

Université IBM i 2018

May 16/17 2018

IBM Client Center Paris



IBM i & Data Science: Introduction to Watson Studio & IBM Datascience Experience (DSX Local)

Benoit MAROLLEAU - Cloud Architect
IBM Client Center Montpellier, France



@ : Benoit.marolleau@fr.ibm.com

 [linkedin.com/in/benoitmarolleau](https://www.linkedin.com/in/benoitmarolleau)

 [MarolleauBenoit](https://twitter.com/MarolleauBenoit)

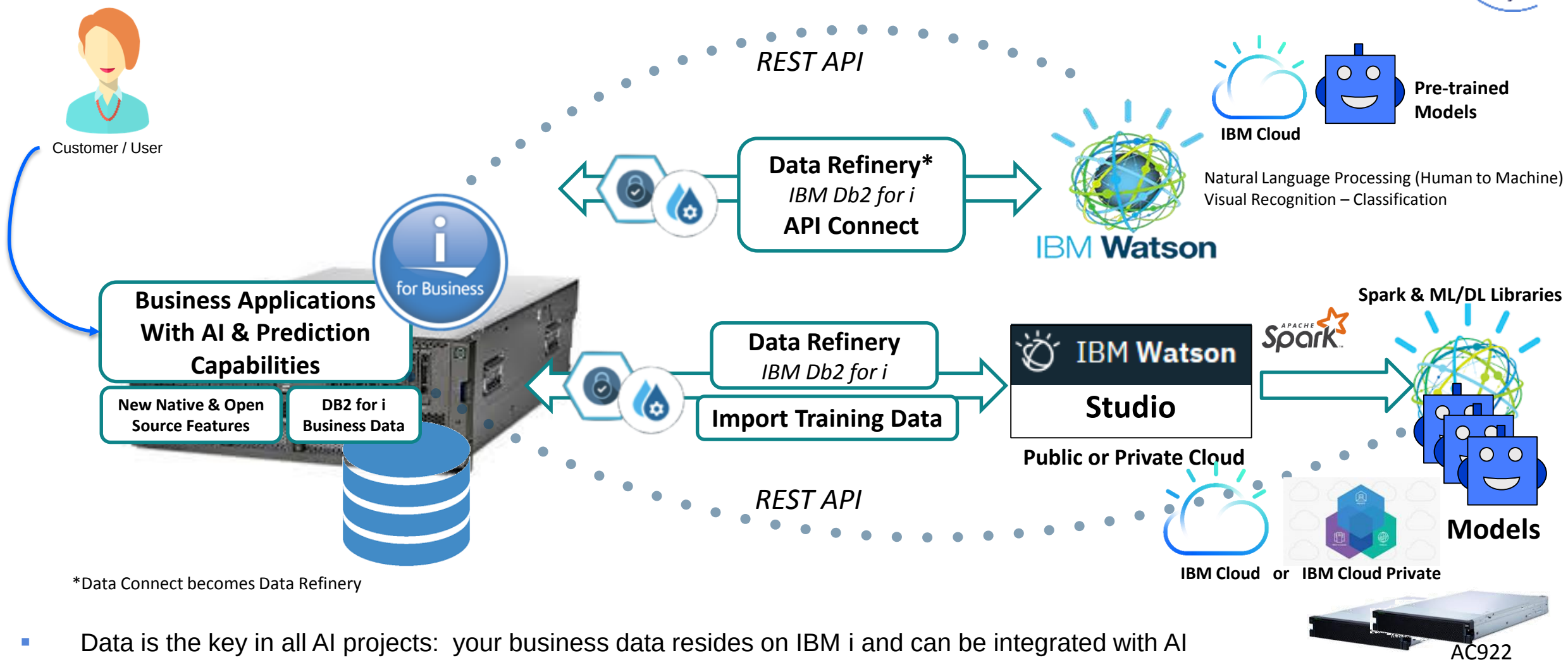
Agenda



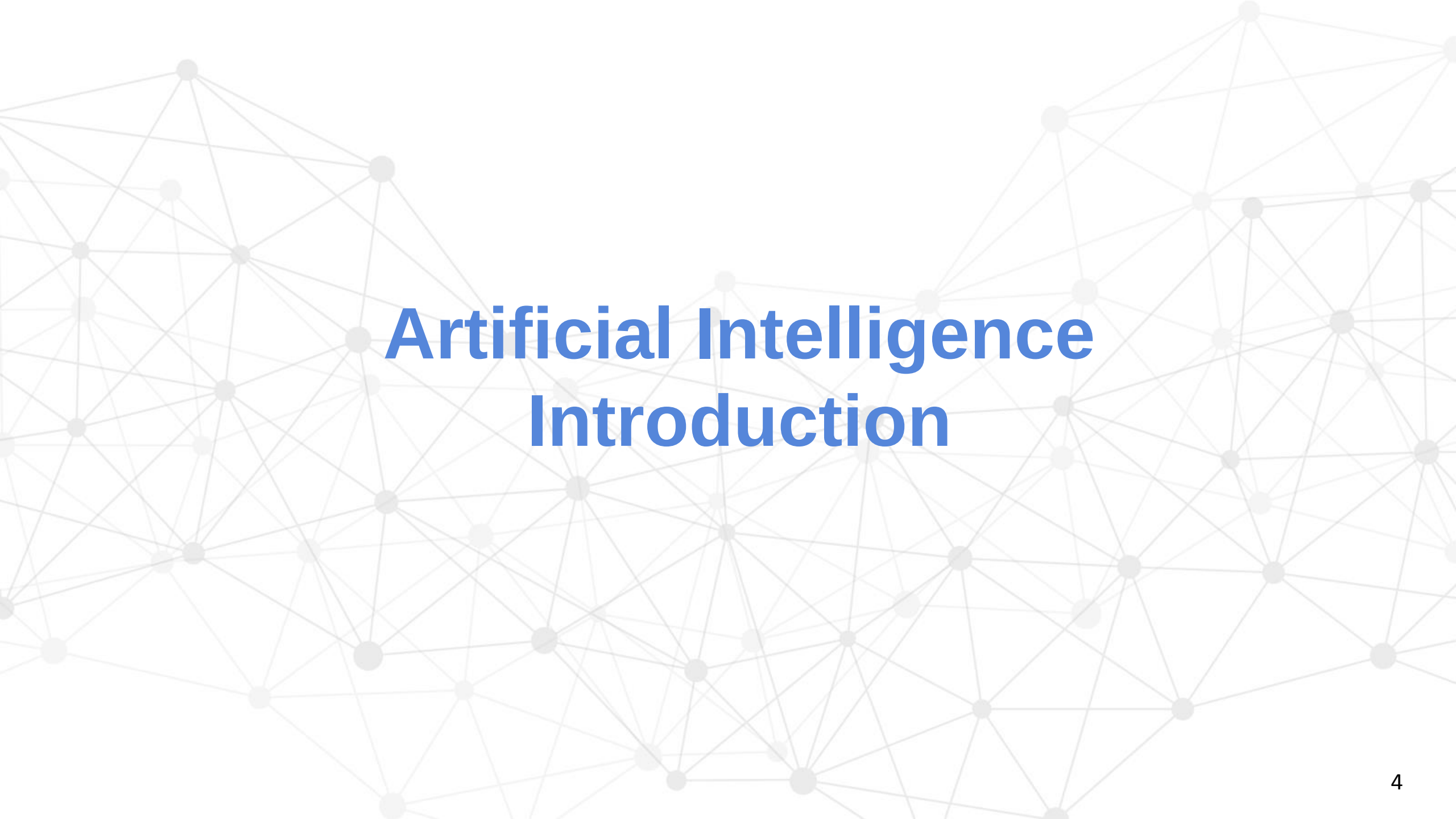
- Introduction: Why AI? What benefits? Approximate vs precise Computing
- AI Enterprise Solutions: IBM Watson, Watson Studio, Data Science Experience
- AI & IBM i
 - Integration example in a data science project
- How to get started?
 - Q&A
 - Contact us! IBM Cognitive Systems Lab

IBM i & Artificial Intelligence

Approximate (AI) & precise (Transactional) computing together



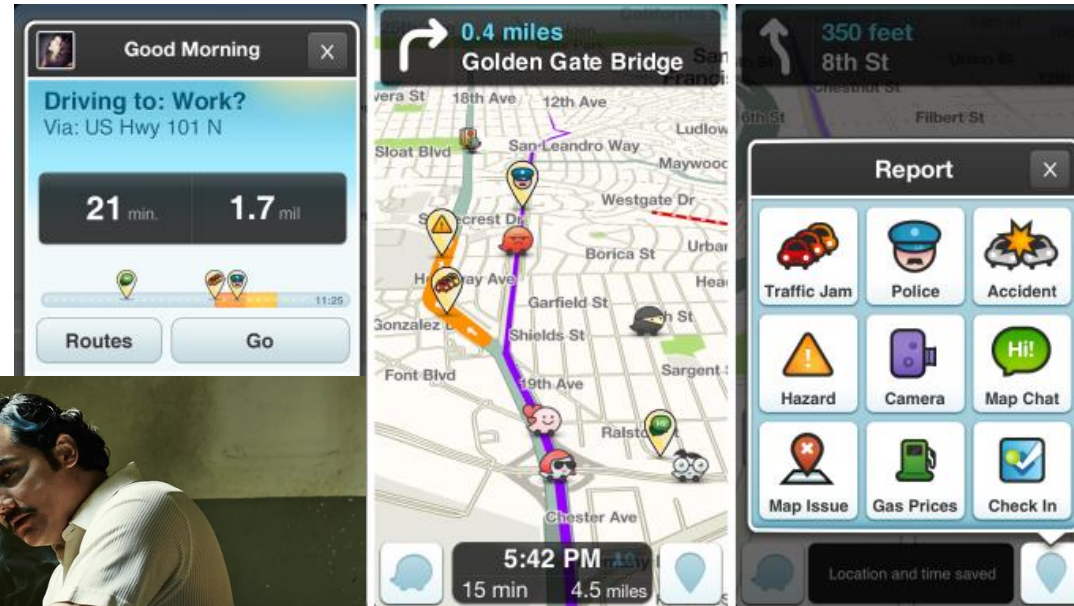
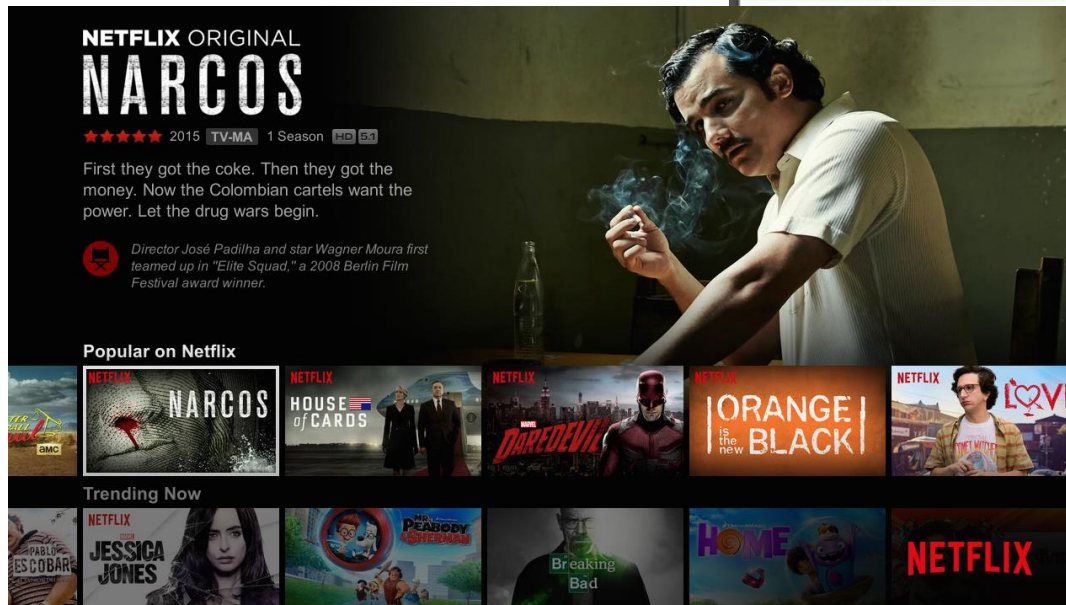
- Data is the key in all AI projects: your business data resides on IBM i and can be integrated with AI
- Use pre-trained & customizable models with IBM Watson (Developer Cloud) services in IBM Cloud
- Build your own use case & business specific models with IBM Watson Studio - IBM Cloud / on premises (DSX Local w/ Cloud Private)



Artificial Intelligence Introduction

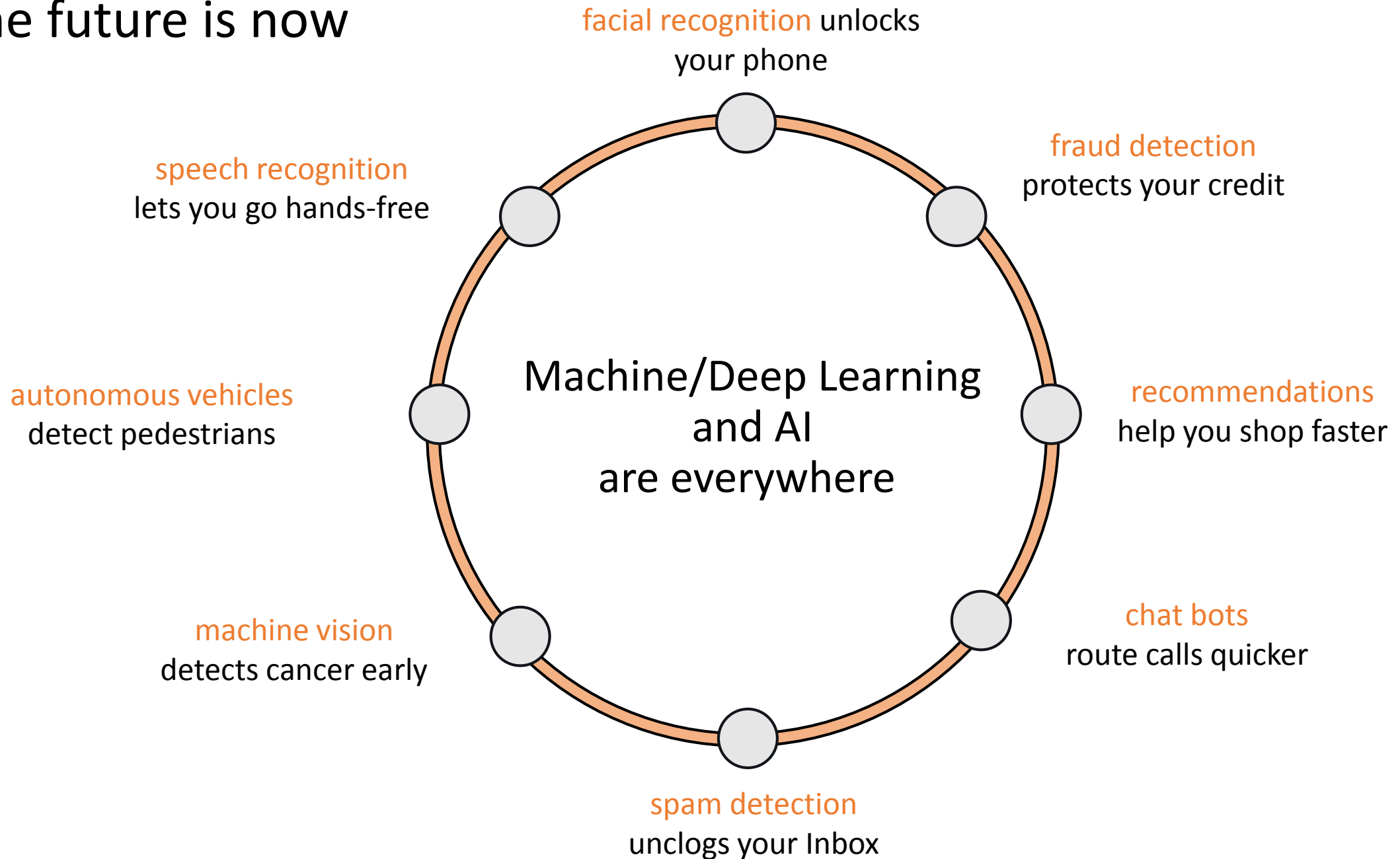
Machine learning is everywhere – influencing nearly everything we do

