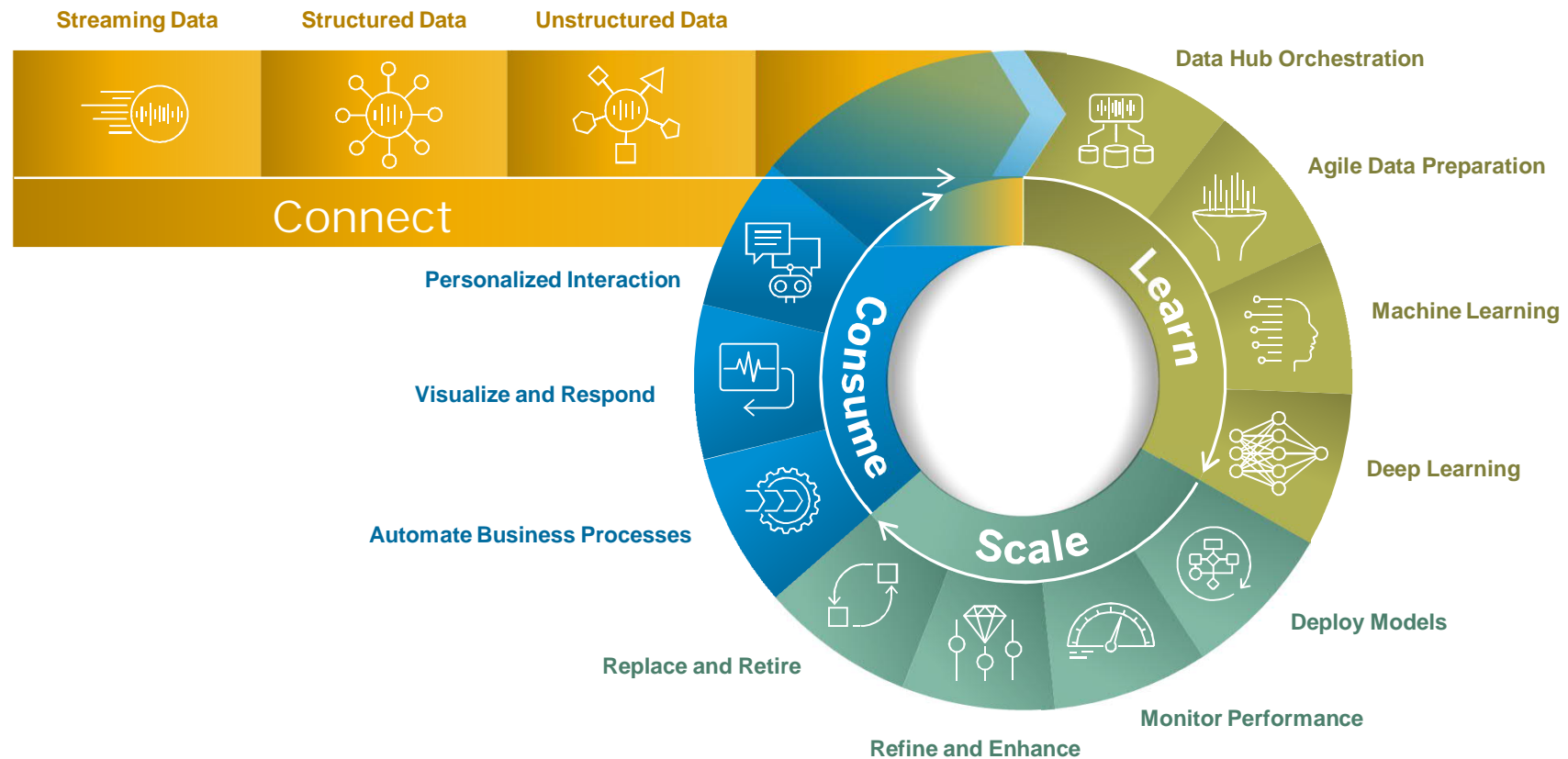




## **SAP Data Intelligence:** **Turn Machine Learning into an Enterprise Process**

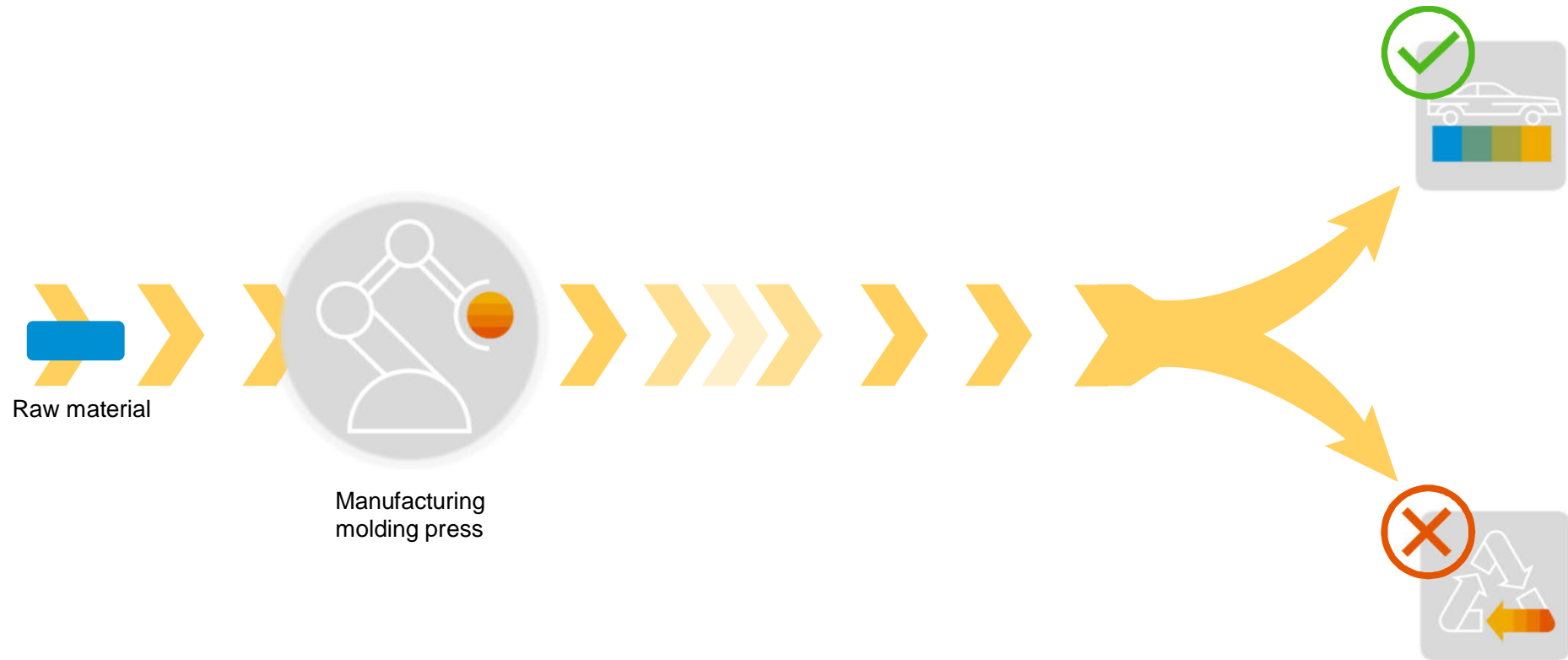


# Data Science Workflow Aided by **SAP Data Intelligence**

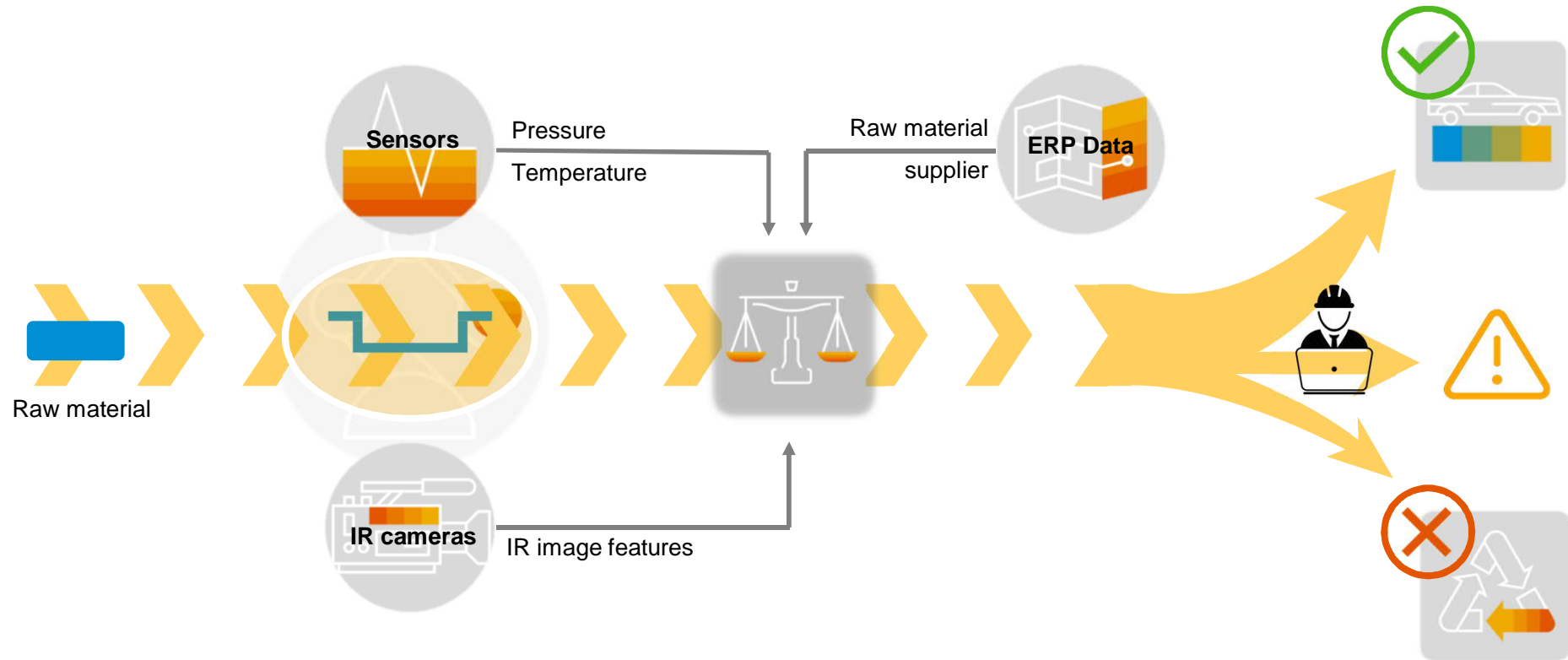




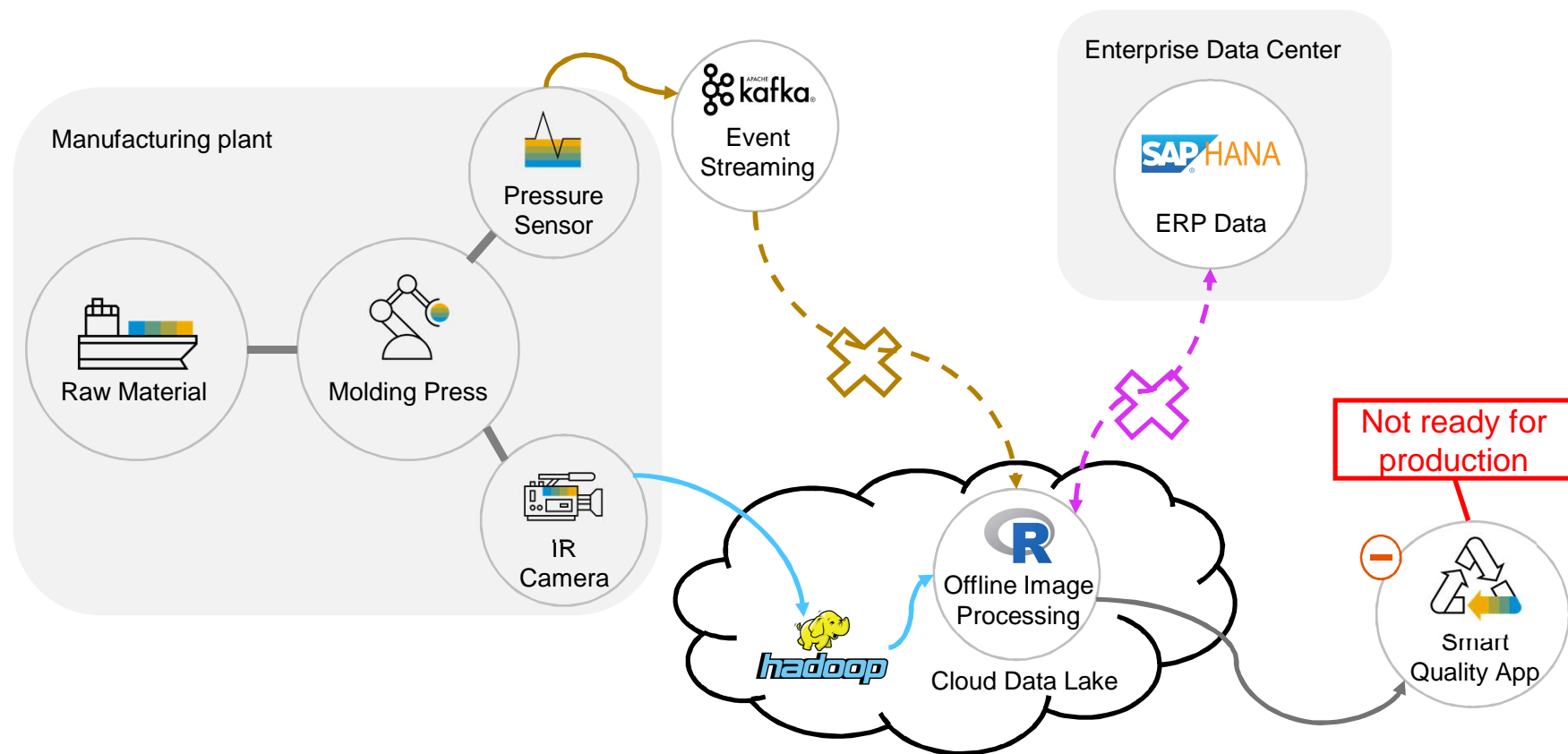
## The situation



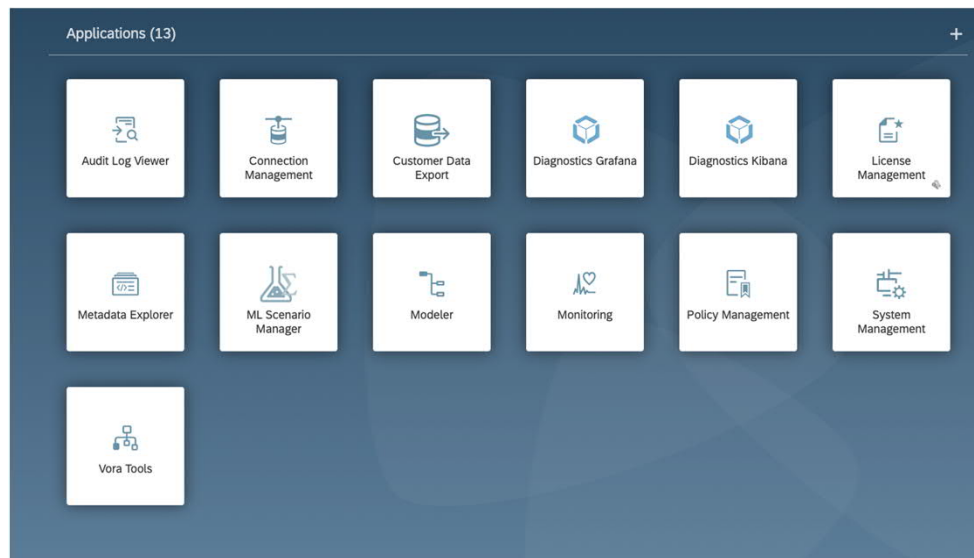
