



DATA INTELLIGENCE PLATFORM:

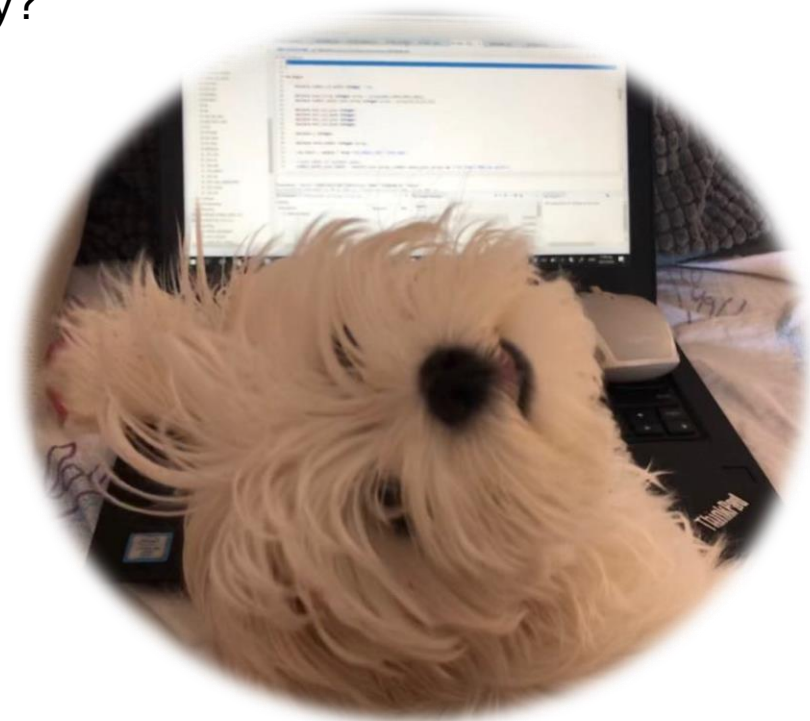
From Algorithms
to Intelligent
Business Processes



Dr. Agamemnon Baltagiannis

Data Science Present & Future

What is changing as the technology hype enters its maturity?



“The factory of the future will have only two employees, a man and a dog. The man will be there to feed the dog. The dog will be there to keep the man from touching the equipment”

Warren Bennis

AI ADOPTION ACROSS INDUSTRIES



47% of 2135 respondents surveyed by McKinsey for the 2018 Global Survey AI Adoption by Industry and Function said they have implemented **at least one AI technology**

Source: AI adoption advances, but foundational barriers remain, McKinsey report, 2018, URL

Business functions in which AI has been adopted, by industry,¹ % of respondents

	Service operations	Product and/or service development	Marketing and sales	Supply-chain management	Manufacturing	Risk	Human resources	Strategy and corporate finance
Telecom	75	45	38	26	22	23	17	15
High tech	48	59	34	23	20	17	21	17
Financial services	49	26	33	7	6	40	9	14
Professional services	38	34	36	19	11	15	16	11
Electric power and natural gas	46	41	15	14	19	14	15	14
Healthcare systems and services	46	28	17	21	9	19	18	13
Automotive and assembly	27	39	15	11	49	2	8	6
Travel, transport, and logistics	51	34	32	18	4	4	2	3
Retail	23	13	52	38	7	9	8	0
Pharma and medical products	31	31	27	13	28	3	6	4

¹This question was asked only of respondents who said their organizations have piloted or embedded at least 1 AI capability in 1 or more functions or business units. Respondents who answered "don't know" or "none of the above" are not shown. For telecom, n = 77; for high tech, n = 215; for financial services, n = 306; for professional services, n = 221; for electric power and natural gas, n = 54; for healthcare systems and services, n = 67; for automotive and assembly, n = 120; for travel, transport, and logistics, n = 55; for retail, n = 46; and for pharma and medical products, n = 65.

McKinsey&Company



IS A DATA SCIENTIST STILL A SCARCE RESOURCE?

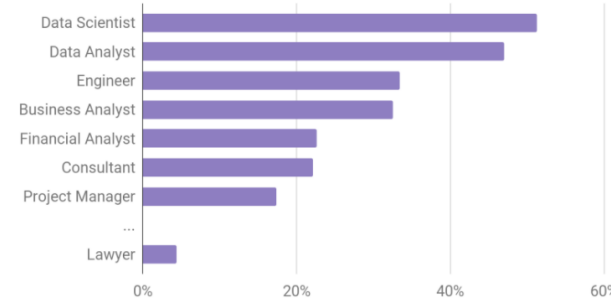


As of 2017, **annual enrollments in data** [in online learning platforms such as Coursera or courses taught by top university and industry partners including Stanford, Princeton, and Google] **exceeded enrollments in any other subject area**

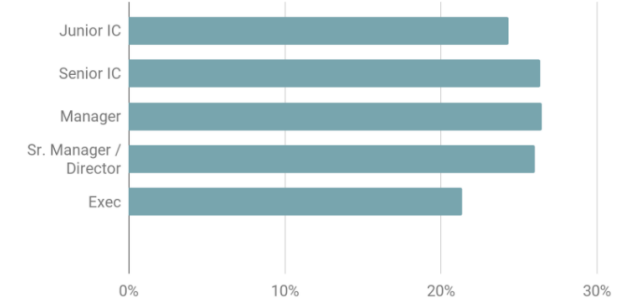
Source: *Data Skills: They're Not Just for Data Scientists*, 2018, URL