Netflix provides personalized movie recommendations



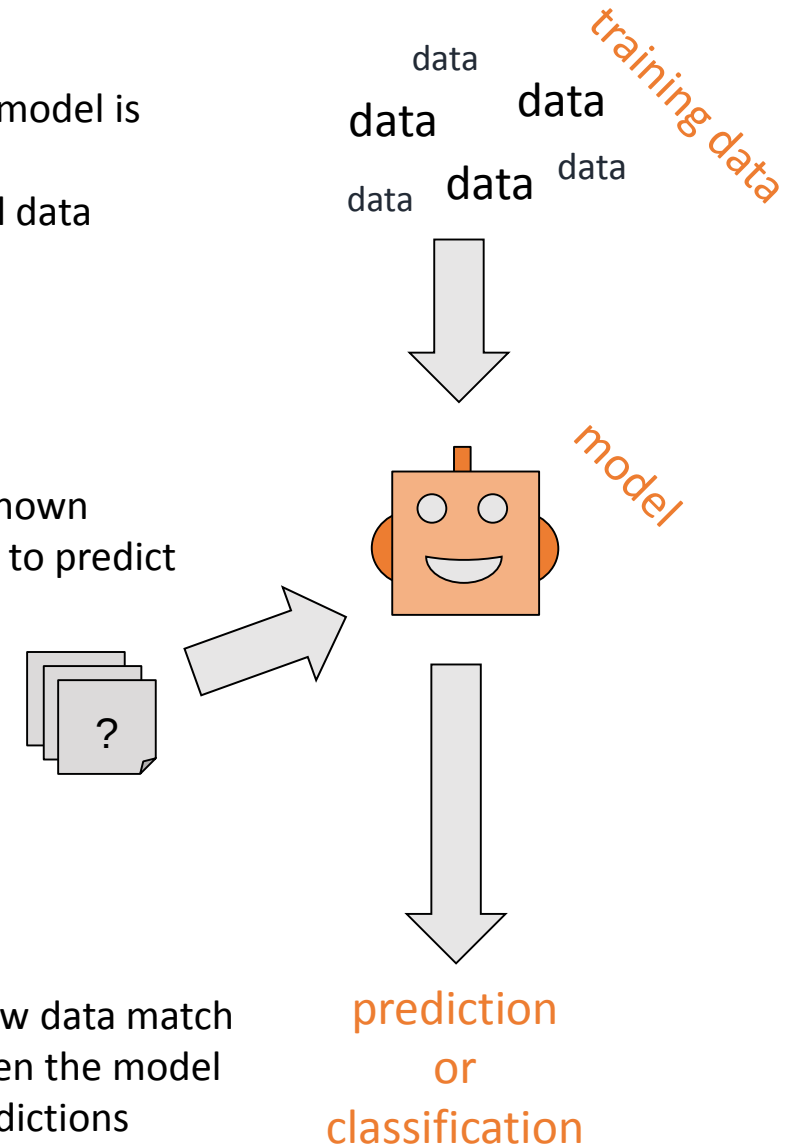
Waze provides a personalized driving experience for its users

The future is now

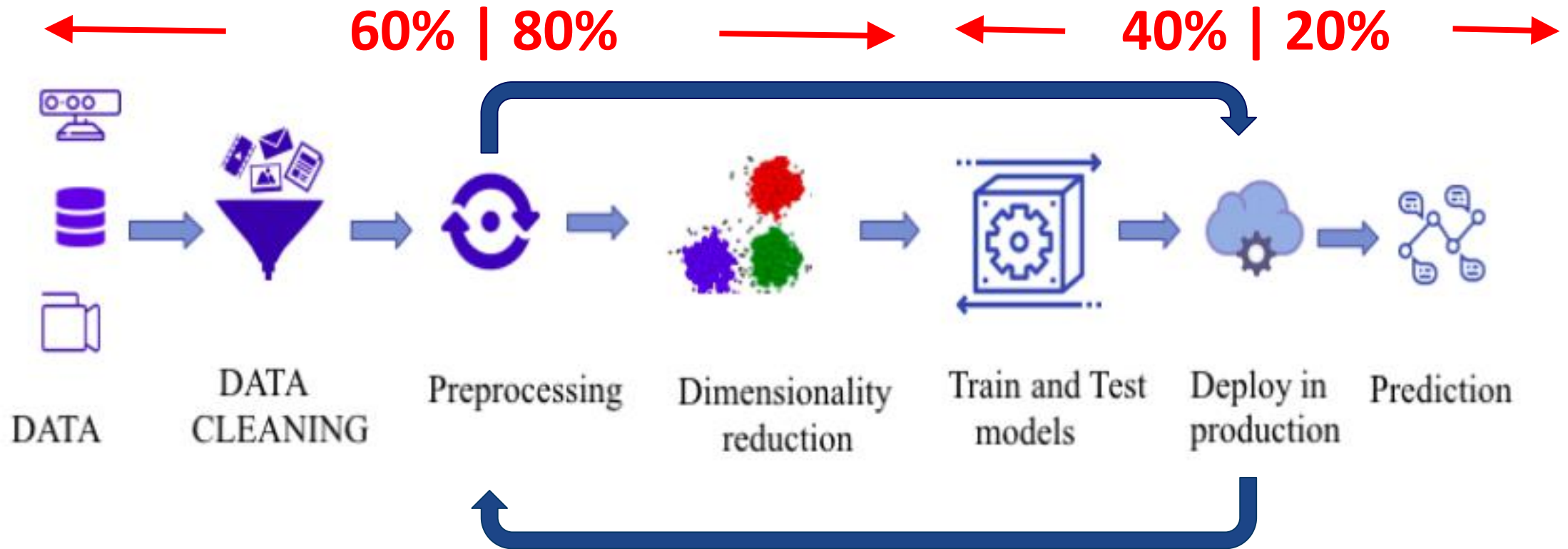




- ③ If patterns in the new data match the training data then the model makes accurate predictions



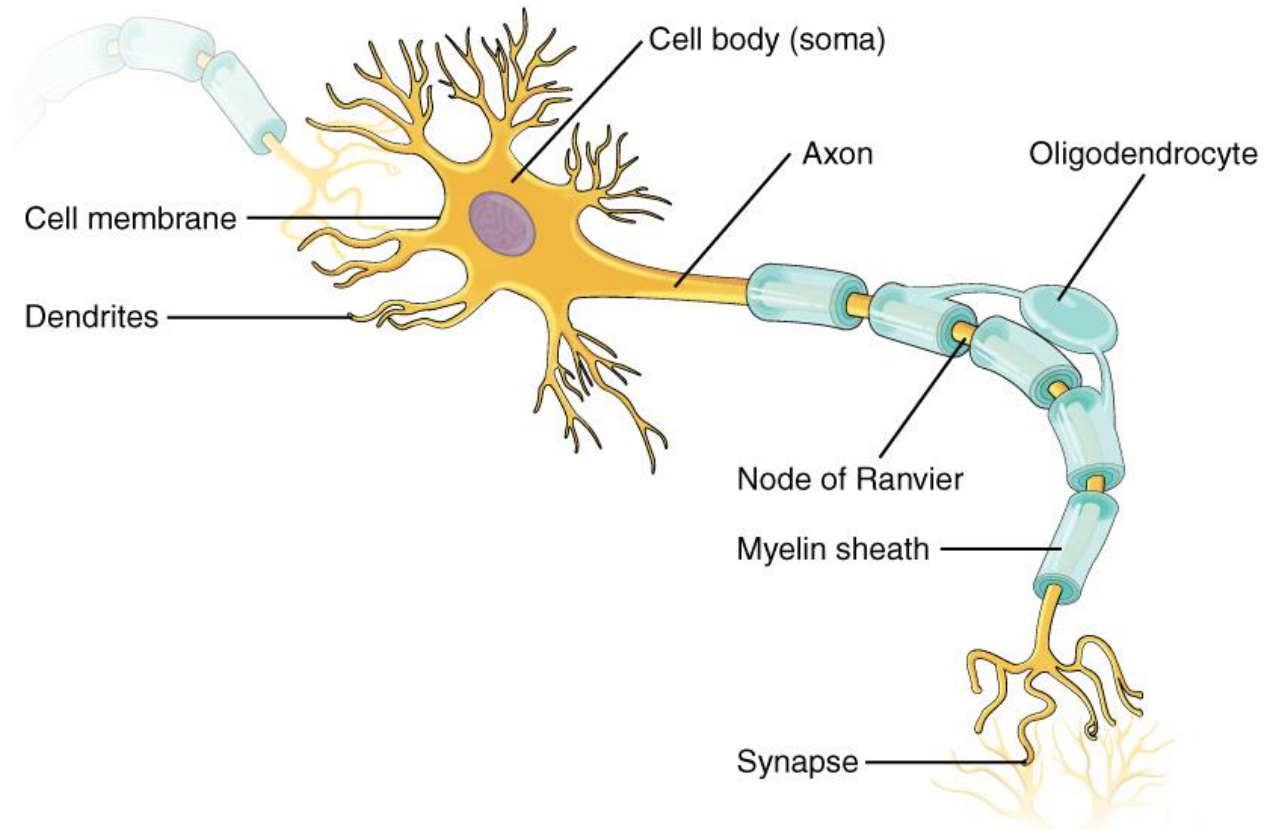
From raw data to AI & Cognitive



A full pipeline to leverage machine learning techniques to solve daily issues

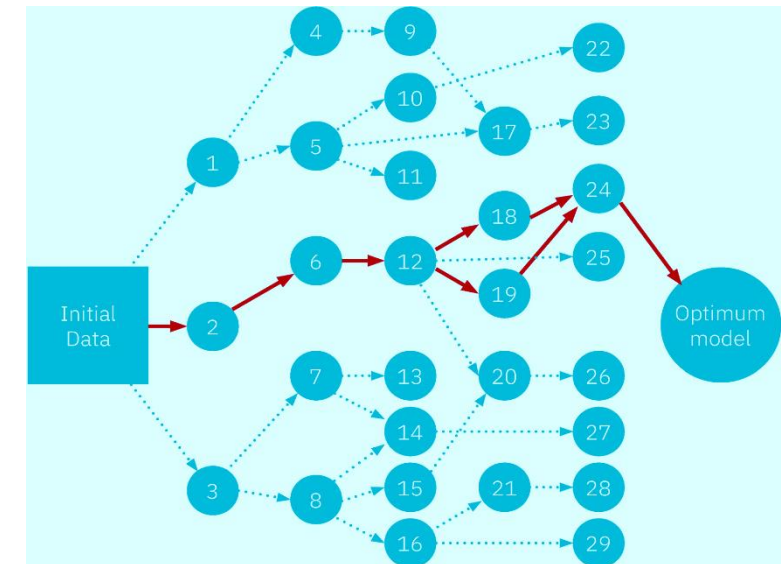
Deep Learning = Training Artificial Neural Networks

Based on biological neurons. Artificial neurons learn by recognizing patterns in data.



A human brain has:

- 200 billion neurons
 - 32 trillion connections between them
- Artificial neural networks have far fewer



Impact of Machine Learning: A simple example



Direct marketing — 1% response rate		
Send marketing mail to 1,000,000 customers at cost of \$2 per mailing to sell a \$220 service.	\$2 x 1,000,000	\$2,000,000
One percent response rate means 10,000 customer will buy service.	\$220 x 10,000	\$2,200,000
Profit*		\$200,000
Predictive direct marketing — 3% response rate		
Send marketing mail to 250,000 customers <i>predicted most likely to buy</i> at cost of \$2 per mailing to sell a \$220 service.	\$2 x 250,000	\$500,000
Three percent response rate means 7,500 customer will buy service.	\$220 x 7,500	\$1,650,000
Profit when using a ML model*		\$1,150,000

Traditional

**Machine
learning**

*Profit calculation does not include other expenses.