## The idea



## The unsuccessful prototype



## The right tool to make it productive - **SAP Data Intelligence**

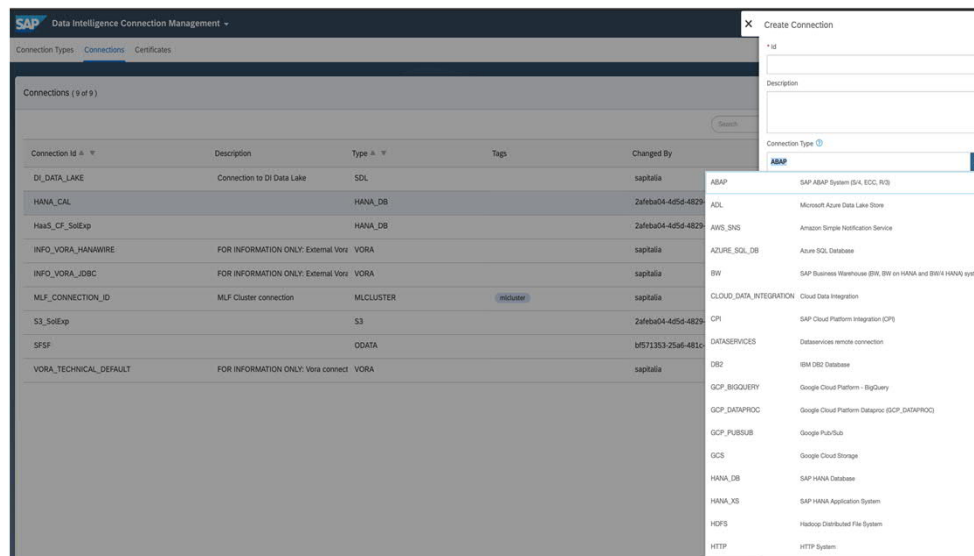


On-premise and consumable  
SAP Cloud Platform service for:

- Centralized Metadata Governance
- Data and Process Pipelining
- Distributed and Heterogeneous Data Processing
- Model Lifecycle Management