Share of Learning in Data, Coursera 2017

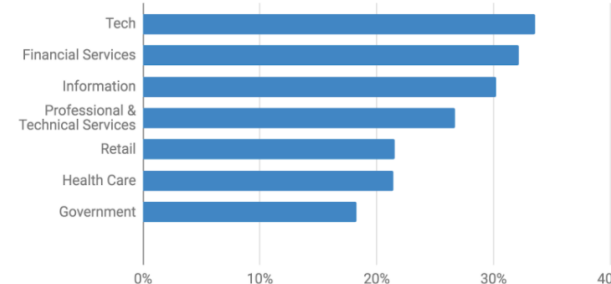
By Role



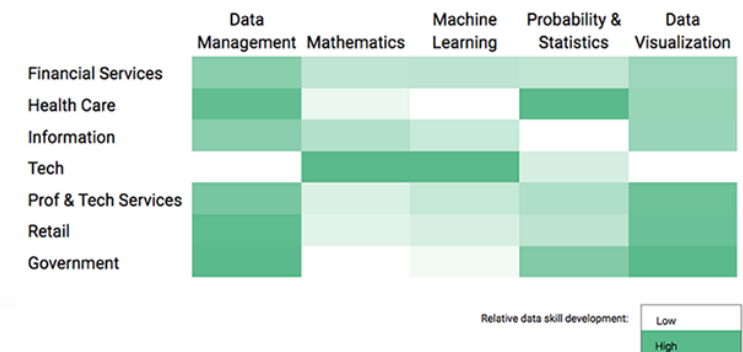
By Level



By Industry



Data Skill Development by Industry



BIGGEST CHALLENGES OF AI

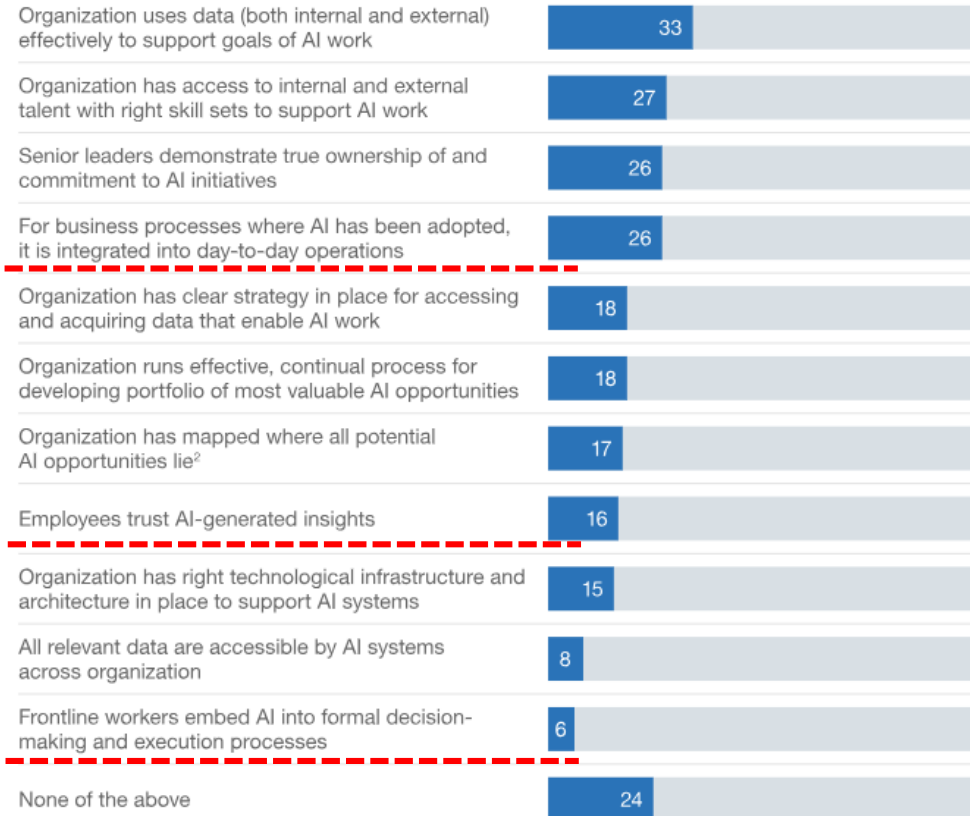


Just **17%** of respondents say their companies have mapped out where, across the organization, all **potential AI opportunities** lie.

And **only 18%** say their companies have a clear **strategy in place** for sourcing the data that enable AI work.

Source: AI adoption advances, but foundational barriers remain, McKinsey report, 2018, URL

Core AI practices in place at organizations,¹ % of respondents



¹This question was asked only of respondents who said their organizations have piloted or embedded AI in 1 or more functions or business units, and they were asked to select all practices that are in place. Respondents who said "don't know" are not shown; n = 1,646.

²Including required level of investment, difficulty of implementation, and potential value at stake.

McKinsey&Company



THE FUTURE OF DATA SCIENCE... (1/2)



Immense Data accumulation: In May of 2012 we uploaded approximately 72 hours of video to YouTube **every minute**. Five years later that number had skyrocketed to 400 hours every minute, which equates to 65 years of video—everyday. Facebook is another huge data generator, with over 2 billion monthly users as of 2017, and generating over **4 million likes every minute**.



IoT and Edge Computing: Every piece of technology is now (or will soon be) a data generator. Currently, we have **approximately 7 billion connected IoT devices** globally, and this number is predicted to increase to **21.5 billion in 7 years**.



Cloud Computing: We'll use cloud-based solutions, even for the simplest projects. Laptops are not evolving fast enough to keep the pace with the **required computational power**.



Data Science/ Machine Learning Automation tools will become regular **commodities** that you don't even need to know how to code to use them



Data cleaning and Feature engineering are still a **bottleneck** for data scientists. Focus will be on ways of advancing and automating modeling, hyperparameter tuning, and cross-validation of models



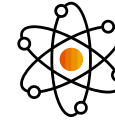
THE FUTURE OF DATA SCIENCE... (2/2)



Ethical awareness will come more to the spotlight. Just as an example, USA courts are already using machine learning algorithms to sentence criminals. If the data is biased, will the algorithm be biased as well?



Interpretable AI: Data Science will have to care even more about **explaining machine learning algorithms**. Explaining the decisions of the so-called machine learning black boxes will be mandatory for some industries.



Data Science Specialization: Netflix, has nine Data-related roles:

Business Analyst	Data Analyst
Quantitative analyst	Analytics engineer
Data engineer	Data scientist
Machine learning scientist	Research scientist

Source: Data Science Trends for 2018, 2017, URL and Data Science Trends for 2019, 2018, URL