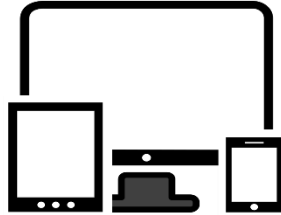
Every industry is changing & can benefit

Leaders everywhere are monetizing data & developing strategies to embed AI in business



Retail

Market Basket Analysis, Next Best Offer, Customer Churn, propensity to buy



Marketing

Discount targeting, email optimization, lifetime client value



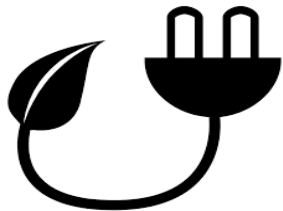
Healthcare

Medicare fraud, AI-assisted diagnosis, drug demand forecast



Manufacturing

Predictive maintenance, process optimization, demand forecast



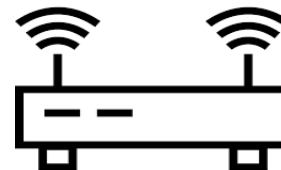
Energy and Utilities

Power usage prediction, smart grid management



Banking

Customer segmentation, credit risk, credit card fraud detection



Security

Malicious activity detection, logs analysis

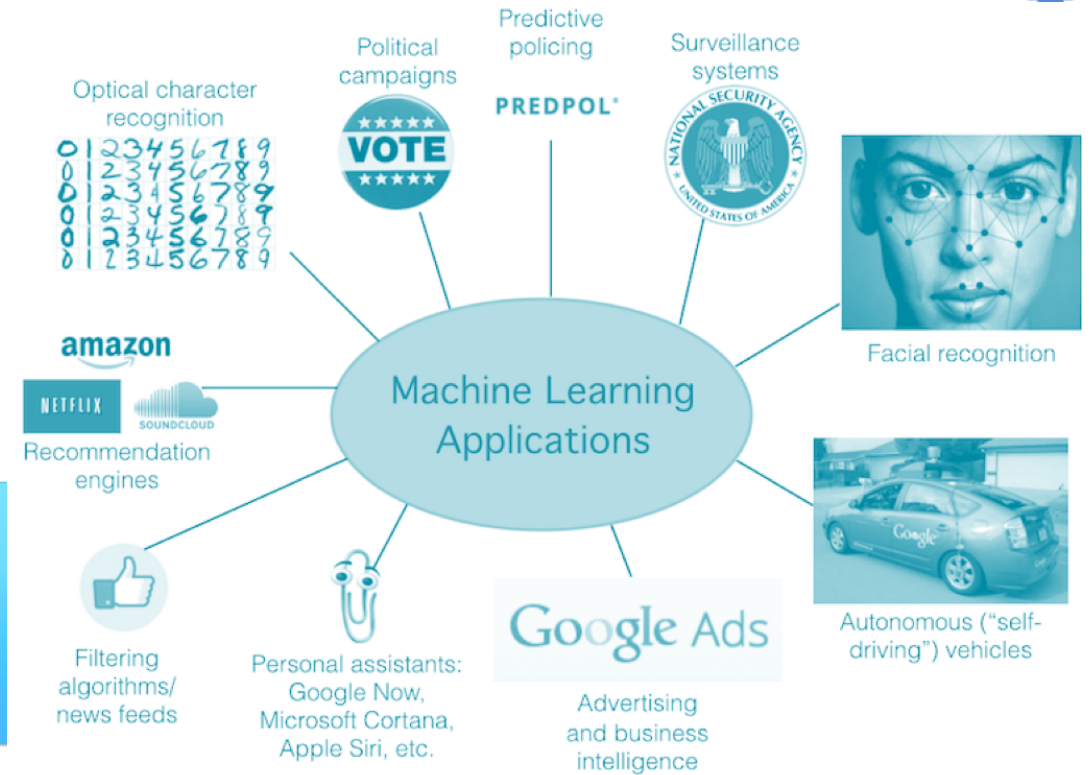
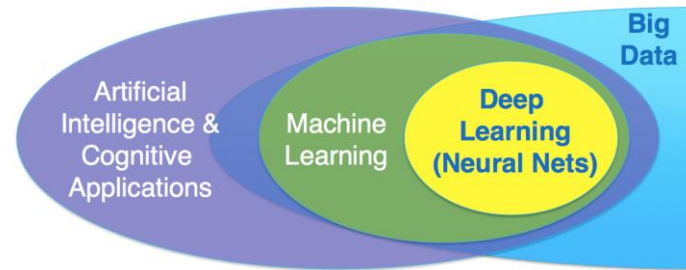


Travel & transportation

Dynamic pricing, call center assistants, tourism forecasting, Self-driving cars

Big Data: Machine Learning techniques

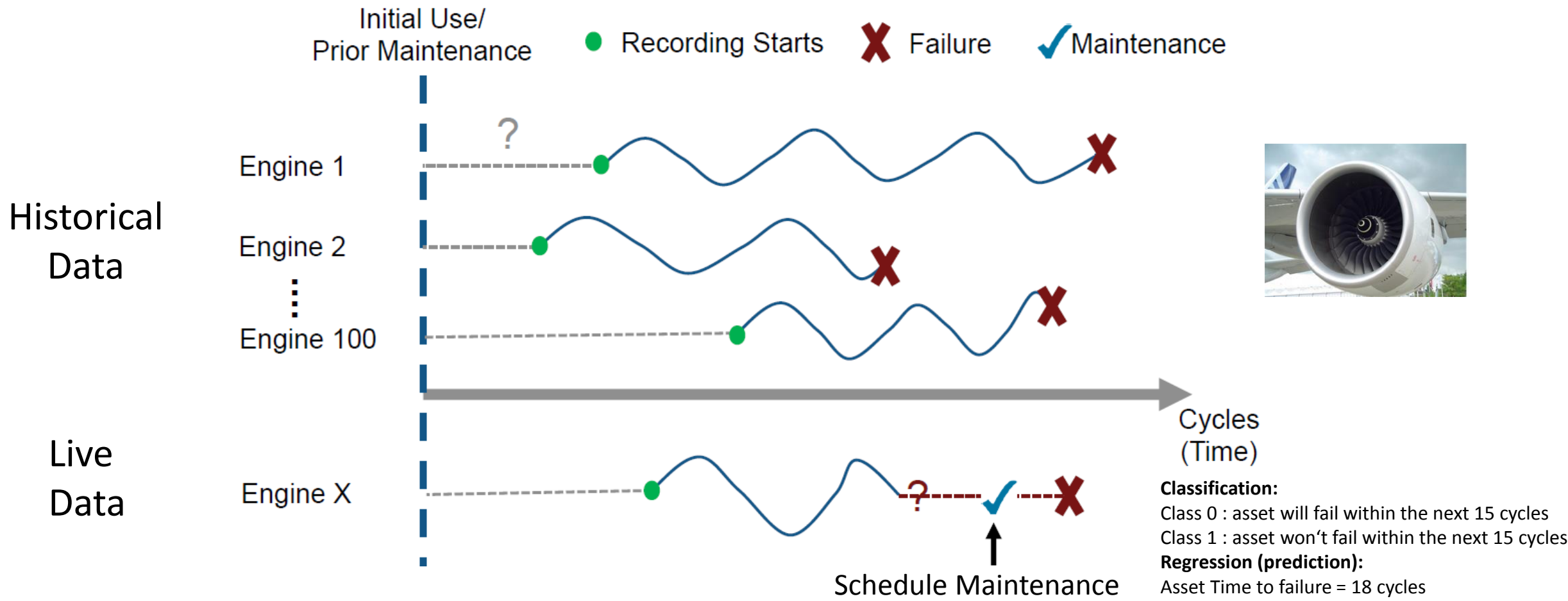
- Classification: predict class from observations
 - - E.g. Spam Email Detection
- Clustering: group observations into “meaningful” groups
 - - E.g. Amazon Recommendations
- Regression (prediction): predict value from observations
 - - E.g. Energy consumption



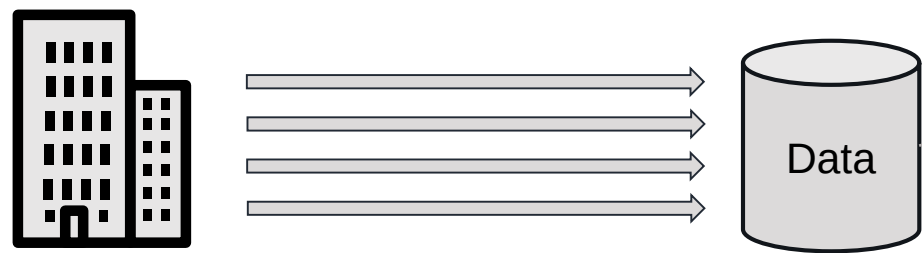
and many different technologies and libraries are available:



Example: Predictive Maintenance



Enterprises generate TONS OF DATA



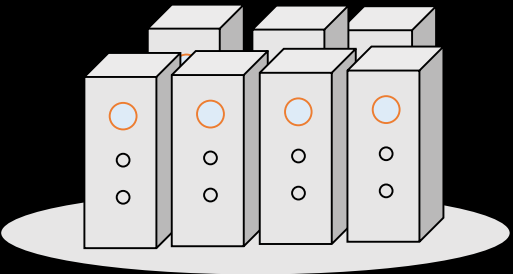
Data that requires governance

Which must be cleaned and shaped for training

...then models must be designed

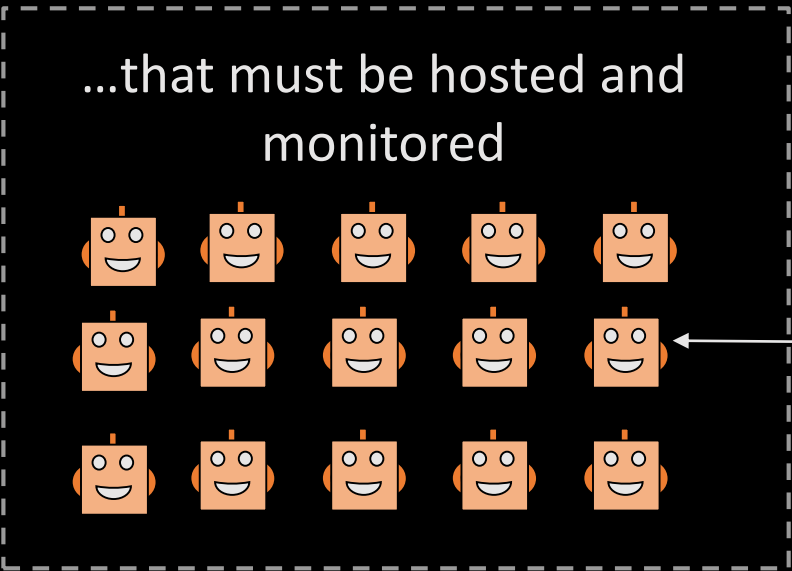


...and trained on high performance compute



To select an optimal model...

IBM Cloud / Watson and Cloud Platform / © 2018 IBM Corporation



Why are enterprises struggling to capture the value of AI?

- | | | | |
|---|---|--|--|
| <ul style="list-style-type: none">• Data• Data resides in silos & difficult to access• Unstructured and external data wasn't considered | <ul style="list-style-type: none">• Governance• If the data isn't secure, self-service isn't a reality• Challenge understanding data lineage and getting to a system of truth | <ul style="list-style-type: none">• Skills• Data Science skills are in low supply and high demand• Nurturing new data professionals is challenging | <ul style="list-style-type: none">• Tools & Infrastructure• Need an environment that enables a "fail fast" approach• Discrete tools present barriers to productivity |
|---|---|--|--|



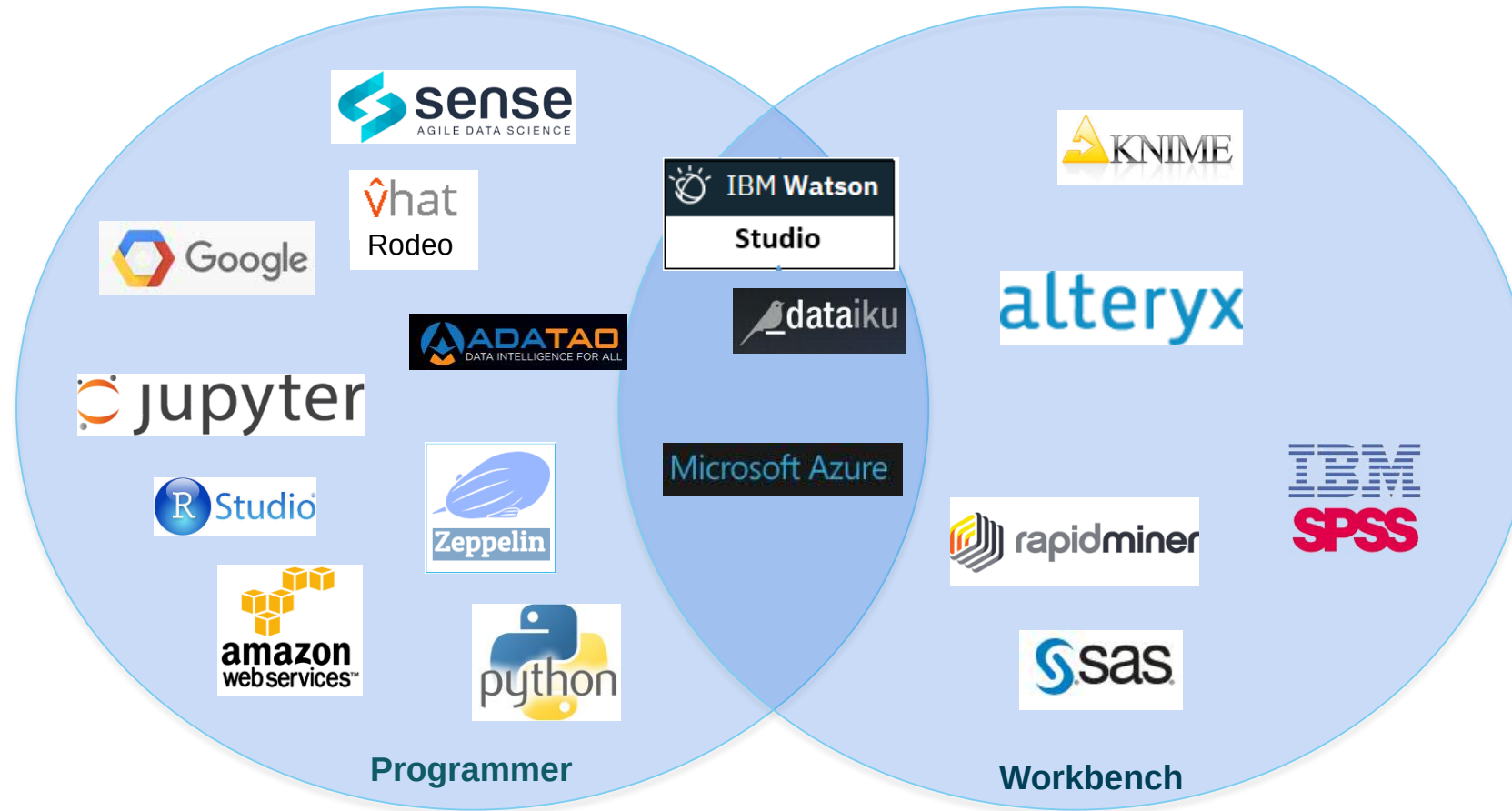
Artificial Intelligence: IBM Solutions

Watson Studio: accelerating value from AI for enterprises

Watson Studio accelerates the machine and deep learning workflows required to infuse AI into your business to drive innovation. It provides a suite of tools for data scientists, application developers and subject matter experts to collaboratively and easily work with data and use that data to build, train and deploy models at scale.