# Connection Management

Connect Learn Scale Consume

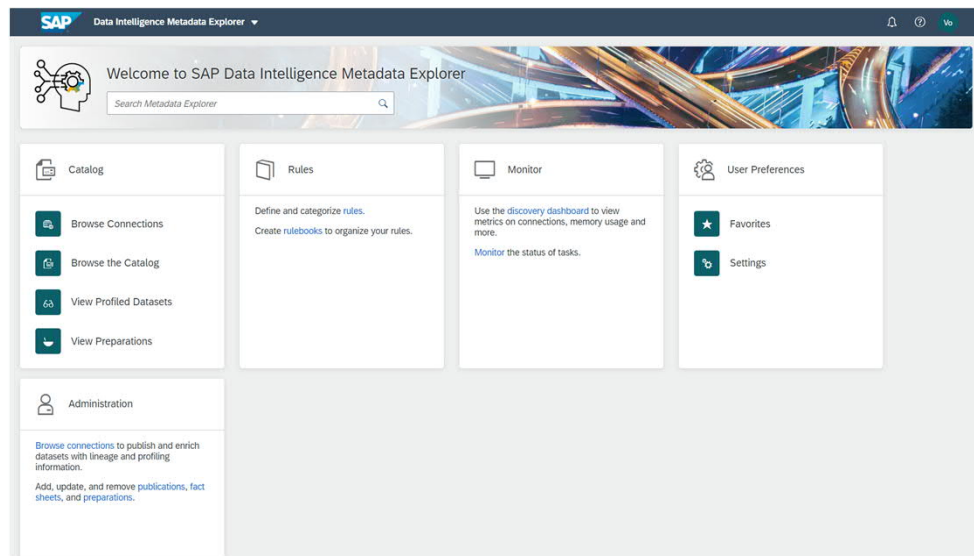


Connect to everything leveraging the access control defined in the source system

- Prebuilt connectors for SAP systems, DB2, MySQL, Oracle, S3, HDFS, OpenData and many more
- Possibility to extend to even more data sources with SAP Open Cloud Connectors and SDK