Source: Beyond Interactive: Notebook Innovation at Netflix, 2018, URL



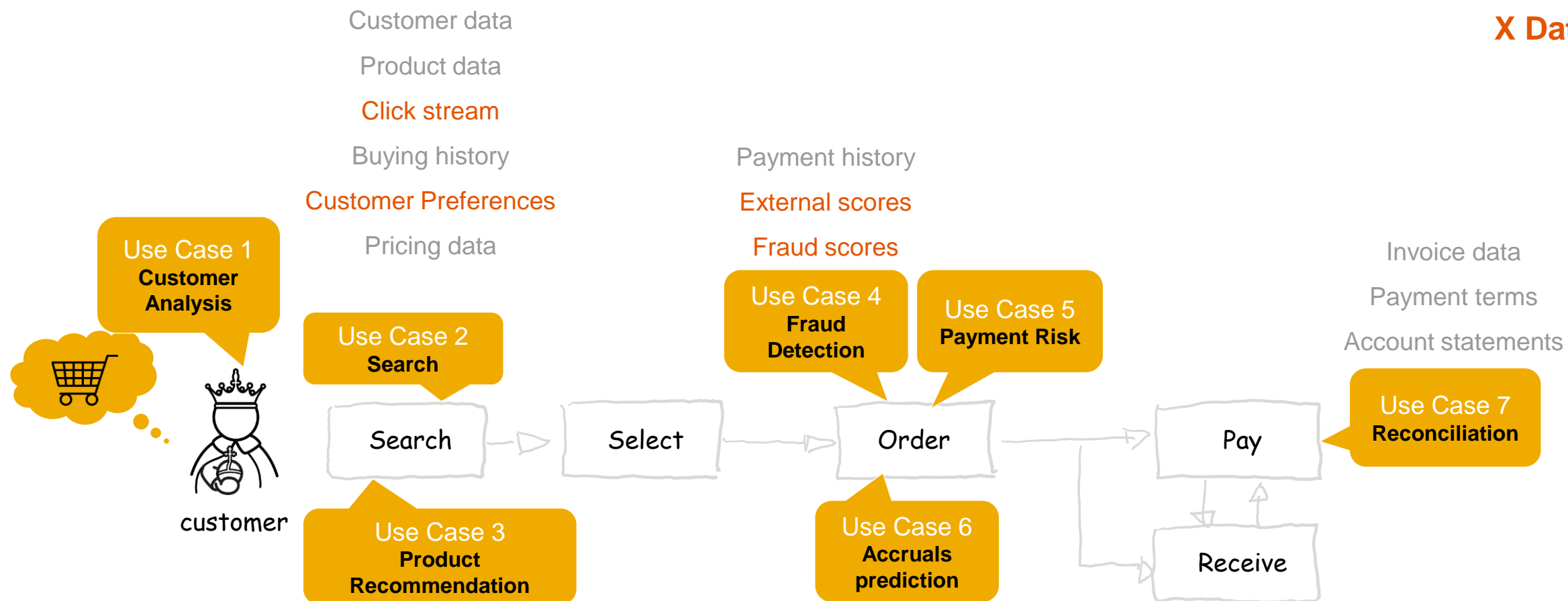
What We See At Our Customers

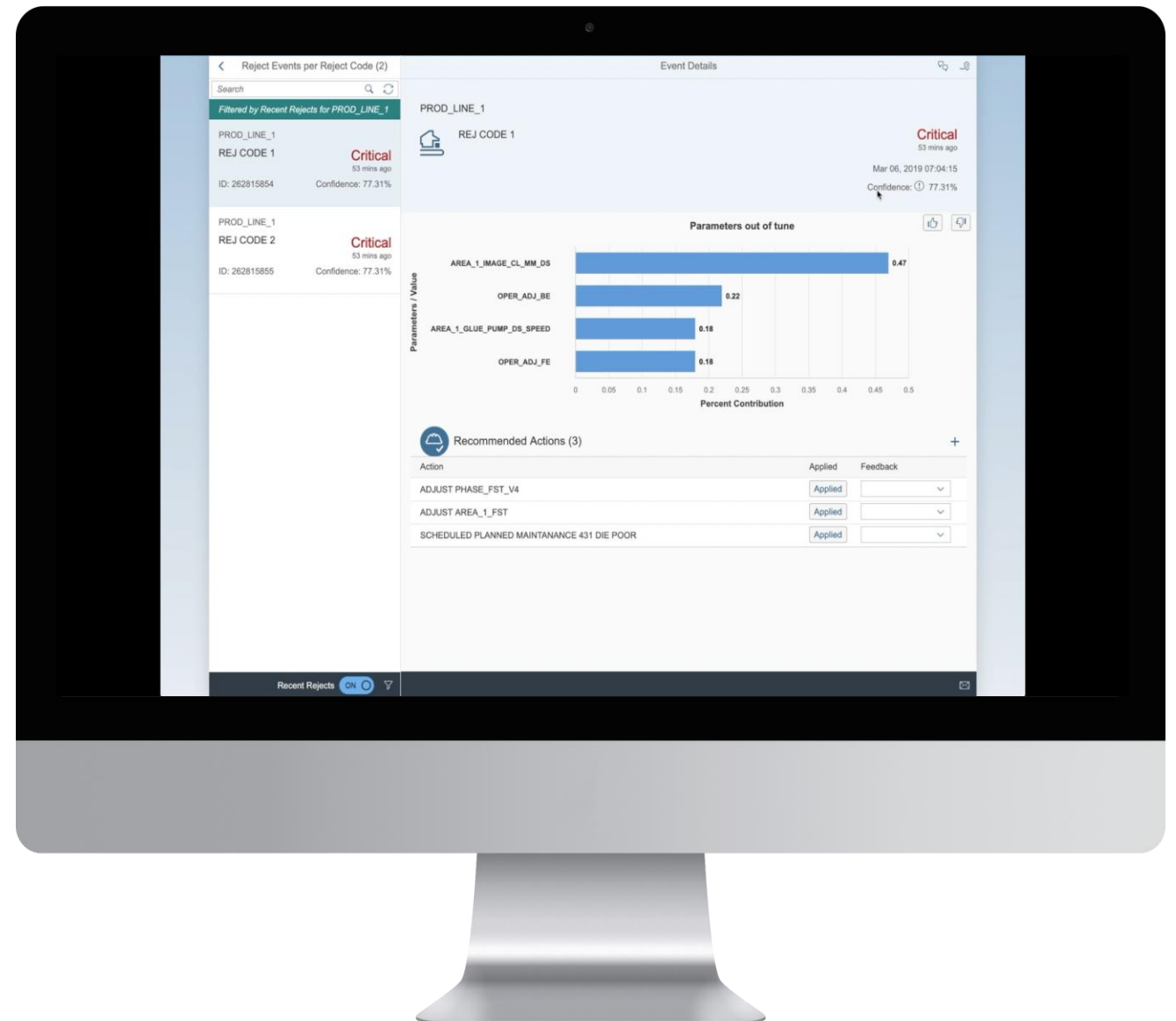
Data Driven Decisions

DATA SCIENCE IS RESHAPING BUSINESS PROCESSES

O Data

X Data







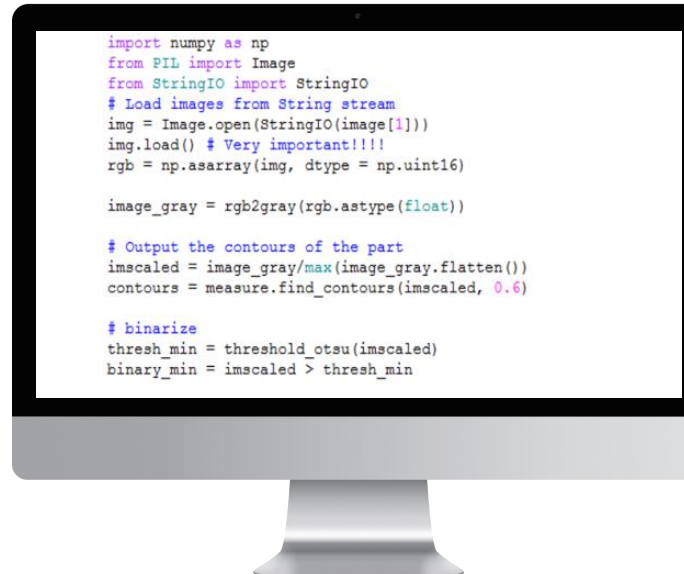
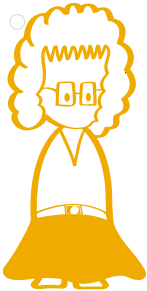
INNOVATION EVENTUALLY MEETS OPERATIONS



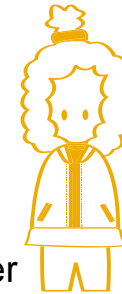
I'm not the one
to productize
this

Here is my
python code!

Data
Scientist



IT
Developer



IT
Operations





INNOVATION EVENTUALLY MEETS OPERATIONS

Head of Digital

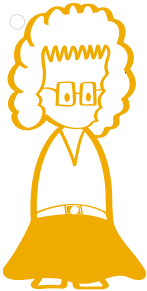


Always the same discussion. It's so slow.
Costs money. Doesn't scale.
And I'm half way fired or in jail if this is not compliant!

I'm not the one to productize this

Here is my python code!

Data Scientist



```
import numpy as np
from PIL import Image
from StringIO import StringIO
# Load images from String stream
img = Image.open(StringIO(image[1]))
img.load() # Very important!!!
rgb = np.asarray(img, dtype = np.uint16)

image_gray = rgb2gray(rgb.astype(float))

# Output the contours of the part
imscaled = image_gray/max(image_gray.flatten())
contours = measure.find_contours(imscaled, 0.6)

# binarize
thresh_min = threshold_otsu(imscaled)
binary_min = imscaled > thresh_min
```

IT Developer



Does our architect know about this?

How to transform/embed this in our application technology?

Did anyone check data access rights?