- AI is not magic
- AI is algorithms + data + team

Data Science Ecosystem



We've been recognized for our vision

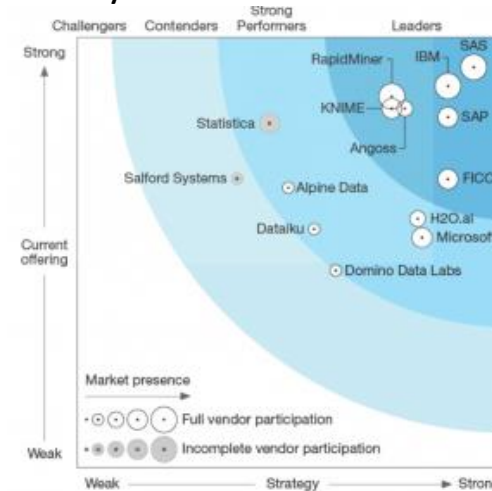
Gartner Magic Quadrant 2017 Data Science Platforms



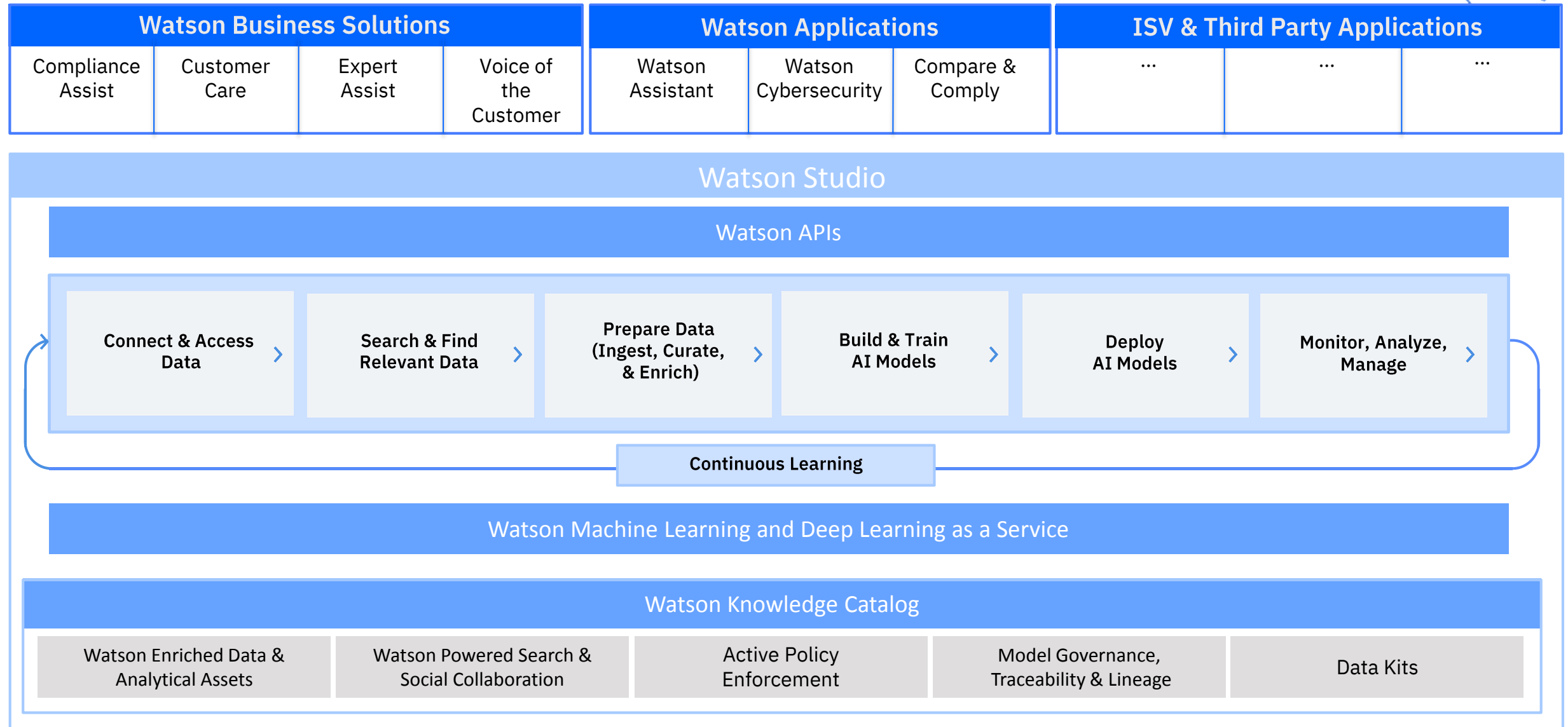
DeveloperWeek 2017 Devie



Forrester Wave 2017 Predictive Analytics & Machine Learning



Watson: AI for Smarter Business



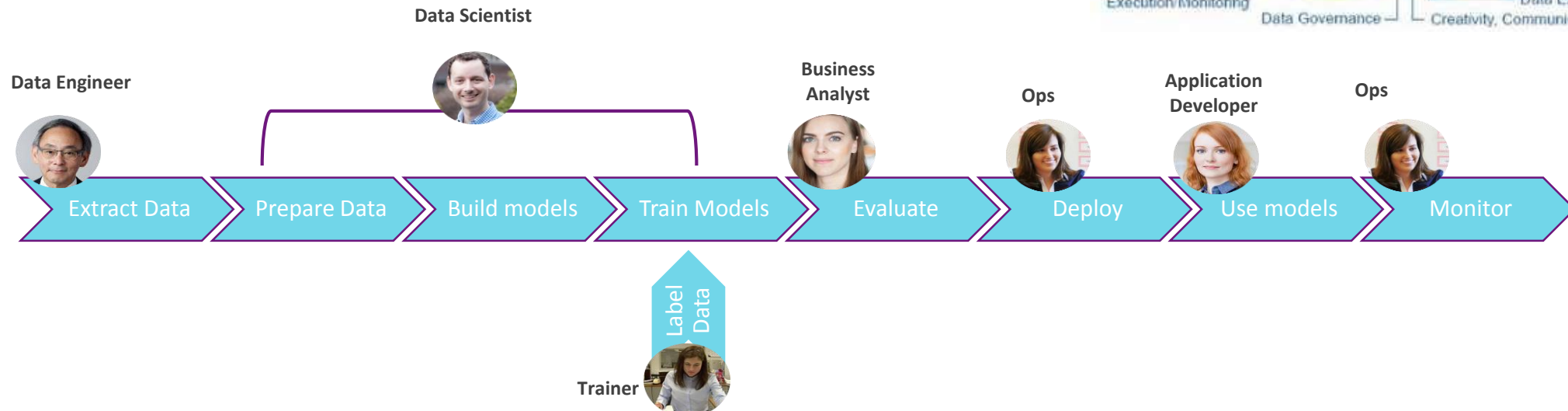
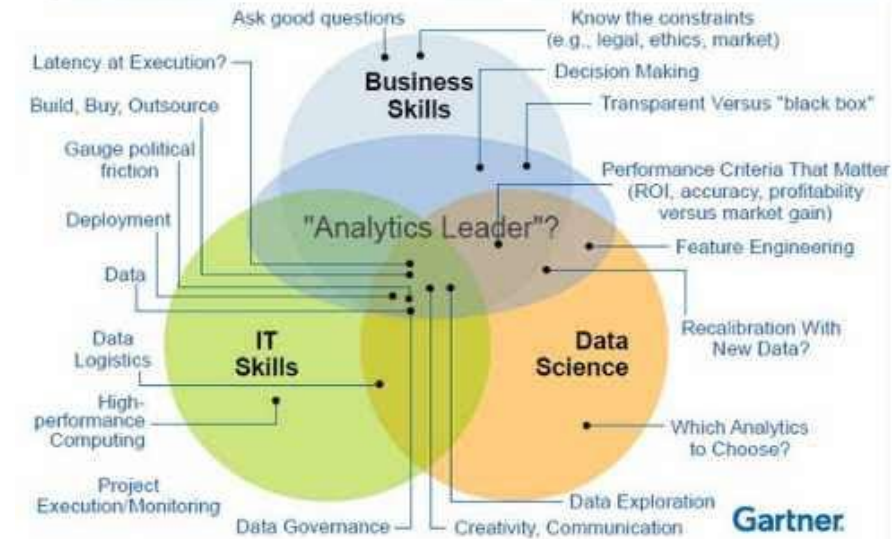
Learns from Small data

Data Science is a Team Sport

■ Building ML-infused apps requires multiple skillsets:

- Define an ML model
- Store, manage, update training data
- Manage lifecycle of the trained model
- Ability to do inferencing on the trained model(s)

Driving the Success of Data Science Solutions: Skills, Roles and Responsibilities ...



Watson Studio

Built for AI teams – enabling team productivity and collaboration



Tanya
Domain Expert

Her Job:

To transfer knowledge to Watson for a successful user experience.

What she does:

- Range of domain knowledge and uses that to teach Watson and develop custom models
- As Tanya gains more experience she optimizes her knowledge to teach Watson to design better end-user experiences.

Sometimes known as:

Subject matter expert, content strategist.



Mike
Data Scientist

His Job:

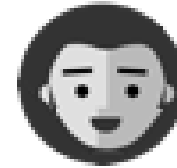
Transform data into knowledge for solving business problems.

What he does:

- Runs experiments to build custom models that solve business problems.
- Use techniques such as Machine Learning or Deep Learning and works with Tanya to validate success of trained models.

Sometimes known as:

ML/DL engineer, Modeler, Data Miner



Ed
Data Engineer

His Job:

Architects how data is organized and ensures operability

What he does:

- Builds data infrastructure and ETL pipelines. Works with Spark, Hadoop, and HDFS.
- Works with data scientist to transform research models into production quality systems.

Sometimes known as:

Data infrastructure engineer



Deb
The Developer

Her Job:

Builds AI application that meet the requirements of the business.

What she does:

- Starts PoCs which includes gathering content, dialog building and model training
- Focus is on app building for the team or company to use. Will handle ML Ops as needed

Sometimes known as:

Front-end, back-end, full stack, mobile or low-code developer

Watson Studio

Supporting the end-to-end AI workflow

Connect &
Access Data

Connect and discover content from multiple data sources in the cloud or on premises. Bring **structured** and **unstructured** data to one toolkit.

Search and Find
Relevant Data

Find data (structured, unstructured) and AI assets (e.g., ML/DL models, notebooks, Watson Data Kits) in the **Knowledge Catalog** with intelligent search and giving the right access to the right users.

Prepare Data
for Analysis

Clean and prepare your data with **Data Refinery**, a tool to create data preparation pipelines visually. Use popular open source libraries to prepare unstructured data.

Build and Train
ML/DL Models

Democratize the creation of ML and DL models. Design your AI models **programmatically** or **visually** with the most popular **open source** and IBM ML/DL frameworks or leverage transfer learning on **pre-trained** models using **Watson tools** to adapt to your business domain. Train at scale on **GPUs** and **distributed** compute

Deploy Models

Deploy your models easily and have them **scale automatically** for online, batch or streaming use cases

Monitor, Analyze
and Manage

Monitor the performance of the models in production and trigger automatic retraining and redeployment of models. Build **Enterprise Trust** with Bias Detection, Mitigation Model **Robustness** and Testing Service Model **Security**.

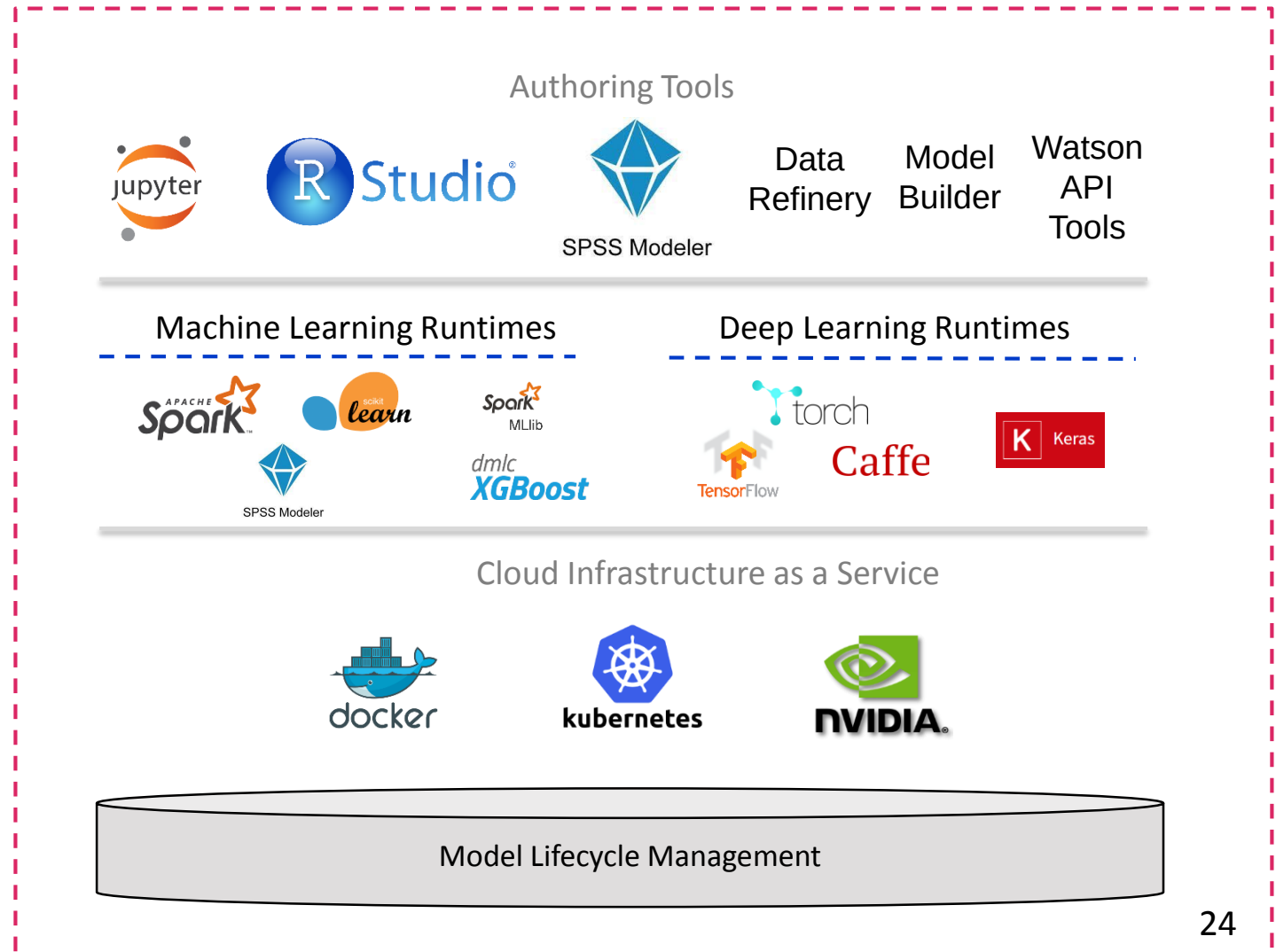
Watson Studio

Comprehensive set of tools for the end-to-end AI workflow

- Create, collaborate, deploy, and monitor
- Best of breed open source & IBM tools
- Code (R, Python or Scala) and no-code/visual modeling tools

- Most popular open source frameworks
- IBM best-in-class frameworks

- Fully managed service
- Container-based resource management
- Elastic pay as you go CPU/GPU power



■ Existing Customer



[Link to Case Study](#)

Geo: Nordics (Europe)

Sector: Commercial – Media & Entertainment

■ Background

- GroupM is the world's largest media investment group with more than \$102bn billings (RECMA, 2016) and 24,000 employees across 81 countries
- They are a broker of digital advertisement, specialised in banner placing in digital media (web, cell phones...)
- Their Nordics team has ~8 data scientists

■ Business Problem

- GroupM needs to know when and where to place advertisements most effectively and how much to charge for it.
- To do so, they were using manual forecasting (based on R) and they were encountering difficulties to scale out the process.
- They were looking for a way to effectively automate their analysis.

■ Solution

- With Watson Studio, Group M was empowered to:
 - To feed data into one single platform, in a structured way.
 - To develop models with common tooling and reuse existing assets to accelerate the development of new use cases.
 - To consume their models as micro-services through rest APIs.