# Metadata Explorer

Connect [Learn](#) Scale Consume

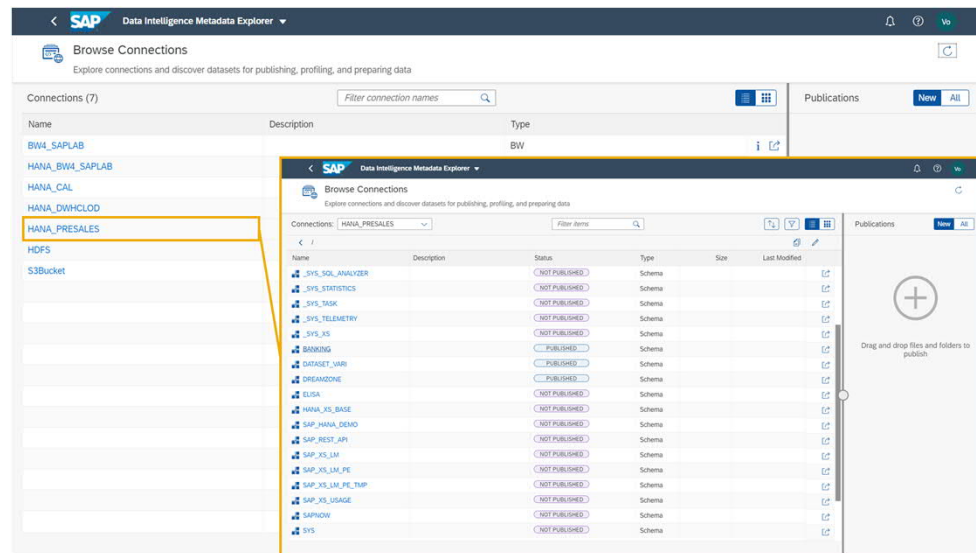


View, understand and organize your data across the IT landscape

- Data Discovery
- Data Profiling
- Data Preparation
- Data Quality Rules
- Metadata Cataloging

# Metadata Explorer - Data Discovery

Connect [Learn](#) Scale Consume

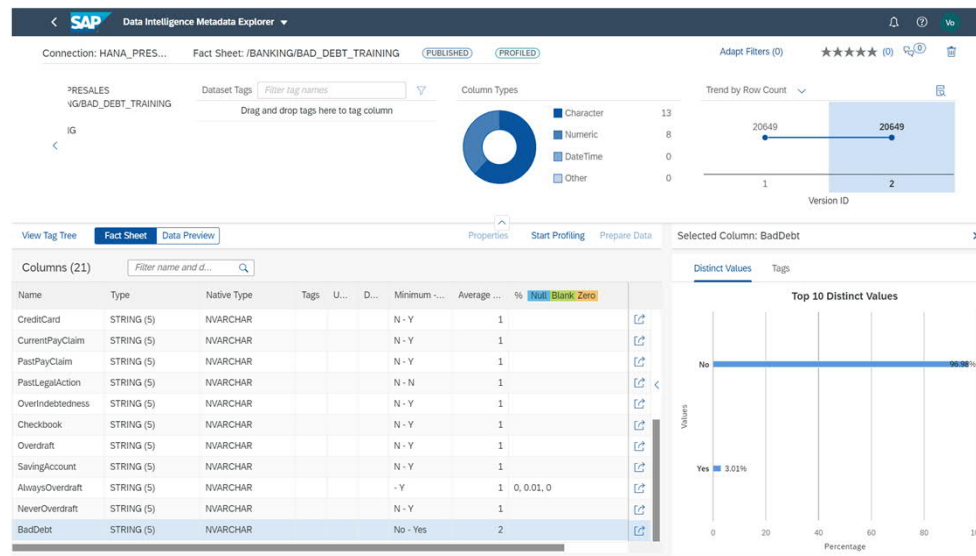


Explore data assets across the connected landscape and use the preview functionality to get a first understanding of your data



# Metadata Explorer - Data Profiling

Connect [Learn](#) Scale Consume



Retrieve more in depth information about your data in order to clarify the structure and understand anomalies

- Growth trend
- Values distribution across the dataset
- Percentage of null, blank and zero across the columns

# Metadata Explorer - Data Preparation

Connect [Learn](#) Scale Consume

The screenshot shows the SAP Data Intelligence Metadata Explorer interface. It features a table with columns: CustomerId, Age, Children, Gender, MaritalStatus, JobCode, JobHistory, Revenue, and MarketingSegment. The 'MaritalStatus' column is selected, and a sidebar on the right shows preparation actions like Change Case, Custom Action, Fill, Trim, Replace, Duplicate, Rename, Split, Combine, Convert To DateTime, Convert To Number, Clear, Reorder, and Remove. The bottom status bar indicates a sample size of 20 columns and 1000 rows.