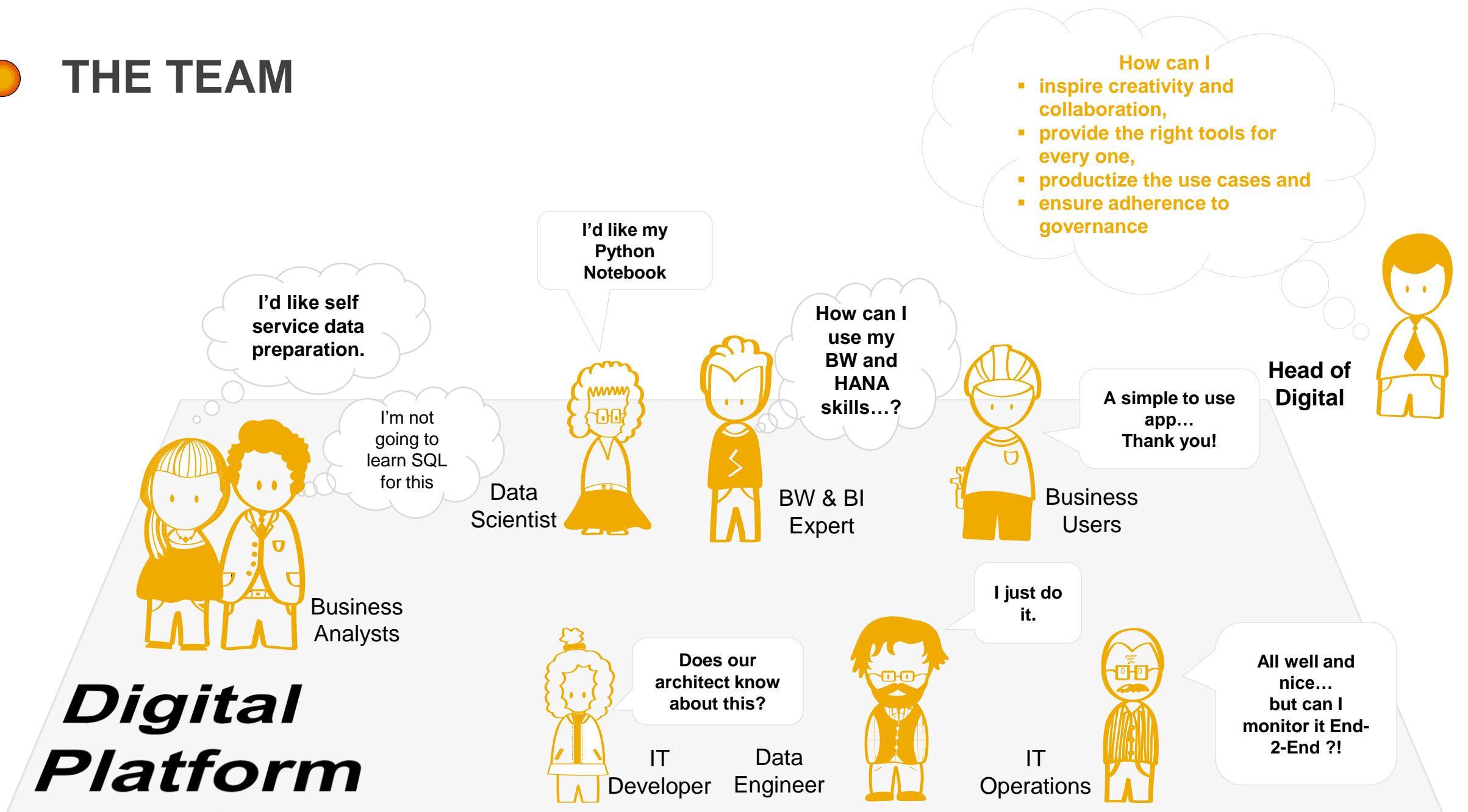
IT Operations



How to automate this?
How to monitor?
How to maintain this?



THE TEAM





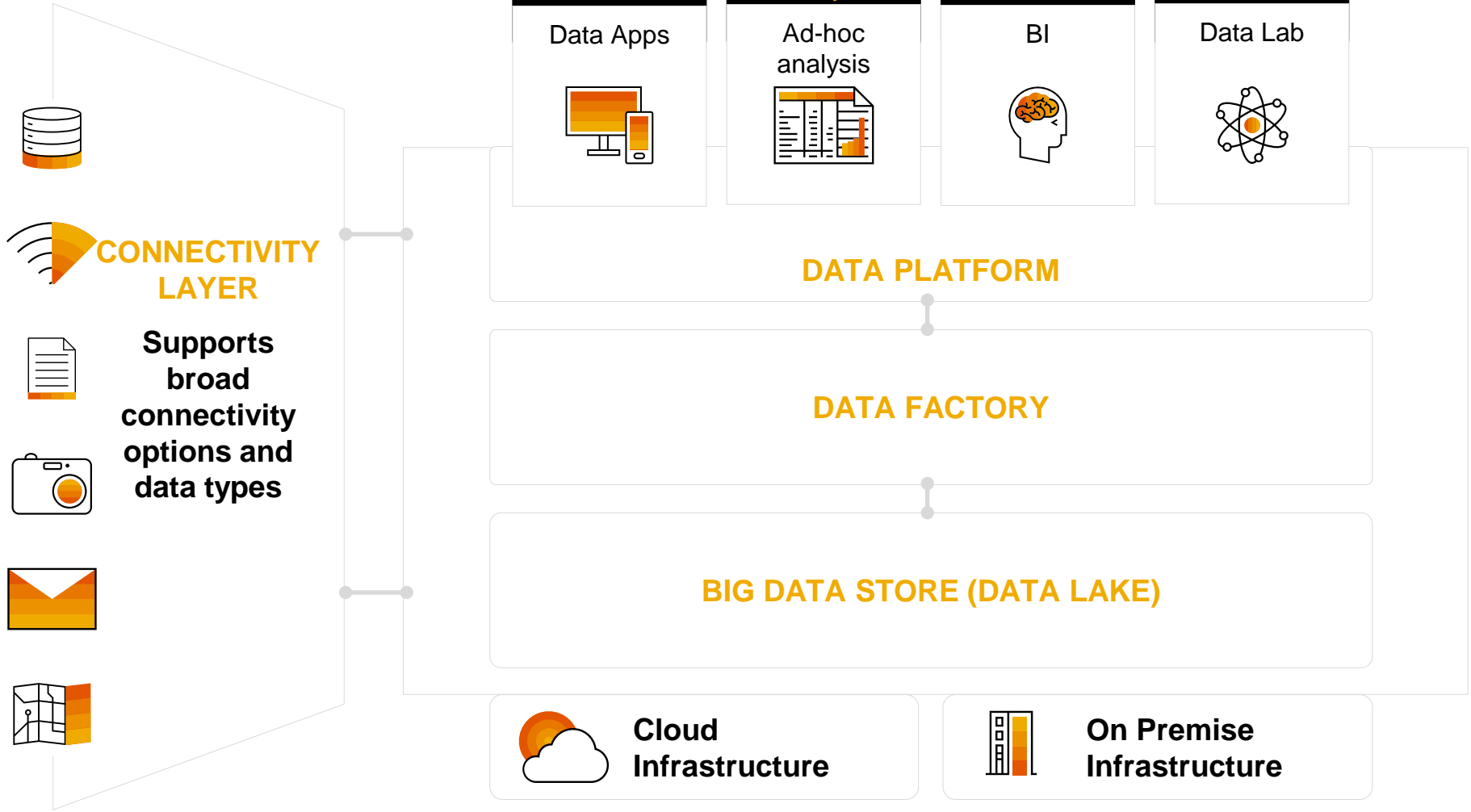
Data Intelligence Platform

How SAP bridges the gap between Innovation and Enterprise Apps

How can I provide all needed functionality in an enterprise ready way...?