Core Attributes of Watson Studio/ Data Science Experience



IBM Data Science Experience

Community

- Find tutorials and datasets
- Connect with Data Scientists
- Ask questions
- Read articles and papers
- Fork and share projects

Open Source

- Code in Scala/Python/R/SQL
- Jupyter Notebooks
- RStudio IDE and Shiny
- Apache Spark
- Your favorite libraries

IBM Added Value

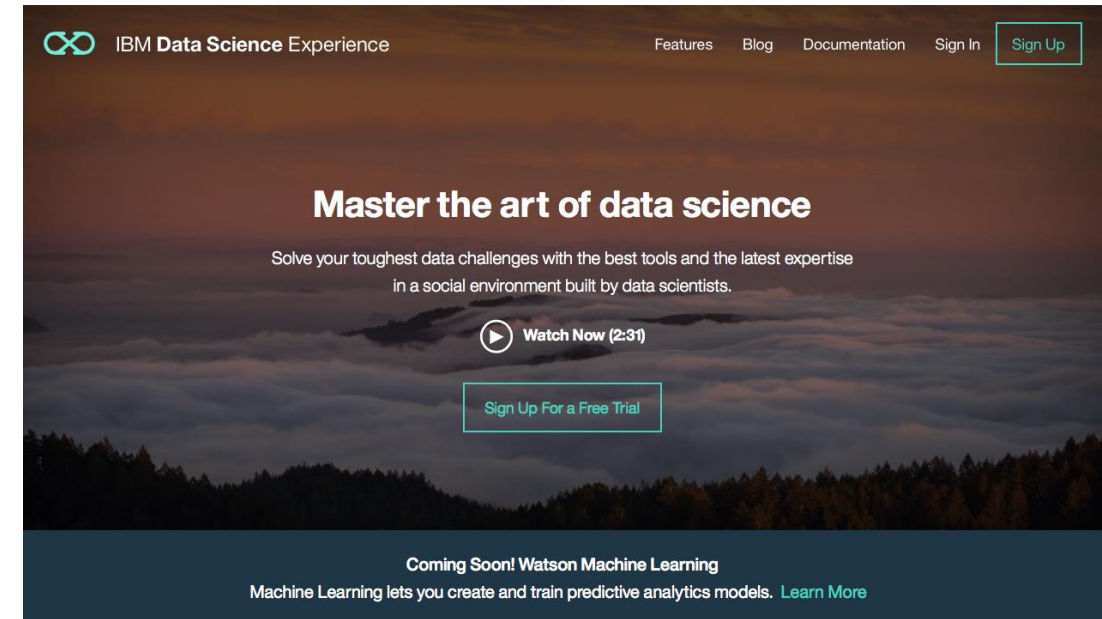
- IBM Machine Learning*
- SPSS Modeler Canvas*
- Prescriptive Analytics - DOpexcloud
- Projects and Version Control
- Managed Spark Service

Kubernetes & Docker based infrastructure (IBM Cloud / IBM Cloud Private)

* beta

Core Attributes of Watson Studio/ Data Science Experience

- **Data Science as a Team Sport**
Lets *data scientists/engineers, analysts*, stakeholders *collaborate* to collect, share, explore, *analyze data* in order to *derive insights and train models*, and *share or deploy resulting assets*
- **Projects** - collaborate as team or work individually
- **Jupyter Notebooks + IBM value add**
 - Integrated in Projects with access control
 - *Spark integration with R/Python/Scala kernels*
 - *Versions, comments, share link, publish to GitHub*
 - *PixieDust, Brunel, ...*
- **Machine Learning** *integrated in Projects:*
Use ML Wizard and Flows to train Models
- **RStudio** *integrated with Spark*
- **DSX Integrates with Data in many places**
 - Object Storage (SWIFT now, new Cloud Object Storage soon)
 - Watson Data Platform Services and WDP Catalog
 - Message Hub and IBM Streaming Analytics
 - Can call any IBM service, e.g. Watson, Quantum, etc
 - Third party data services on premise or on other clouds
- **Built on the IBM Cloud platform**



Learn

Get started or get better with built-in learning.



Create

Use the best of open source tooling with IBM innovation.



Collaborate

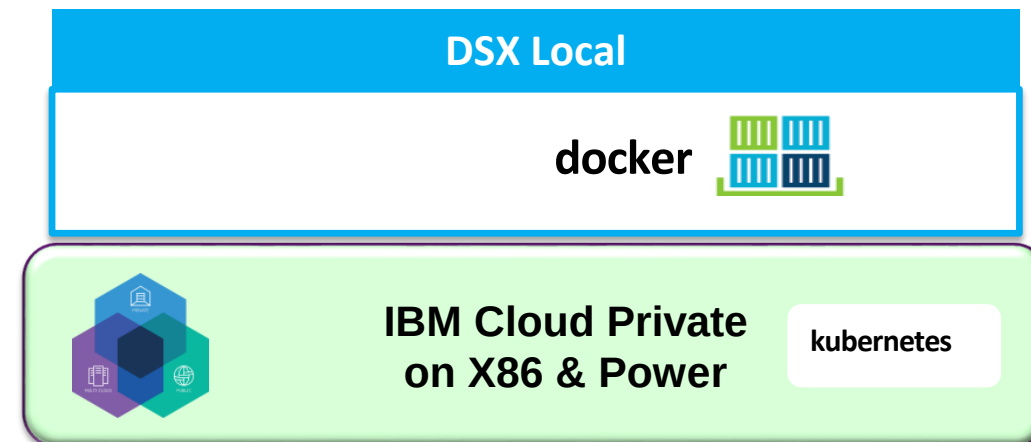
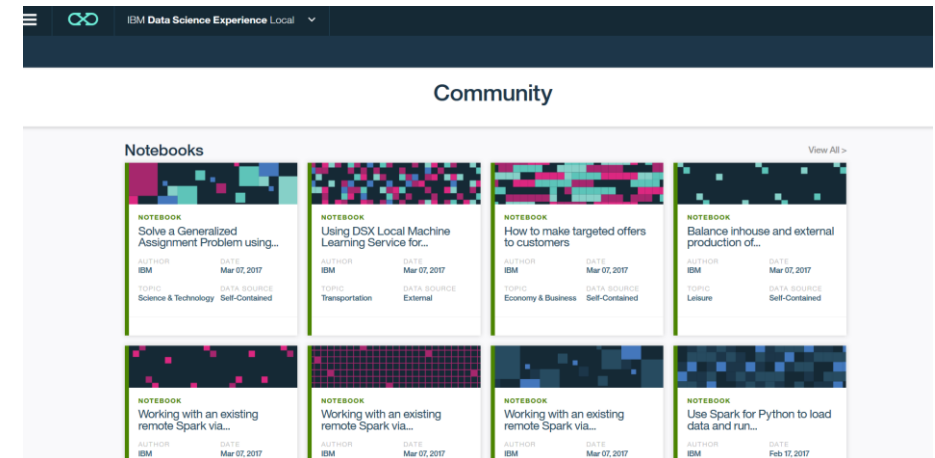
Work smarter using community, work faster with your team.

Try it yourself at <https://datascience.ibm.com>

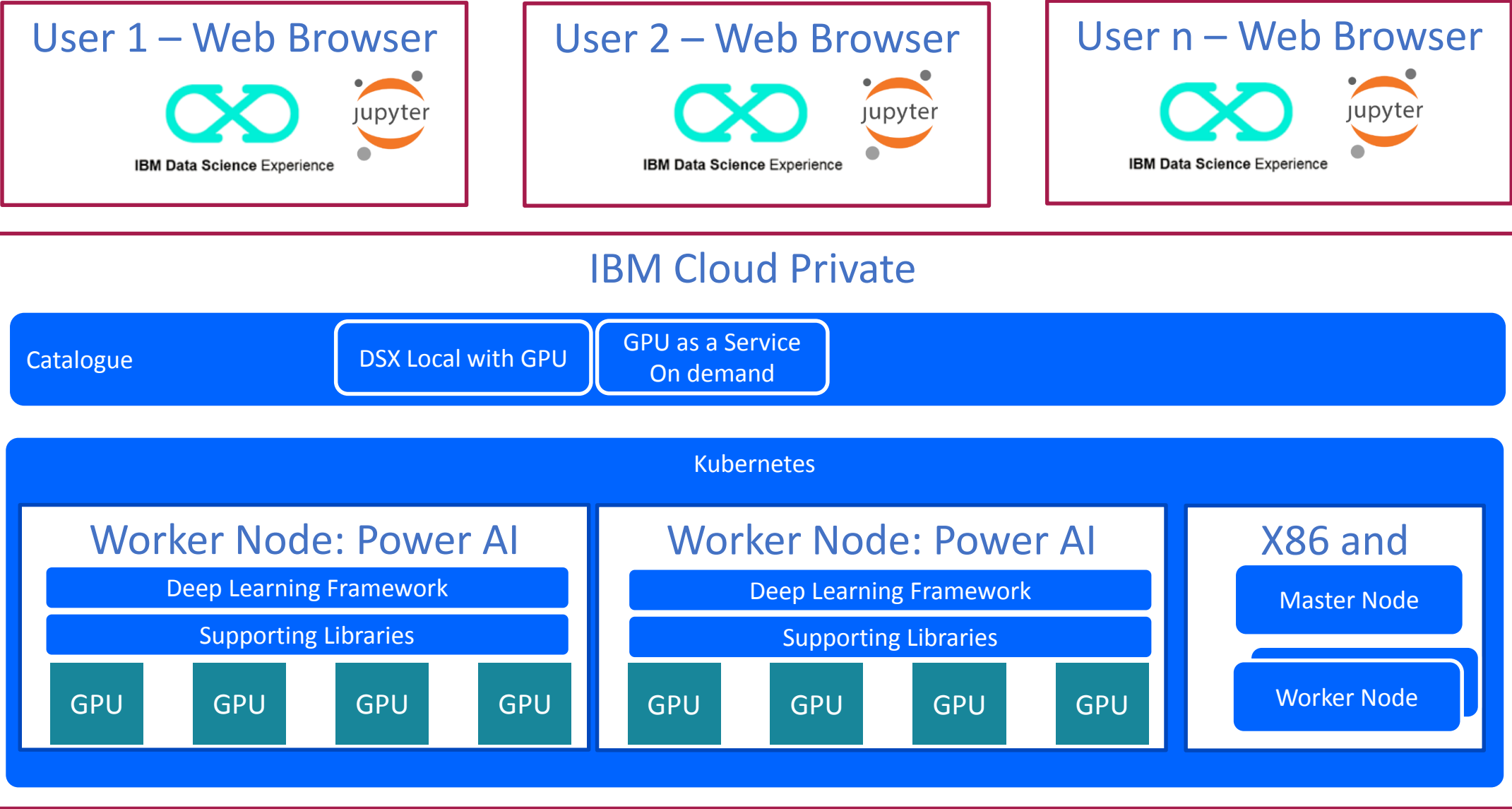
Core Attributes of Watson Studio/ Data Science Experience

- IBM Watson Studio (aka DSX) is available

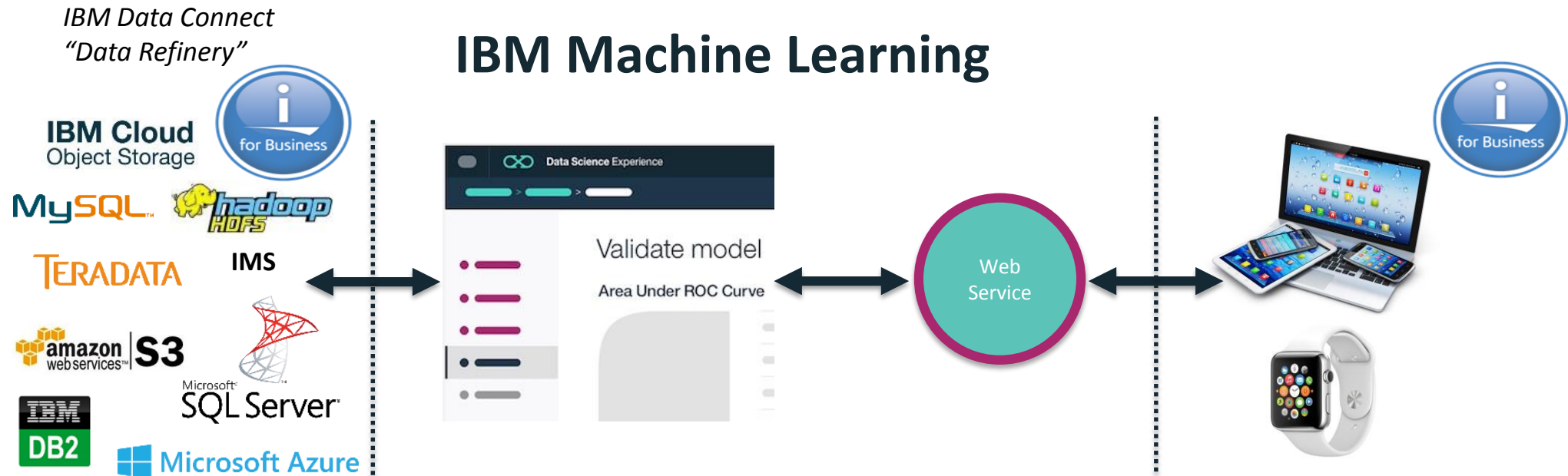
- ☐ As a cloud offering aka **Watson Studio**
- ☐ As a desktop application
 - Free, disconnected mode
- ☐ As an on-premises solution
 - ☐ **DSX Local** on x86/Power
 - ☐ Power: Scale-out LC Systems with PowerAI + GPU / Nvlink acceleration
 - ☐ Possible private cloud deployment with IBM Cloud Private



Deployment Example: IBM Private Cloud & DSX Local



Core Attributes of Watson Studio/ Data Science Experience



Data Access:

- Easily connect to Behind-the-Firewall and Public Cloud Data
- Catalogued and Governed Controls through Watson Data Platform

Creating Models:

- Single UI and API for creating ML Models on various Runtimes
- Auto-Modelling and Hyperparameter Optimization

Web Service:

- Real-time, Streaming, and Batch Deployment
- Continuous Monitoring and Feedback Loop

Intelligent Apps:

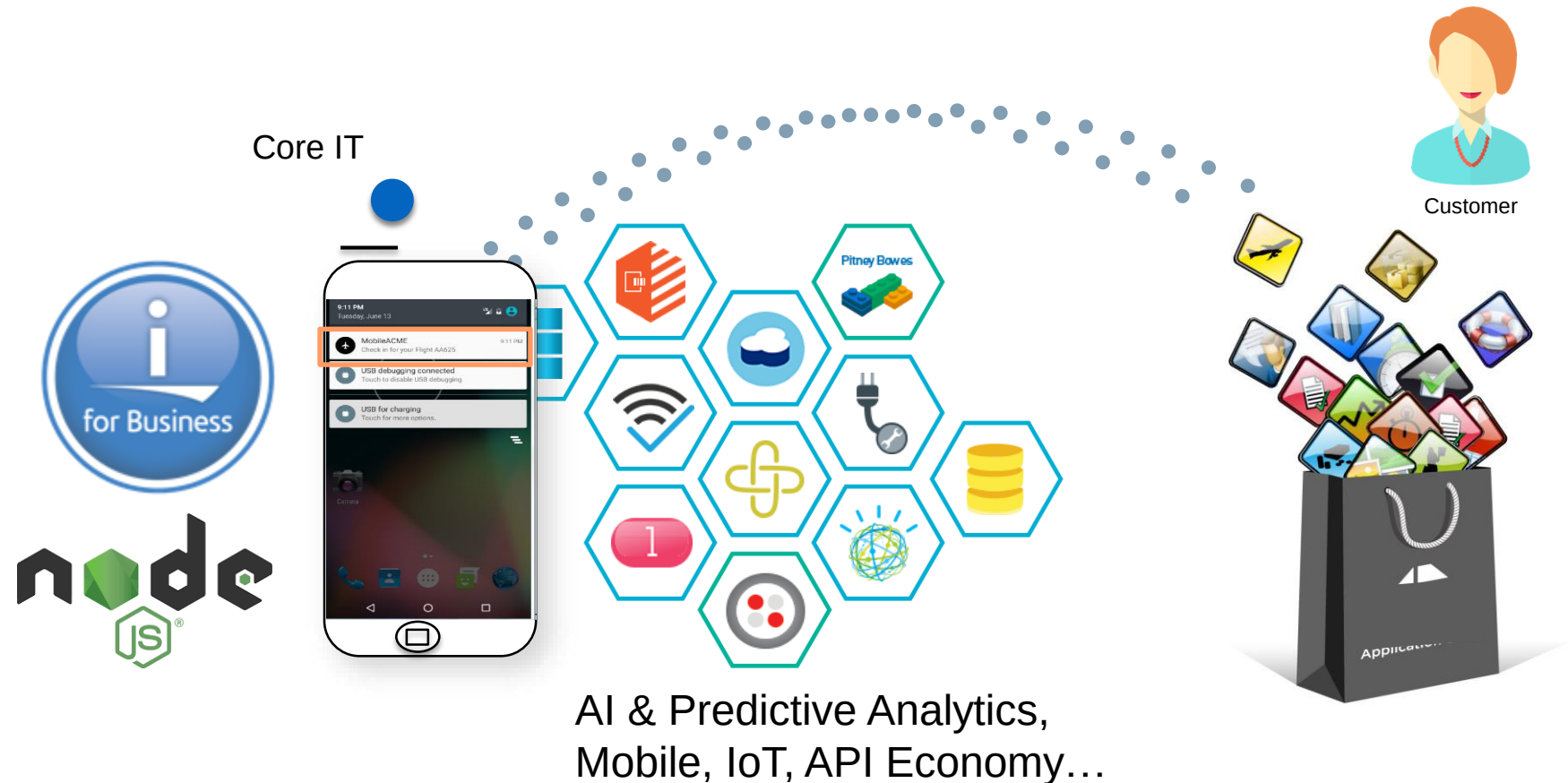
- Integrate ML models with apps, websites, etc.
- Continuously Improve and Adapt with Self-Learning

A background network diagram consisting of numerous grey circular nodes connected by thin grey lines, forming a complex web-like structure.

IBM i & Watson Studio / DSX + Demonstration

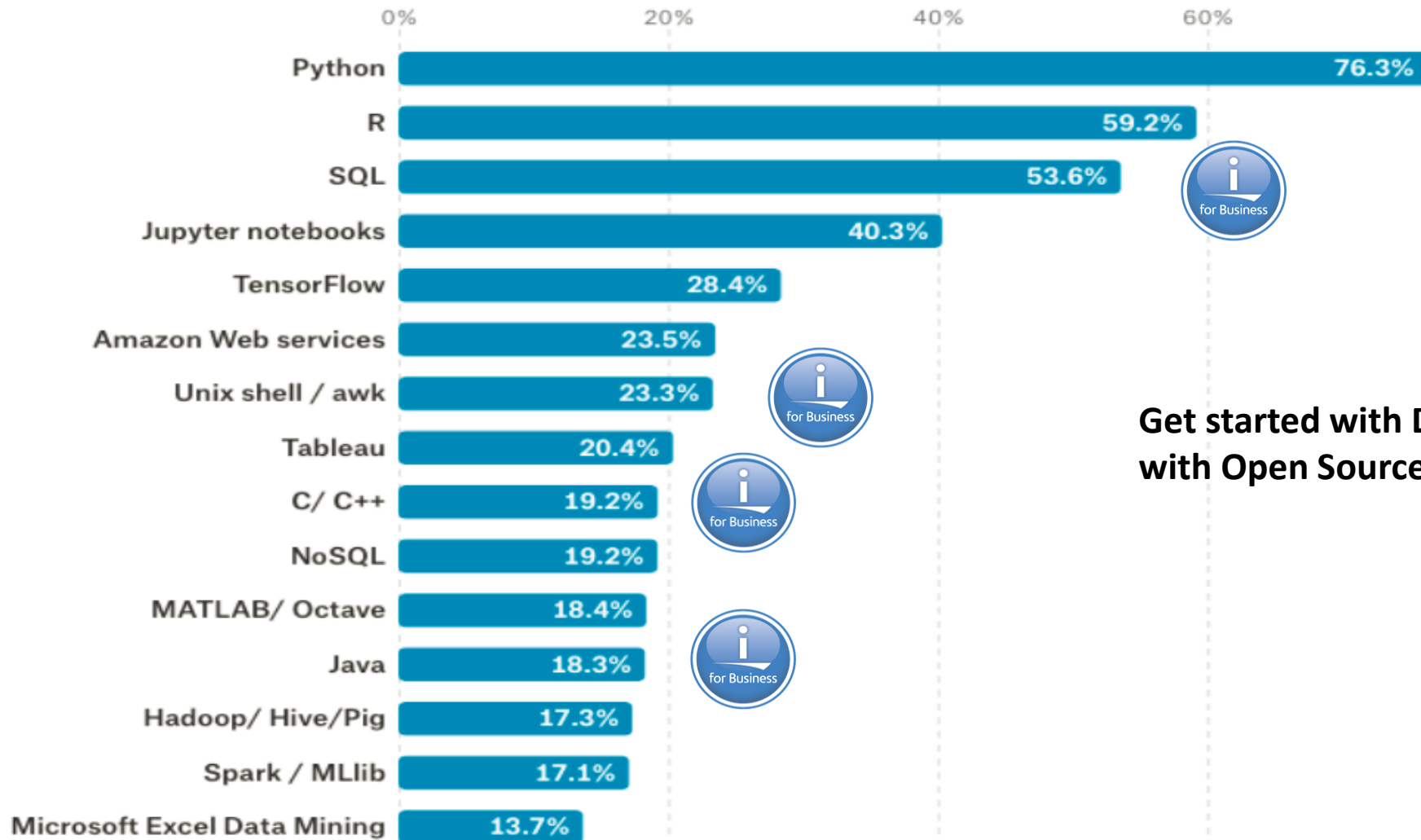
IBM i , AI & Data Science

- ❑ AI = Algorithms + Data (including Data in Db2 for i)
- ❑ Native (RPG) & New Db2 features & Languages available on i facilitates the IBM i $\leftrightarrow$ Watson / AI Integration
- ❑ **Work & prepare directly your data using SQL, Python, Java, etc. directly on IBM i !**
- ❑ Build intelligence & predictive capabilities using Watson Data Platform (including Studio & Data Refinery) & Machine learning techniques



Data Science tools & technologies

Kaggle 2017 Data Science Tools Survey



Get started with Data Science on IBM i
with Open Source Technologies

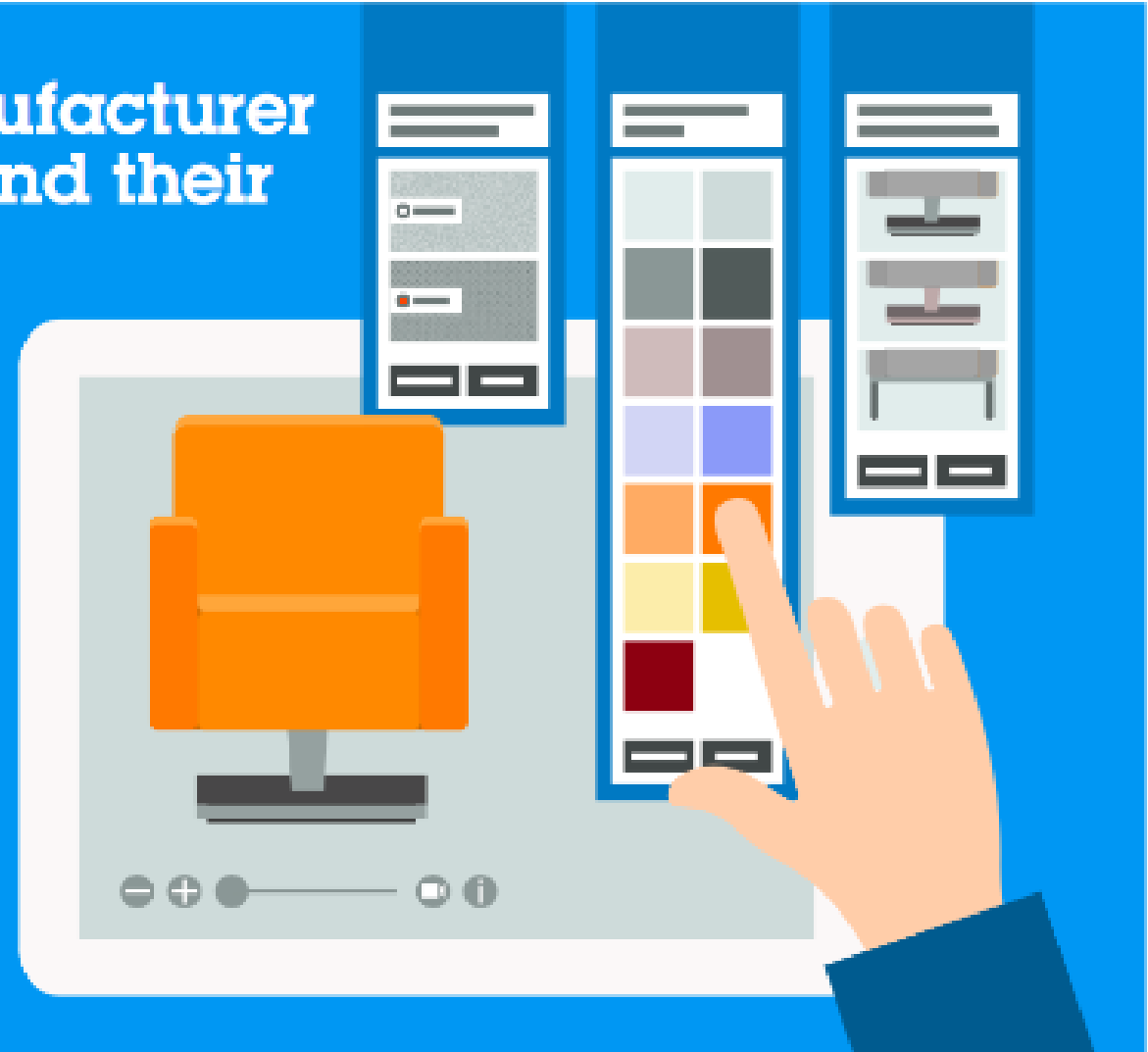
Ready for AI: Connecting your business to the future



The art of fine seating

How does luxury manufacturer JORI help customers find their perfect furniture?

-  Offers 100+ seating frames, fabrics and finishes
-  Wanted to help customers visualize possible combinations
-  Created a 3D configurator with open source software on IBM® i
-  Inspires customers to create their ideal designs
-  50% faster deliveries, as configurator accelerates manufacturing
-  Will use IBM Watson® cognitive technology to help consumers find their preferred fabric



Ready for AI: Connecting your business to the future



The art of fine seating

- Offers 100+ seating frames, fabrics and finishes
- Wanted to help customers visualize possible combinations
- Created a 3D configurator with open source software on IBM® i
- Inspires customers to create their ideal designs
- 50% faster deliveries, as configurator accelerates manufacturing
- Will use IBM Watson® cognitive technology to help consumers find their preferred fabric