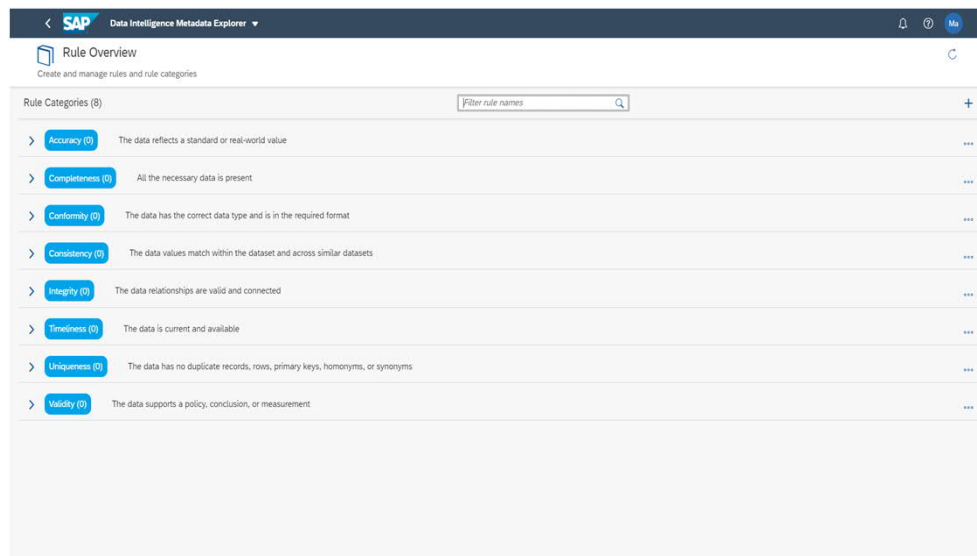
CustomerId	Age	Children	Gender	MaritalStatus	JobCode	JobHistory	Revenue	MarketingSegment
150344	23	0	F	Single	220	3	1667	1
78428	26	0	F	Single	101	3	1167	1
71592	30	0	M	Single	284	3	917	1
77275	27	0	M	Single	16	3	1042	1
84507	23	0	M	Single	87	3	1083	1
78091	31	0	F	Single	229	3	990	2
83867	26	0	M	Single	101	3	1000	1
73041	28	0	M	Single	35	3	2083	0
124170	26	0	F	Single	21	3	765	1
76948	28	0	F	Single	114	3	861	1
189278	25	0	F	Single	333	3	1125	1
115552	24	0	F	Single	321	3	842	1
81103	48	0	M	Single	17	3	1350	1
84214	25	0	F	Single	220	3	683	1
84477	22	0	F	Single	300	3	2071	1
83059	27	0	M	Single	16	3	1042	1
185394	31	0	F	Single	229	3	990	2
70501	21	0	M	Single	20	3	1850	1
70447	26	0	F	Single	101	3	833	1
77376	30	0	M	Single	284	3	917	1
72045	25	0	F	Single	96	3	883	1

## End-to-end self-service data preparation for business users

- Access a dataset or upload a flat file to prepare the data based on a sample dataset
- Filter, replace, combine, convert and more in just a few clicks without any technical scripting skills
- Save the results in one of your connected systems

# Metadata Explorer - Data Quality Rules

Connect **Learn** Scale Consume

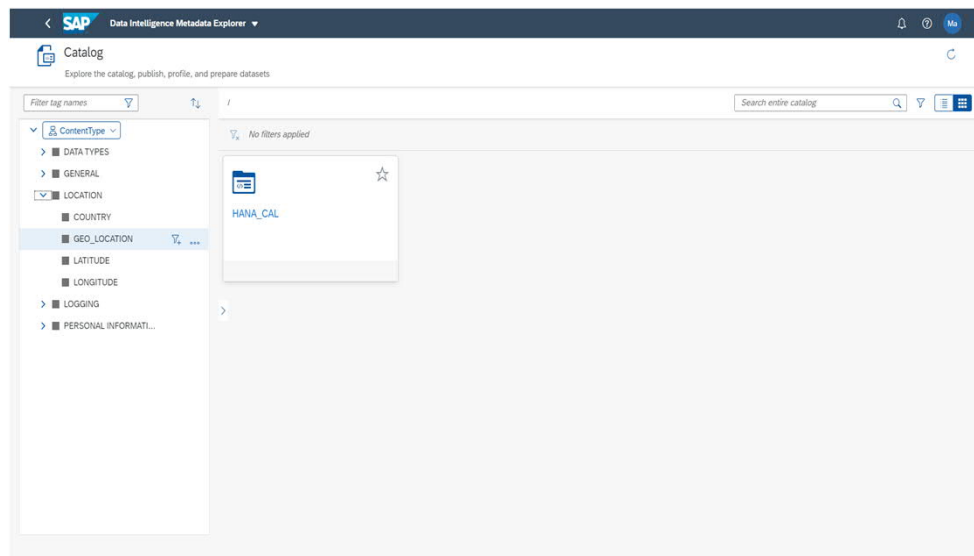


Understand the quality of your data to help you in data preparation and modeling

- Define quality rules through pre-defined or custom formats
- Define thresholds and rulebooks
- Run rules against datasets
- Build data quality dashboards

# Metadata Explorer - Metadata Cataloging

Connect [Learn](#) Scale Consume



Define, govern and manage your metadata assets across enterprise systems

- Automatically index metadata based on content type and enabled text based search
- Use labels at dataset and column level to classify your data assets and use labels as filters in your data catalog

# ML Scenario Manager

Connect Learn Scale Consume

The screenshot displays the SAP ML Scenario Manager interface. The main panel, titled 'ML Scenarios', contains a search bar and a date range filter set to 'Jan 10, 2020 - Jan 20, 2020'. Below this is a table listing four scenarios:

Name	Business Question	Changed By	Changed On
Identify Species from Iris Dataset <small>Created with AutoML</small>		b/571353-25a6-481c-a302-49ecb01014bf	1/20/20, 11:23 AM
Marathon time prediction	Basic Regression example	39b1d9bf-d423-4a8c-afbe-c574d484471b9	1/16/20, 10:58 AM
Human Resources - Employee Attrition		b/571353-25a6-481c-a302-49ecb01014bf	1/15/20, 5:41 PM
Demo Similarity Scoring		2af6ba04-4d5d-4829-9a09-df1ba0d4a2	1/10/20, 11:59 AM

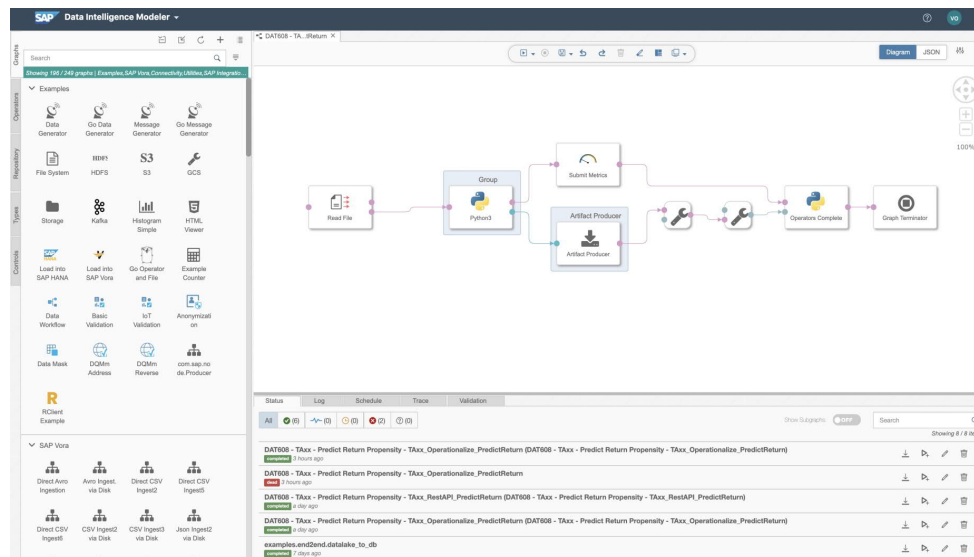
The right-hand panel, titled 'Version History', shows the version details for the selected 'Demo Similarity Scoring' scenario. It lists two versions:

- Version 2** by 2af6ba04-4d5d-4829-9a09-df1ba0d4a2, dated 1/10/20, 12:10 PM.
- Version 1** by 2af6ba04-4d5d-4829-9a09-df1ba0d4a2, dated 1/10/20, 11:59 AM.