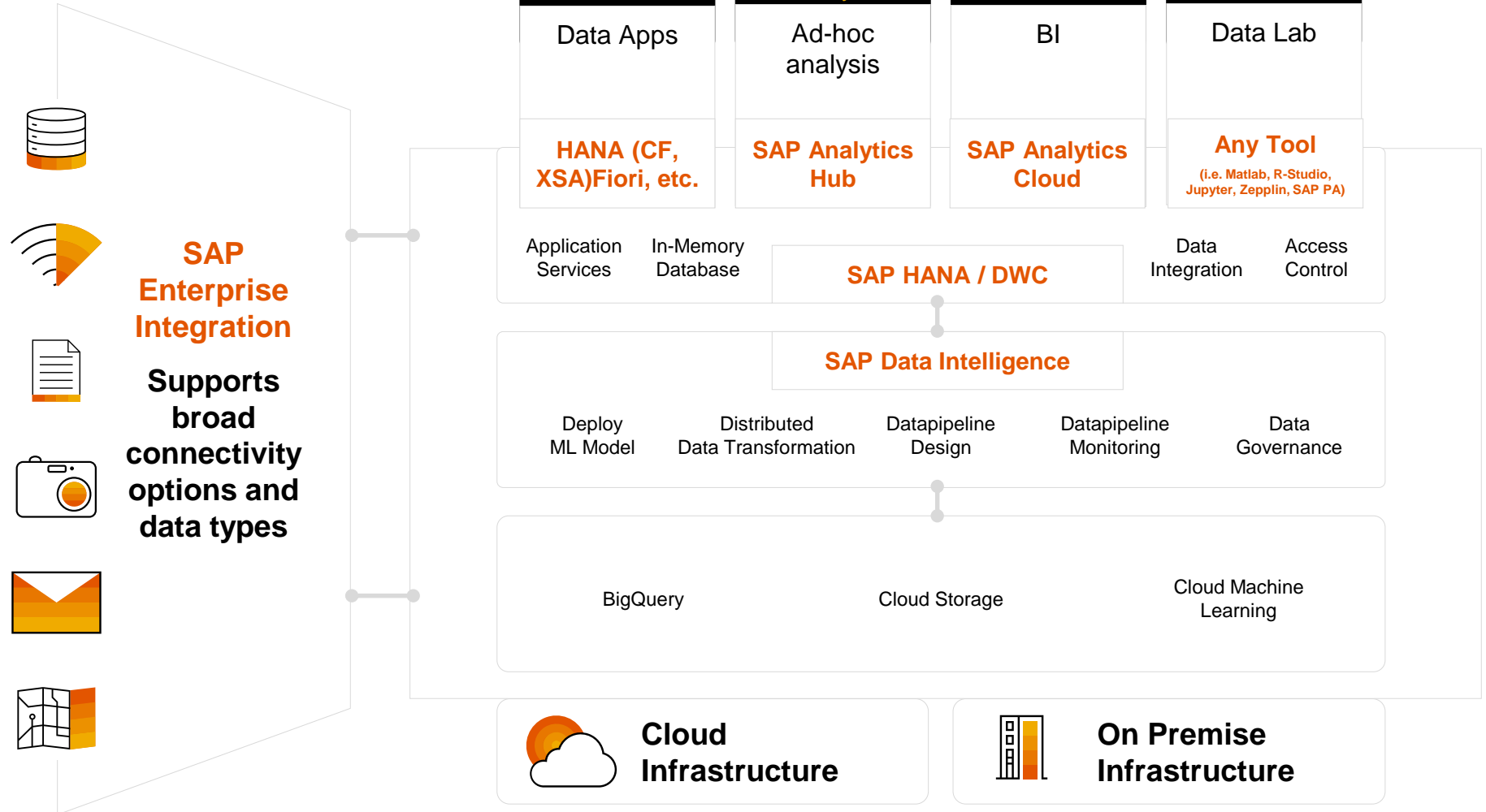


Head of Digital

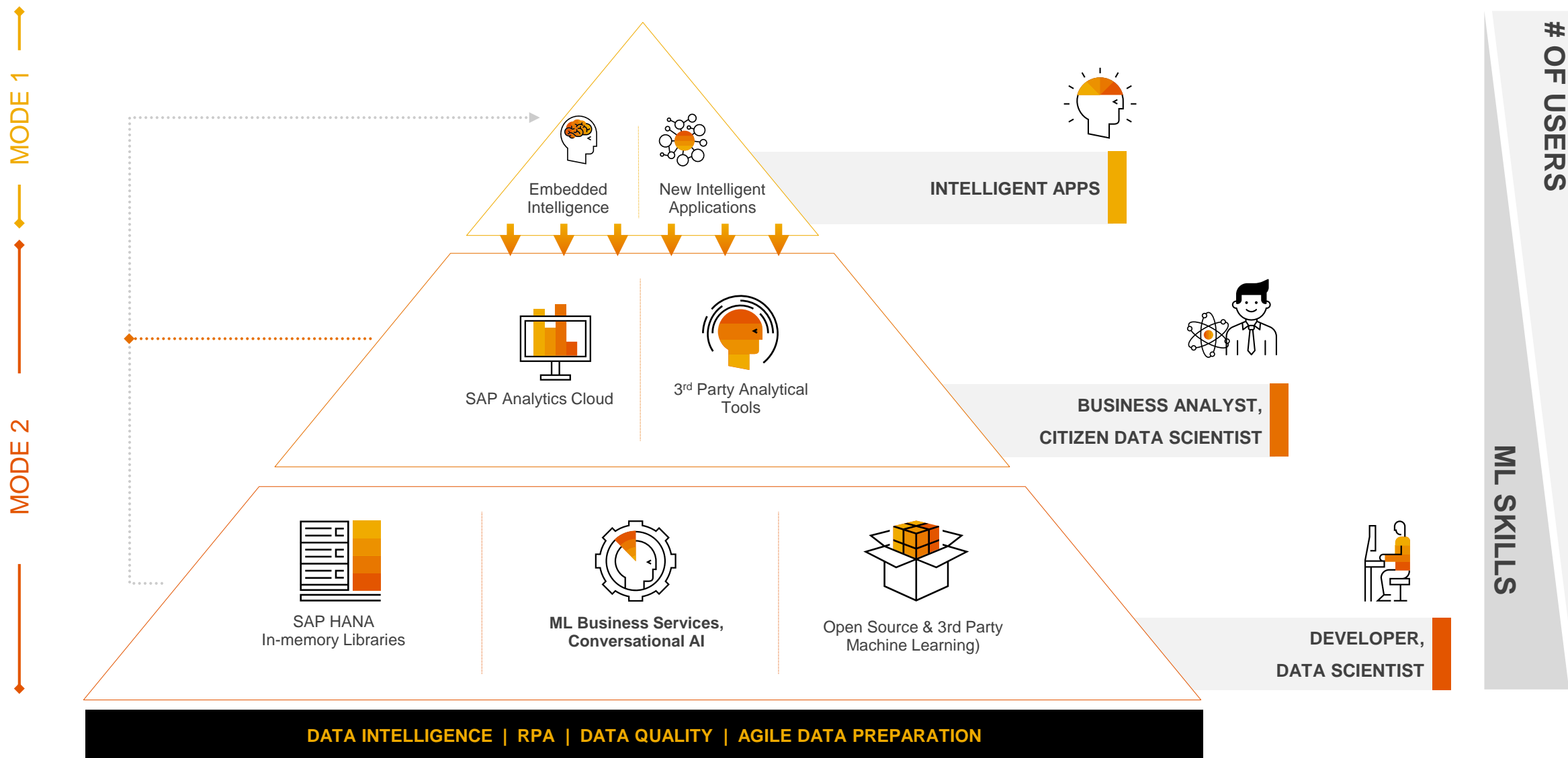




SAP'S PERSPECTIVE

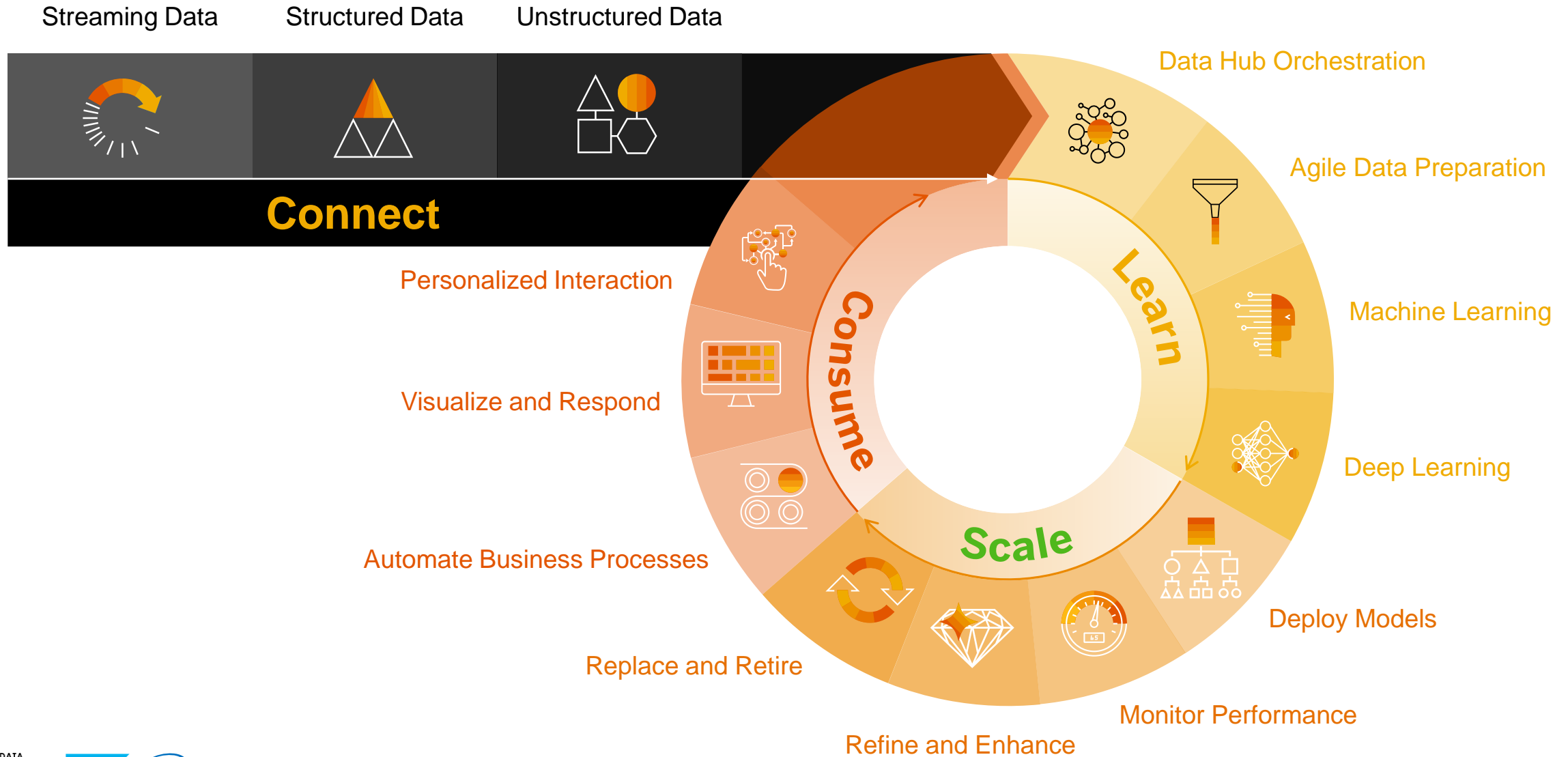


● HOW IS SAP ENABLING THE INTELLIGENT ENTERPRISE





SAP DATA INTELLIGENCE: THE FOUNDATION OF THE AI ASSEMBLY LINE





CREATE AND DEPLOY

The Right Tools for the Job

Work in a JupyterLab or Pipeline context so your Data Scientists can use the tools they know, and IT can understand what is being built

AI Labs as a Service

Spin up any lab environment supporting major SAP and Open Source languages and libraries so no time is wasted waiting for infrastructure to be ready

Deployment Made Simple

Promote Lab environments into production environments with just a few clicks

The screenshot displays a JupyterLab environment. On the left, a file browser shows a directory structure with files like 'Clickstream to Payment Prediction.ipynb' and 'credit_features.csv'. The main area is split into two panes. The top pane shows a code editor with Python imports for pandas, matplotlib, numpy, and sklearn. The bottom pane shows the output of a Jupyter cell, which includes a text description 'Credit