The most integrated data platform for business

- Integrated Database (Db2 for i)
- Integrated Web Services

Modern, Open Source applications development

- Python, Node.JS, Ruby
- PHP
- Mobile Application Development



Data Connect
IBM Db2 for i

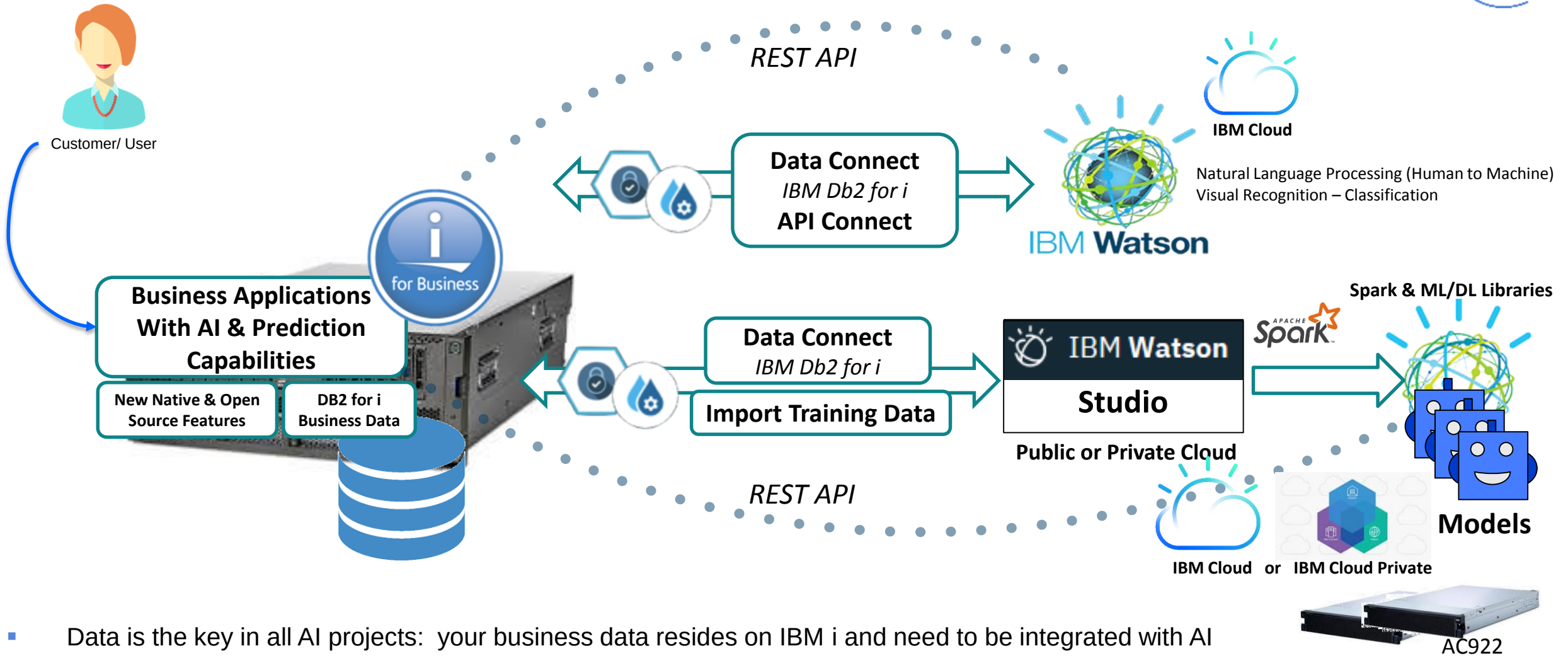
Secure Connectors to Watson for AI

- Data Connect for Db2 for i
- SQL
- Python, Node.JS
- Free form RPG



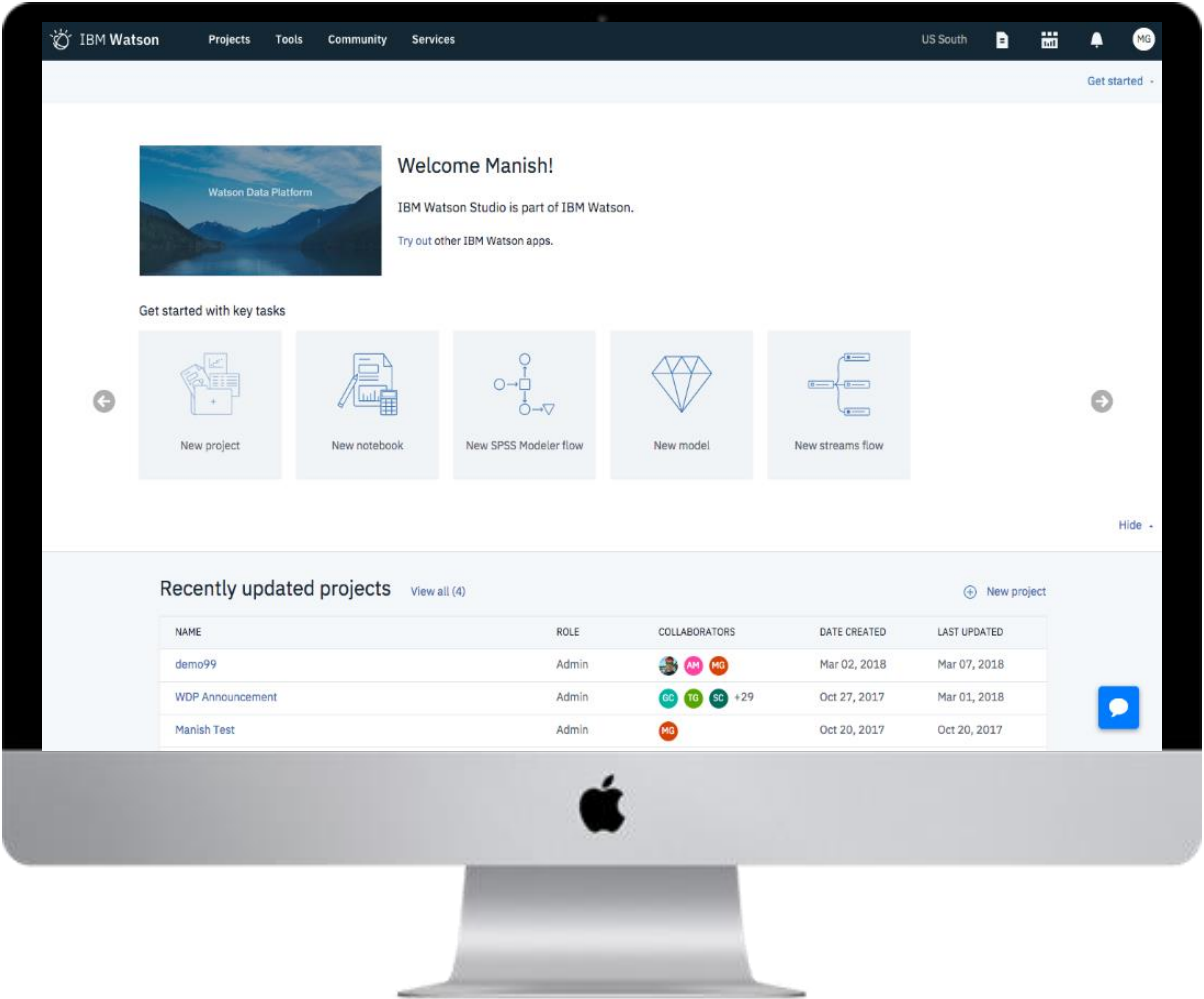
IBM Watson

IBM i & Artificial Intelligence



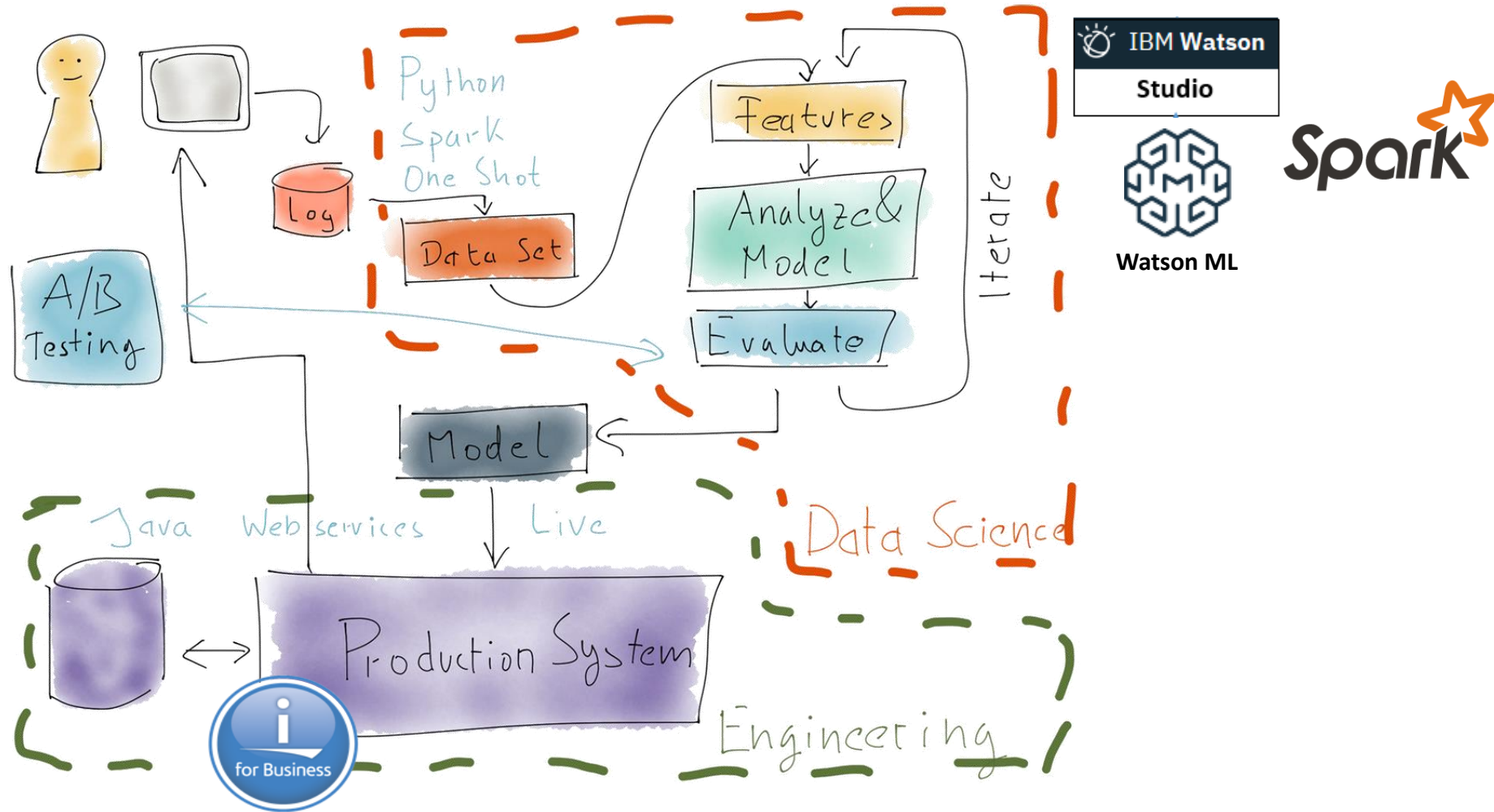
- Data is the key in all AI projects: your business data resides on IBM i and need to be integrated with AI
- Use pre-trained & customizable models with IBM Watson (Developer Cloud) services in IBM Cloud
- **Build your own use case & business specifics models with IBM Watson Studio** - IBM Cloud / on prem (DSX Local w/ Cloud Private)

Demo



Machine Learning & IBM i Demo:

Predict outdoor equipment purchase

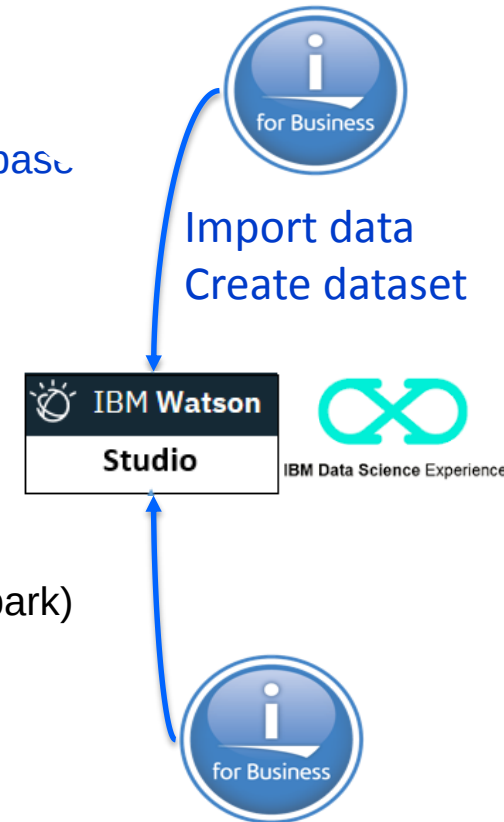


Machine Learning & IBM i Demo:

Predict outdoor equipment purchase



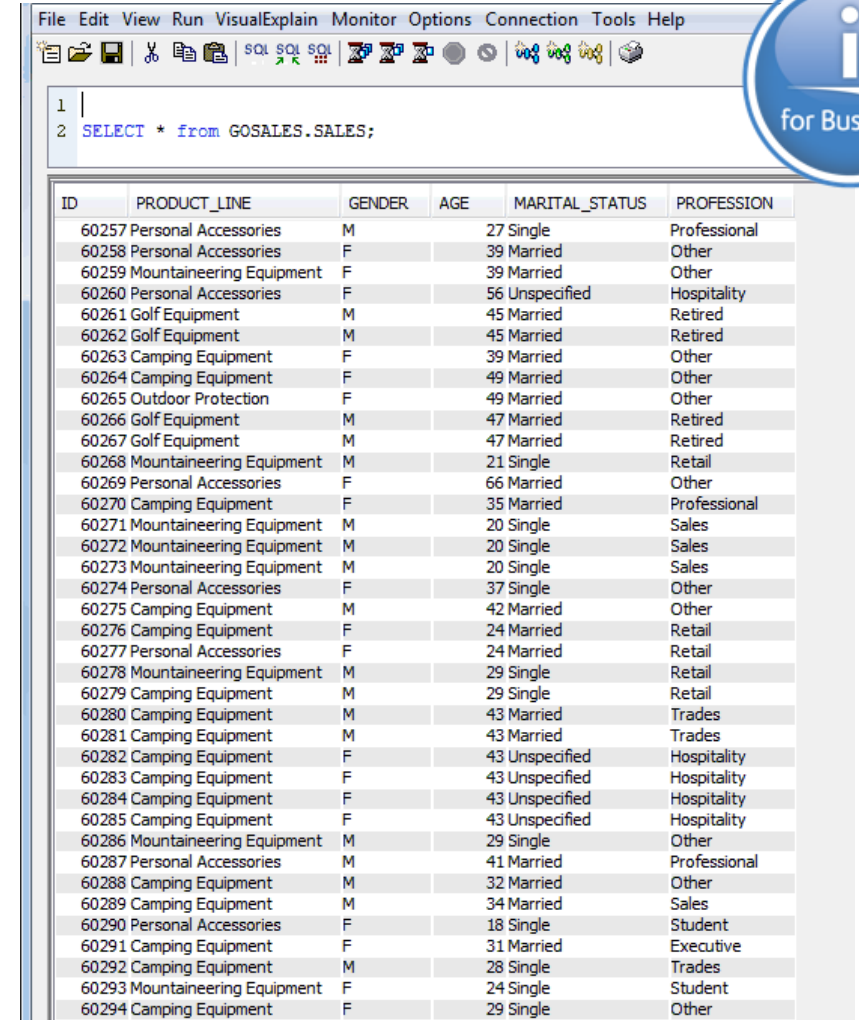
1. Export or connect your data on IBM i
 - Use Access Client Solutions (ACS) for CSV Export or create a jdbc Connector to your Db2 for i database.
 - In that demonstration, one table (GOSALES/SALES) containing historical data - outdoor equipment purchases -- Alternative: Direct connection to your IBM i with a DSX "Connection" !
2. Work on your Data with Watson Studio / DSX Local
 - Data **visualization**, **cleaning**, Data **Refinery**, feature engineering – Complement Watson Explorer
 - Jupyter (R, Scala, Python) or R Studio - Data science & ML/DL Libraries (Spark ML, Tensorflow, Panda, etc)
3. Create & Evaluate your Machine Learning Models
 - Demo: Machine learning with IBM Machine Learning with Automatic Model Builder or Jupyter (PySpark)
 - Machine Learning techniques: Classification - Purchase prediction based on client features
4. Deploy your predictive models & publish it as a REST API
5. Augment your IBM i applications
 - REST API Calls for any programs. In our case, Node-RED & Node.js (5733OPS)
6. Monitor & re-evaluate your models



Original Tutorial : [here](#)

AI = Data, Data, and Data

- ❑ AI usually requires quantity & quality
- ❑ Depends on your business objectives & required precision
- ❑ ML (including DL) techniques choice impact the result
- ❑ AI = Data, Data, Data & skills (Data Science)
- ❑ Watson Studio / DSX can assist you in that modeling & training phases – demo
- ❑ In our simple case, we choose a classification algorithm for predicting the next purchase (label = PREDICT_LINE column) of a customer based on his characteristics (features = GENDER, AGE, MARITAL_STATUS, PROFESSION)
- ❑ Training the model ⇔ executing the classification algorithm against our historical Db2 data and compare it to the real PRODUCT_LINE value (supervised training). The training framework will adjust parameters to make it more accurate. At the end, the model is trained based on this data.
- ❑ It doesn't mean that the predictive function is precise enough. This relevance has to be determined by the Line of business.



The screenshot shows a SQL query interface with a menu bar (File, Edit, View, Run, Visual Explain, Monitor, Options, Connection, Tools, Help) and a toolbar. The query editor contains two lines of SQL code:

```
1 |  
2 SELECT * from GOSALES.SALES;
```

Below the query editor is a table with the following columns: ID, PRODUCT_LINE, GENDER, AGE, MARITAL_STATUS, and PROFESSION. The table contains 28 rows of data, representing customer records with their purchase history and demographic information.