One single interface to see all AI models across the organization and their versioning in a unified workspace

# Modeler & Monitoring

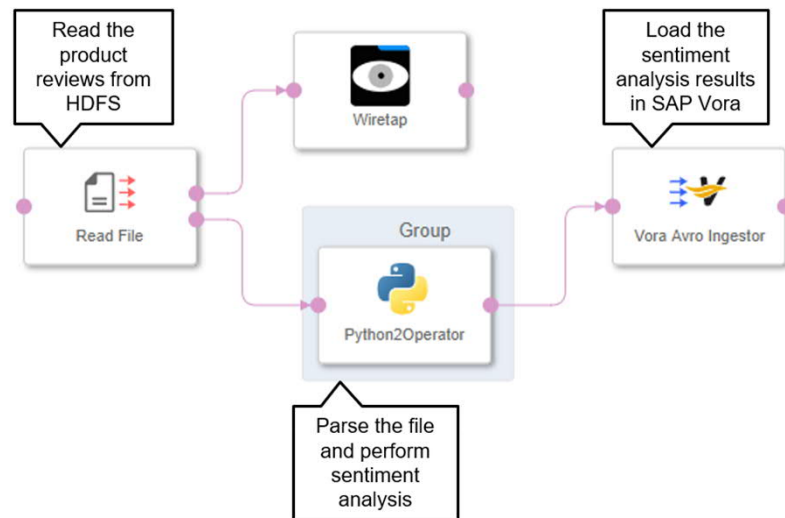
Connect Learn Scale Consume



- Rapidly develop complex data workflows as chains of operators, replacing coding and scripting with an intuitive graphical user interface
- Schedule, monitor and observe the executions of your workflows

# Modeler - Pipelines

Connect Learn Scale Consume



## Extensible

- Standard operators
- Custom/Partner operators with the possibility to use as base Python, R, JavaScript, Node.js, Go, ABAP & more

## Scalable

- Distributed
- Containerized

## Re-Usable

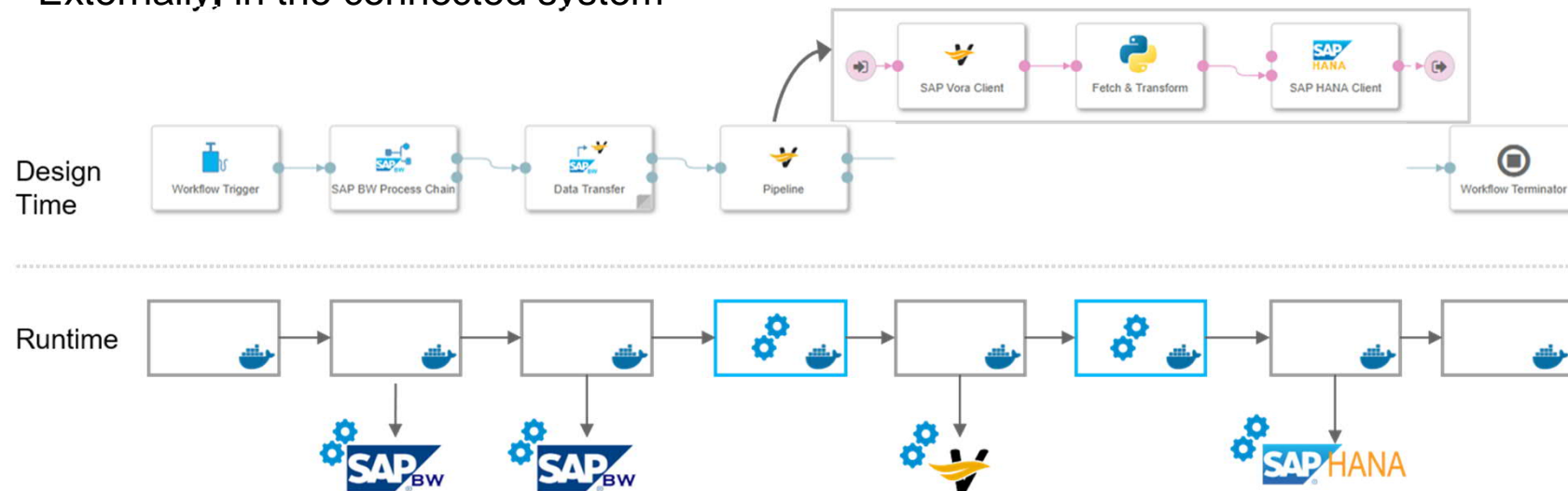


## Modeler - Orchestration

Connect Learn **Scale** Consume

The heavy lifting / logic execution happens either:

- Internally, in the operator container
- Externally, in the connected system



## Model Creation

Connect [Learn](#) Scale Consume



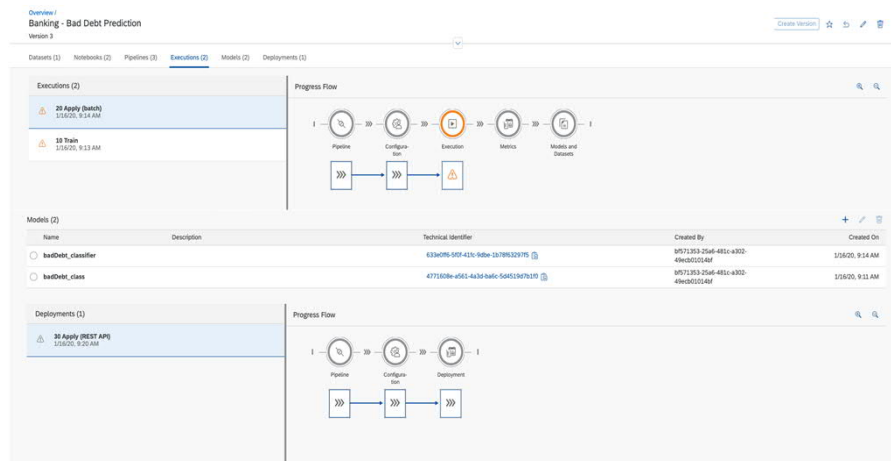
Keep total flexibility for data scientists