features from historic payments (Data Lake)', a code snippet for reading a CSV file, a preview of the first 5 rows of the data (a table with 24 columns), and a scatter plot titled 'Explore Credit Feature Space' showing a distribution of data points.

```
[30]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import sklearn
```

Credit features from historic payments (Data Lake)

```
[5]: df = pd.read_csv("credit_features.csv", header=-1)
df.iloc[:5,]
```

	0	1	2	3	4	5	6	7	8	9	...	14	15	16	17	18	19	20	21	22	23
0	20000	2	2	1	24	2	2	-1	-1	-2	...	0	0	0	0	689	0	0	0	0	1
1	120000	2	2	2	26	-1	2	0	0	0	...	3272	3455	3261	0	1000	1000	1000	0	2000	1
2	90000	2	2	2	34	0	0	0	0	0	...	14331	14948	15549	1518	1500	1000	1000	1000	5000	0
3	50000	2	2	1	37	0	0	0	0	0	...	28314	28959	29547	2000	2019	1200	1100	1069	1000	0
4	50000	1	2	1	57	-1	0	-1	0	0	...	20940	19146	19131	2000	36681	10000	9000	689	679	0

5 rows × 24 columns

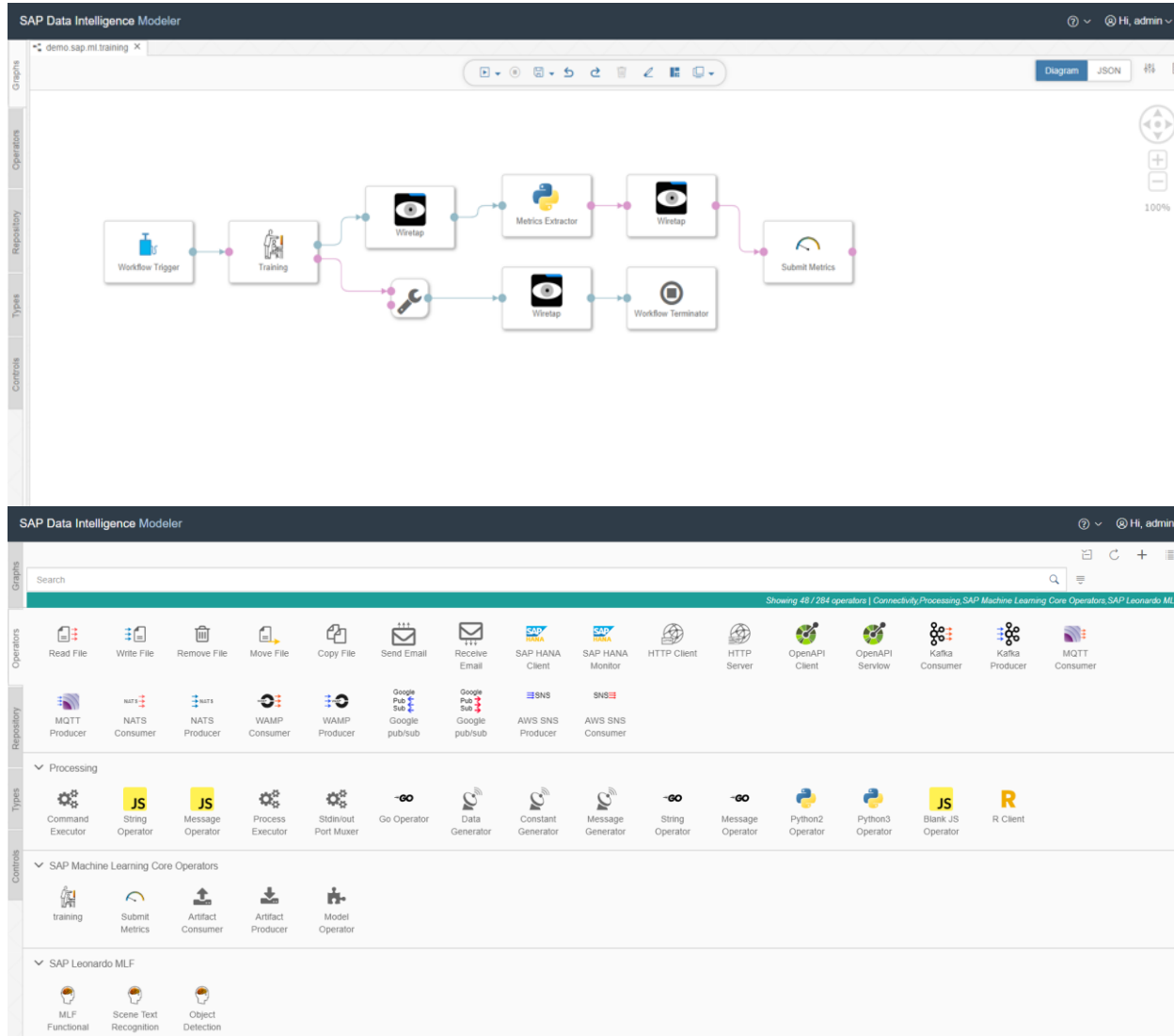
Explore Credit Feature Space

```
[16]: w = df.iloc[:10000, 15] / np.max(df.iloc[:10000, 15])
w[w == 0] = 0.1
w = w * 60
plt.scatter(df.iloc[:10000, 14], df.iloc[:10000, 17], c=df.iloc[:10000, 23], s=w)
```