- JupyterLab integration to ML Scenario Manager
- Use your preferred language to deal with data exploration and model creation and all the latest open-source libraries
- Leverage all HANA predictive libraries through the configurable operator, in an R or Python-like syntax or by direct scripting
- Consume SAP Leonardo functional services using the pre-defined operator

# ML Scenario Manager & Modeler

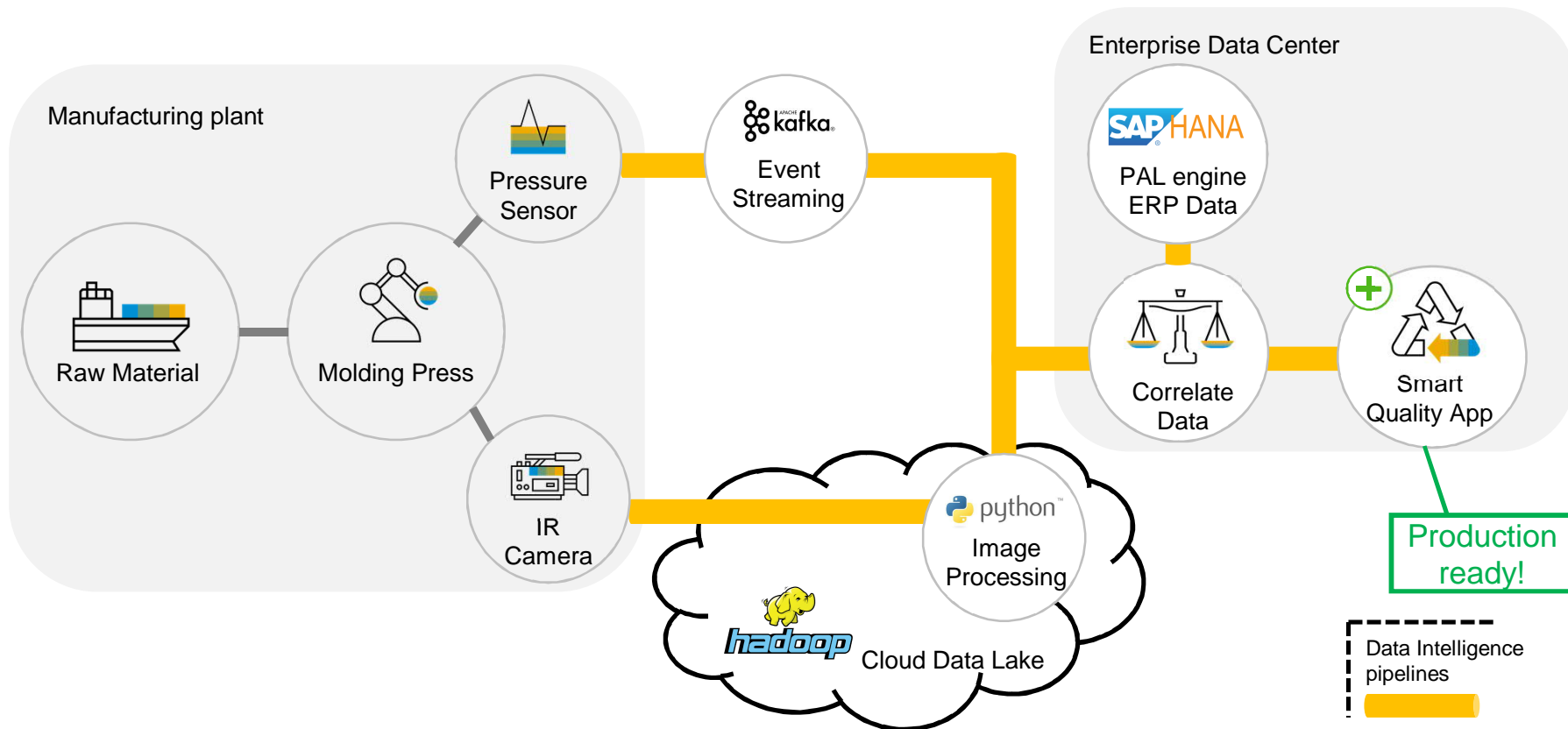
## Model Lifecycle Management

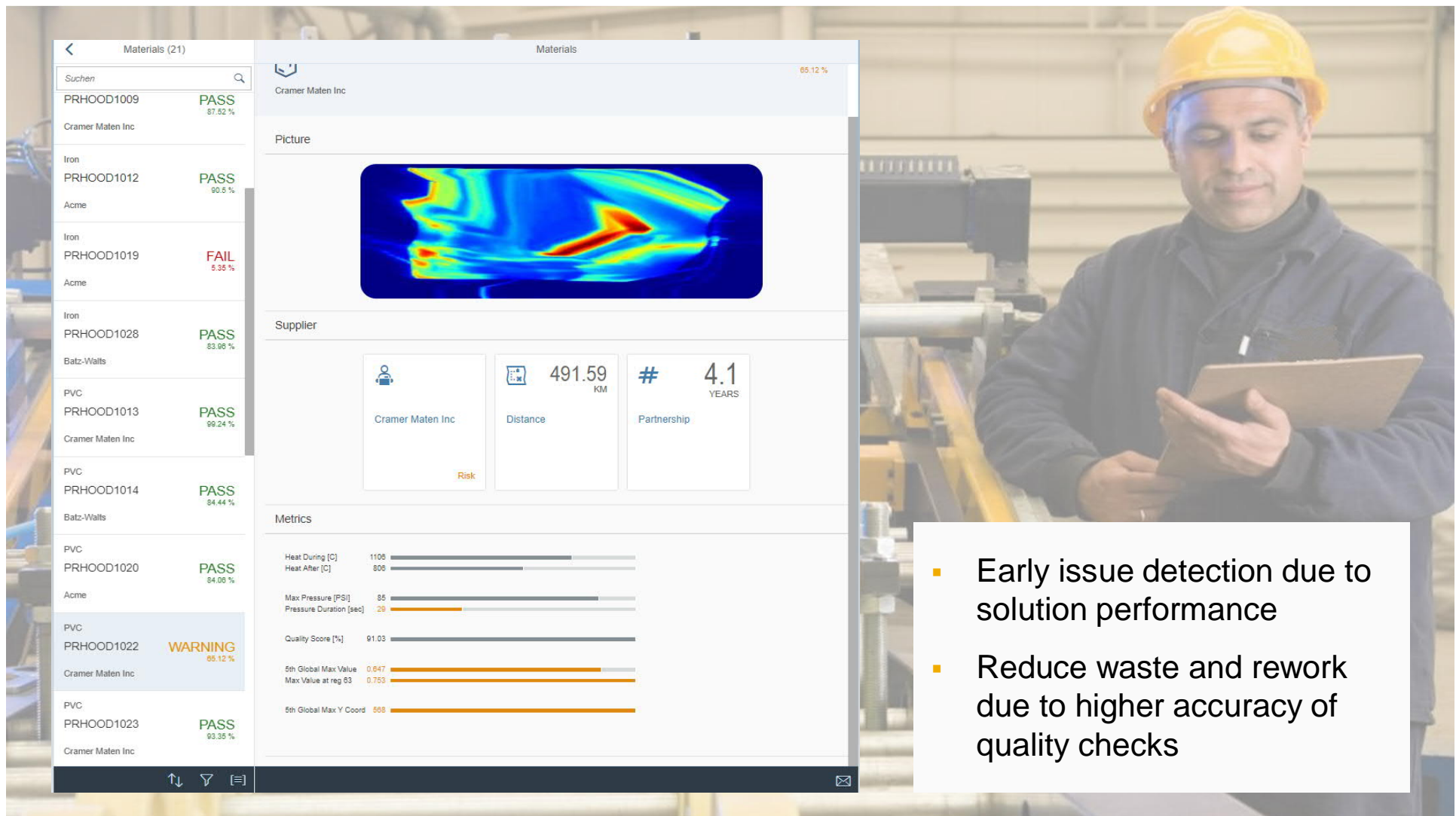
Connect Learn Scale Consume



- Build pipelines for training and inference starting from scratch or using templates, and link them to the related ML scenario
- Retrieve model metrics and related artifacts from pipelines executions
- Deploy your models as RESTful APIs to make them available in enterprise applications
- Manage and monitor all of this centrally through the ML Scenario Manager

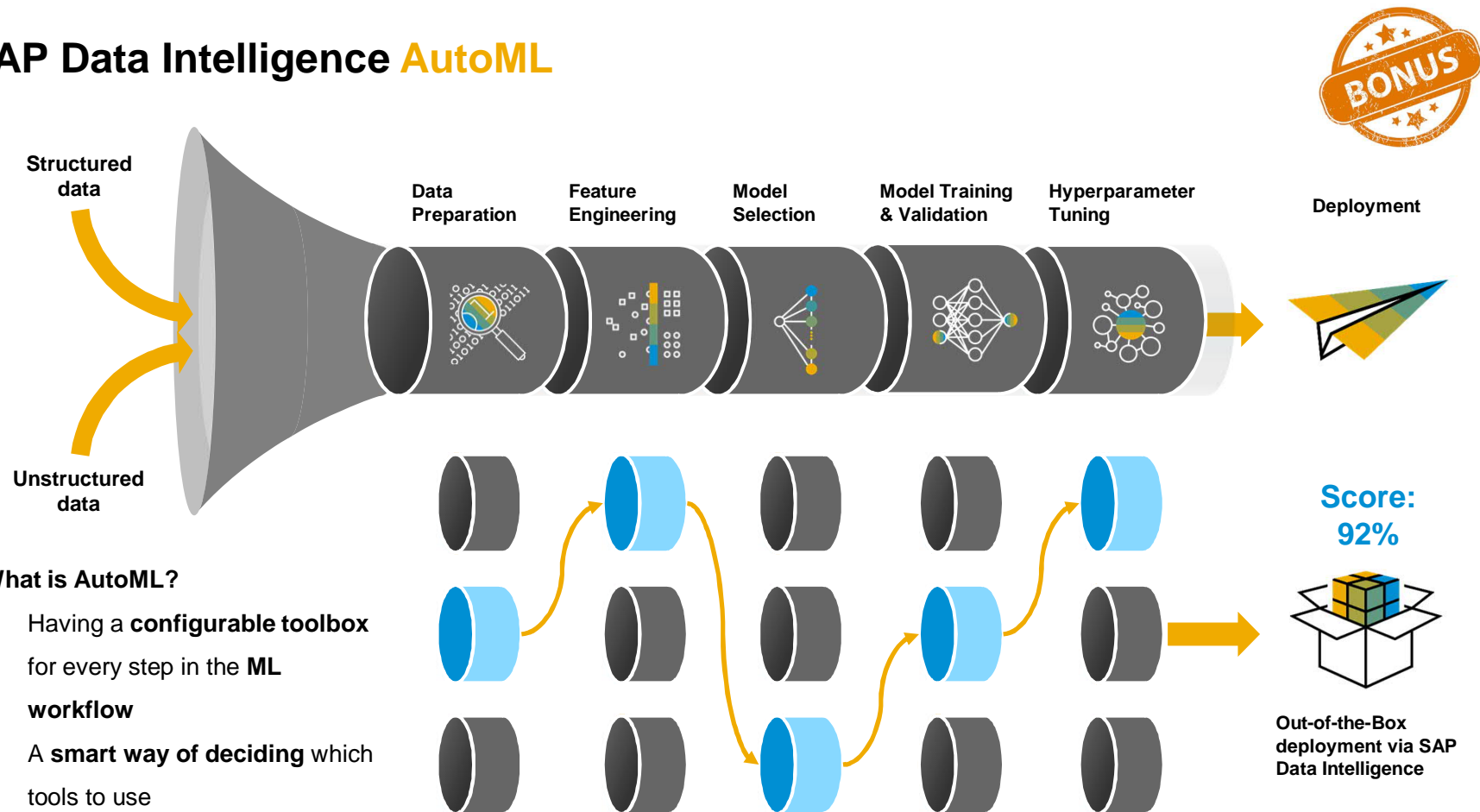
## The new as-is: Predictive quality workflow in manufacturing plant





- Early issue detection due to solution performance
- Reduce waste and rework due to higher accuracy of quality checks

# SAP Data Intelligence **AutoML**



# Thank you

**Elisa Masetti**

[elisa.masetti@sap.com](mailto:elisa.masetti@sap.com)



Follow us



[www.sap.com/contactsap](http://www.sap.com/contactsap)

© 2020 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

The information contained herein may be changed without prior notice. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have 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 SE's or its affiliated companies' strategy and possible future developments, products, and/or platforms, directions, and functionality are all subject to change and may be changed by SAP SE or its affiliated companies 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. 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, and they should not be relied upon in making purchasing decisions.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

See [www.sap.com/copyright](http://www.sap.com/copyright) for additional trademark information and notices.