[16]: <matplotlib.collections.PathCollection at 0x7f5ff53c4358>



UNDERSTAND AND EXPLAIN



Understand Model Performance

Track and trace all elements of the machine learning process to understand model performance and ensure trustworthy results

Explain data impact

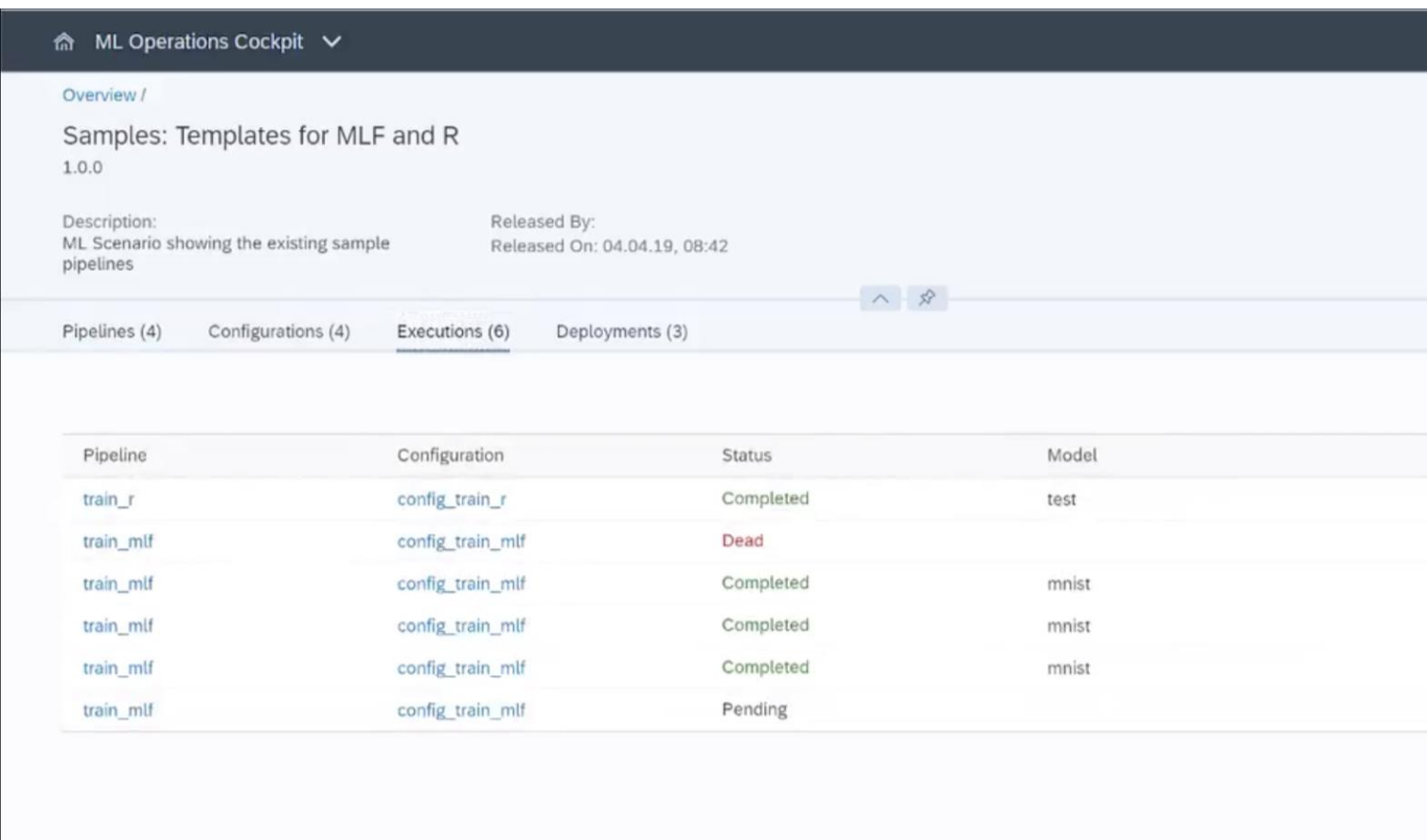
Profile and prepare data prior to model creation

Track data lineage to understand data impact

Reproduce Results

Review metadata and version history to ensure reproducibility and explicable results

AUTOMATE AND SCALE



ML Operations Cockpit

Overview /

Samples: Templates for MLF and R
1.0.0

Description: ML Scenario showing the existing sample pipelines
Released By: Released On: 04.04.19, 08:42

Pipelines (4) Configurations (4) Executions (6) Deployments (3)

Pipeline	Configuration	Status	Model
train_r	config_train_r	Completed	test
train_mlf	config_train_mlf	Dead	
train_mlf	config_train_mlf	Completed	mnist
train_mlf	config_train_mlf	Completed	mnist
train_mlf	config_train_mlf	Completed	mnist
train_mlf	config_train_mlf	Pending	

Automate Low Value Tasks

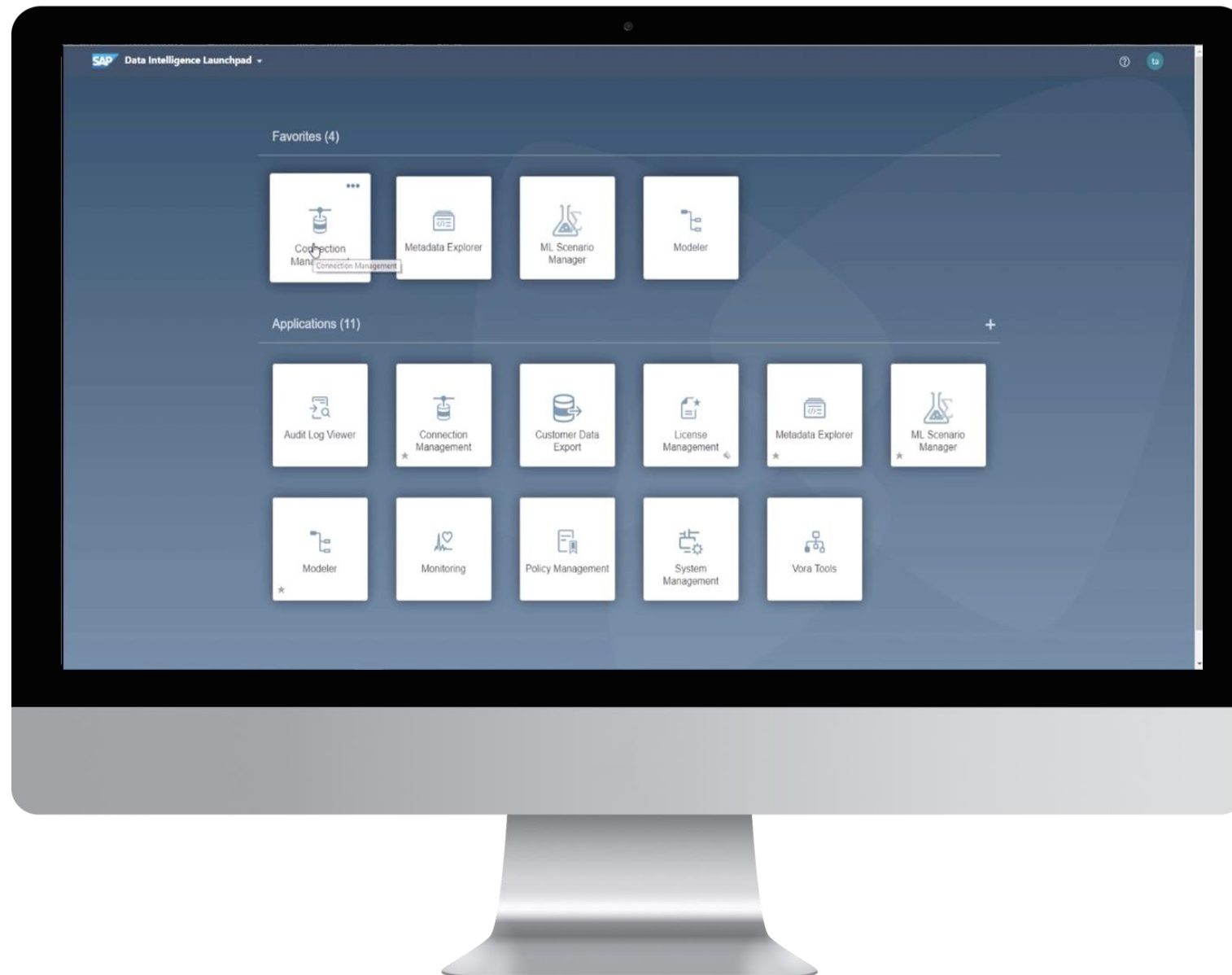
Model performance and lifecycle is automatically managed so IT and Data Scientists can focus on the tasks that matter

Manage Everything at Once

One single interface to see all AI models across the organization in a unified workspace with tools to help manage governance, auditability, and transparency

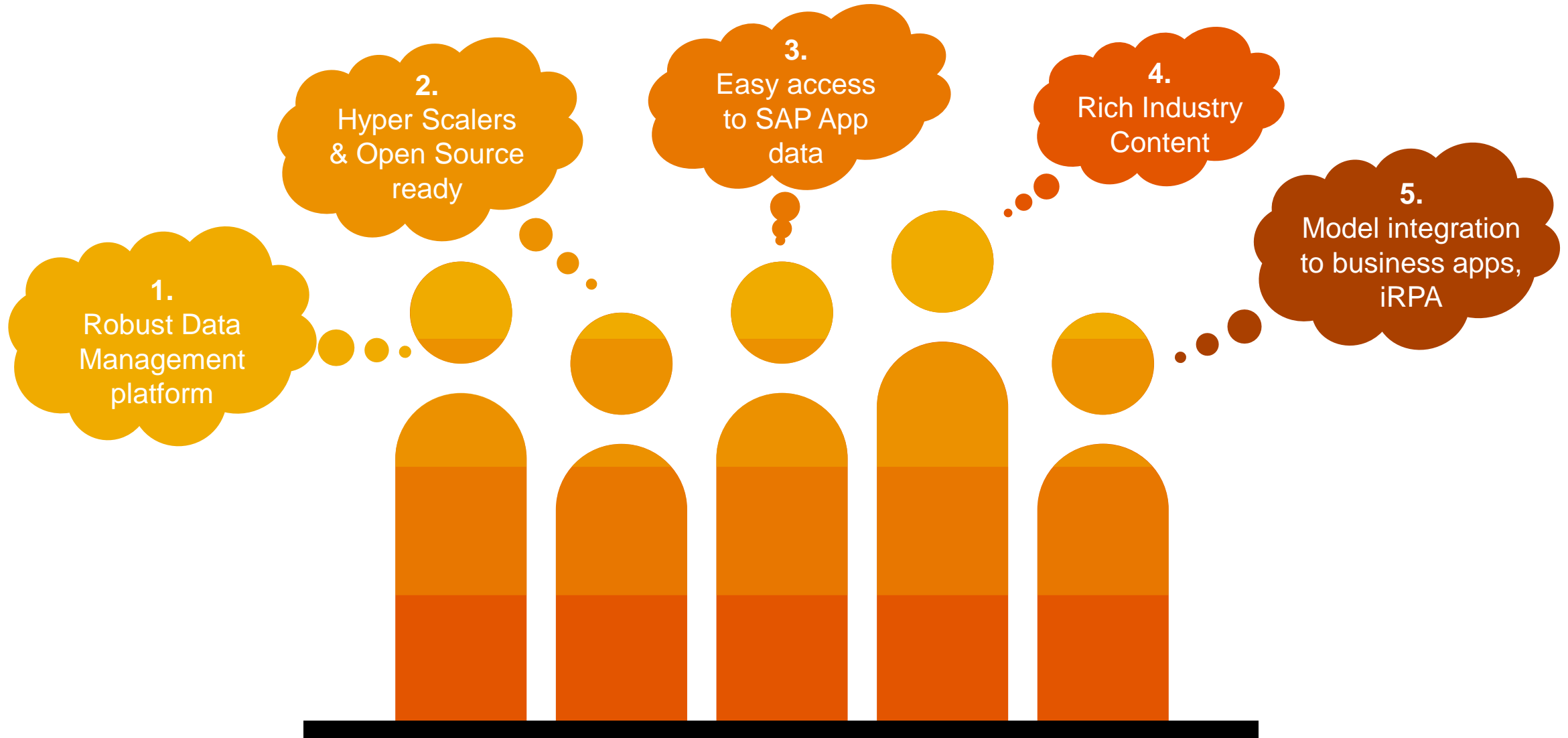
Reduce Infrastructure Costs

Unique serverless cloud architecture means you only use and pay for what you need. No More, No Less.





LOOKING AHEAD: SAP'S PRIORITIES IN DATA SCIENCE



A circular inset image on the left side of the slide. It shows a woman and a man smiling and looking at each other. A large white number "4" is overlaid on the left side of the circle.

How Can SAP Help My Company Start

Data Science Experience Workshop



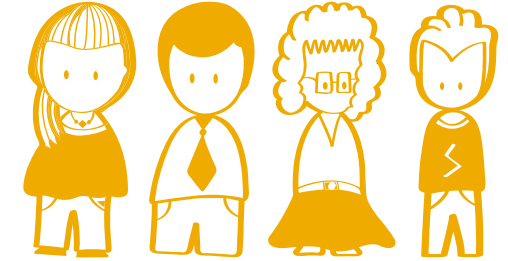
DATA SCIENCE EXPERIENCE WORKSHOP



**SAP presents DS capabilities
with live demos (Day 1)**



Team will transfer into ideas (Day 2)



- BI Manager
- Business Analysts
- Data Scientists
- Big Data Engineers
- Digital/Innovation Officers
- IT Architects
- Business Experts

Participants



KEY TAKEAWAYS



Despite the hype, businesses are still lagging behind in AI mass adoption



Business users want simple and trusted solutions with clear ROI; not bleeding edge technology



SAP is investing heavily in demystifying and automating the E2E process: from data, to insights to business applications

“

All the impressive achievements of deep learning amount to just curve fitting.

Judea Pearl, AI pioneer and philosopher