ID	PRODUCT_LINE	GENDER	AGE	MARITAL_STATUS	PROFESSION
60257	Personal Accessories	M	27	Single	Professional
60258	Personal Accessories	F	39	Married	Other
60259	Mountaineering Equipment	F	39	Married	Other
60260	Personal Accessories	F	56	Unspecified	Hospitality
60261	Golf Equipment	M	45	Married	Retired
60262	Golf Equipment	M	45	Married	Retired
60263	Camping Equipment	F	39	Married	Other
60264	Camping Equipment	F	49	Married	Other
60265	Outdoor Protection	F	49	Married	Other
60266	Golf Equipment	M	47	Married	Retired
60267	Golf Equipment	M	47	Married	Retired
60268	Mountaineering Equipment	M	21	Single	Retail
60269	Personal Accessories	F	66	Married	Other
60270	Camping Equipment	F	35	Married	Professional
60271	Mountaineering Equipment	M	20	Single	Sales
60272	Mountaineering Equipment	M	20	Single	Sales
60273	Mountaineering Equipment	M	20	Single	Sales
60274	Personal Accessories	F	37	Single	Other
60275	Camping Equipment	M	42	Married	Other
60276	Camping Equipment	F	24	Married	Retail
60277	Personal Accessories	F	24	Married	Retail
60278	Mountaineering Equipment	M	29	Single	Retail
60279	Camping Equipment	M	29	Single	Retail
60280	Camping Equipment	M	43	Married	Trades
60281	Camping Equipment	M	43	Married	Trades
60282	Camping Equipment	F	43	Unspecified	Hospitality
60283	Camping Equipment	F	43	Unspecified	Hospitality
60284	Camping Equipment	F	43	Unspecified	Hospitality
60285	Camping Equipment	F	43	Unspecified	Hospitality
60286	Mountaineering Equipment	M	29	Single	Other
60287	Personal Accessories	M	41	Married	Professional
60288	Camping Equipment	M	32	Married	Other
60289	Camping Equipment	M	34	Married	Sales
60290	Personal Accessories	F	18	Single	Student
60291	Camping Equipment	F	31	Married	Executive
60292	Camping Equipment	M	28	Single	Trades
60293	Mountaineering Equipment	F	24	Single	Student
60294	Camping Equipment	F	29	Single	Other

Watson Studio / DSX Dashboard



Projects

Data Assets
from import or
Direct datasource
connection

Code (Notebooks)
Work on the data
programmatically

Models
(Machine Learning)

Data Flows
(Data Refinery)
from CSV
or direct Connections

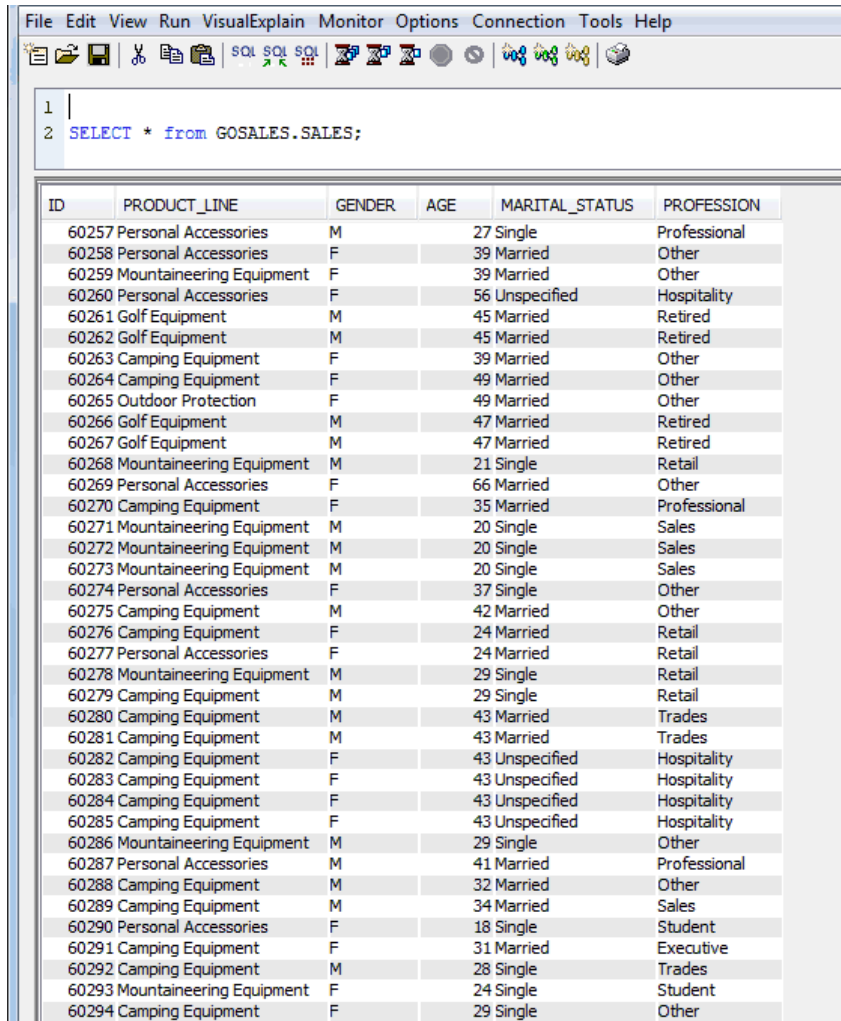
The screenshot displays the IBM Watson Studio / DSX Dashboard. The top navigation bar includes 'IBM Watson', 'Projects', 'Tools', 'Community', and 'Services'. Below the navigation bar, the 'My Projects' section is active, showing 'Project IBM i'. The main content area is divided into several sections:

- Data assets:** A table listing various data assets and connections. The table has columns for NAME, TYPE, SERVICE, CREATED BY, LAST MODIFIED, and ACTIONS. Assets include connections to IBM COS and various CSV files.
- Visual recognition models:** A section for managing visual recognition models, with a '+ New visual recognition model' button.
- Notebooks:** A table listing notebooks. The table has columns for NAME, SHARED, SCHEDULED, STATUS, LANGUAGE, LAST EDITOR, LAST MODIFIED, and ACTIONS. A notebook titled 'From spark ml model to online scoring with python' is listed.
- Streams flows:** A section for managing streams flows, with a '+ New streams flow' button.
- Dashboard:** A section for managing dashboards, with a '+ New dashboard' button. A message states 'you currently have no dashboard'.
- Models:** A section for managing machine learning models, with a '+ New model' button. The table has columns for NAME, STATUS, TYPE, RUNTIME, LAST MODIFIED, and ACTIONS.

Data flows					+ New data flow
NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS	
APPDB-GOSALES-SALES_flow	Data flow	BENOIT MAROLLEAU	11 May 2018, 11:47:10 am	⋮	
GoSalesTraining.csv_flow_Final	Data flow	BENOIT MAROLLEAU	9 Apr 2018, 1:34:59 pm	⋮	
GoSales_Tx_NaiveBayes_IBMi.csv_flow	Data flow	BENOIT MAROLLEAU	9 Apr 2018, 1:24:32 pm	⋮	

Extract Data

- The goal is to extract data for building & training our predictive model.
- The data is in the GOSALES library, table SALES, containing past sales summary (not a normalized table: requires SQL Preprocessing)
- Extraction using ACS & SQL

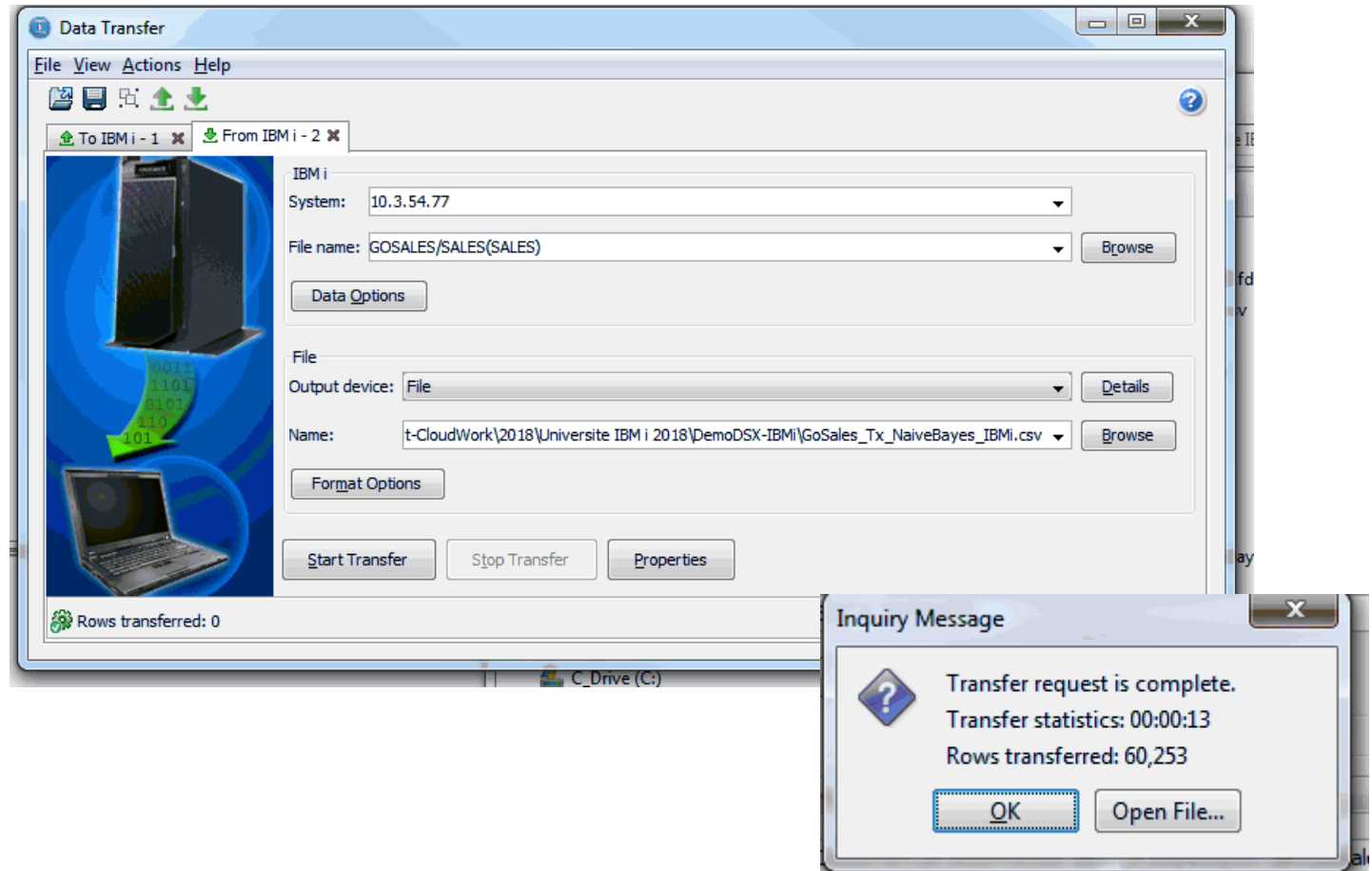


The screenshot shows a SQL client window with a menu bar (File, Edit, View, Run, Visual Explain, Monitor, Options, Connection, Tools, Help) and a toolbar. The query editor contains the following SQL statement:

```
1 |  
2 SELECT * from GOSALES.SALES;
```

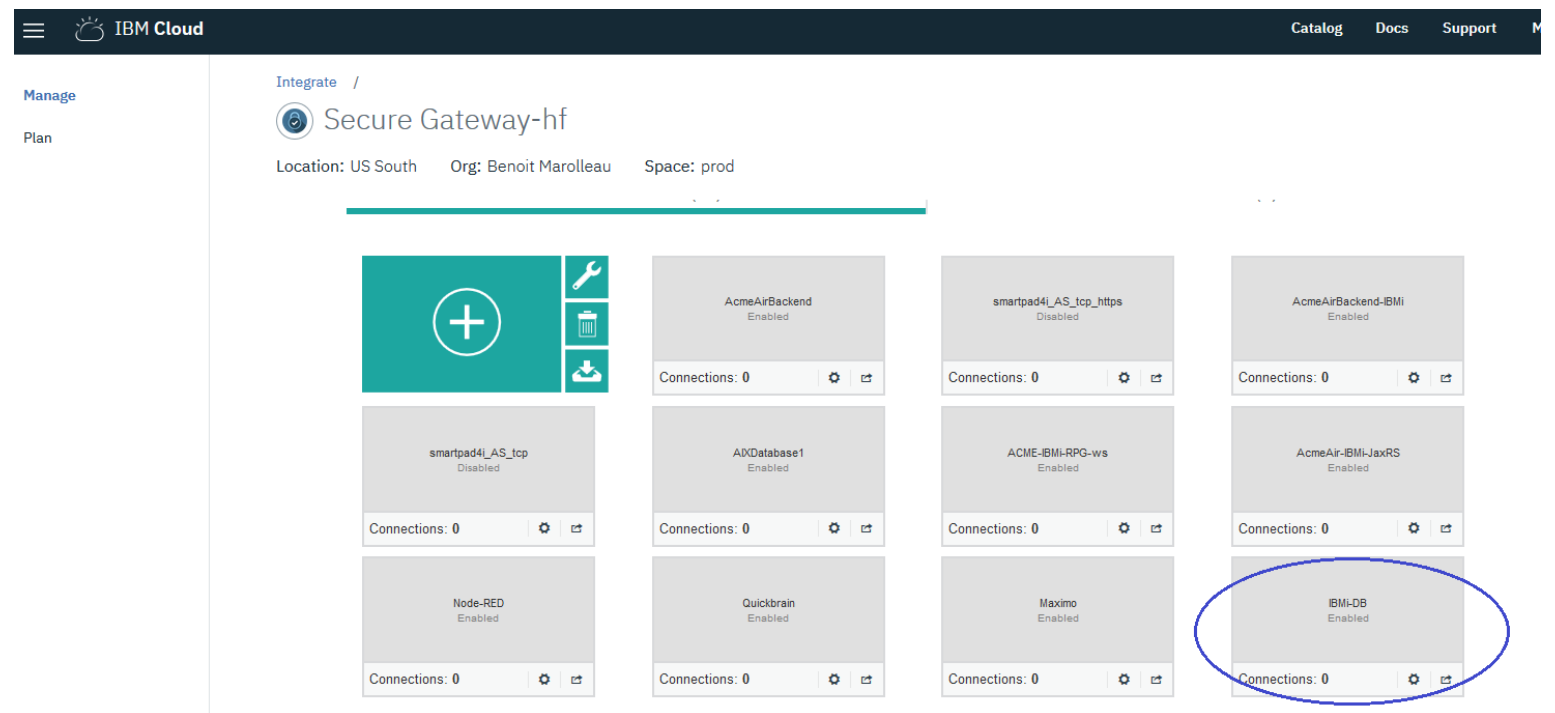
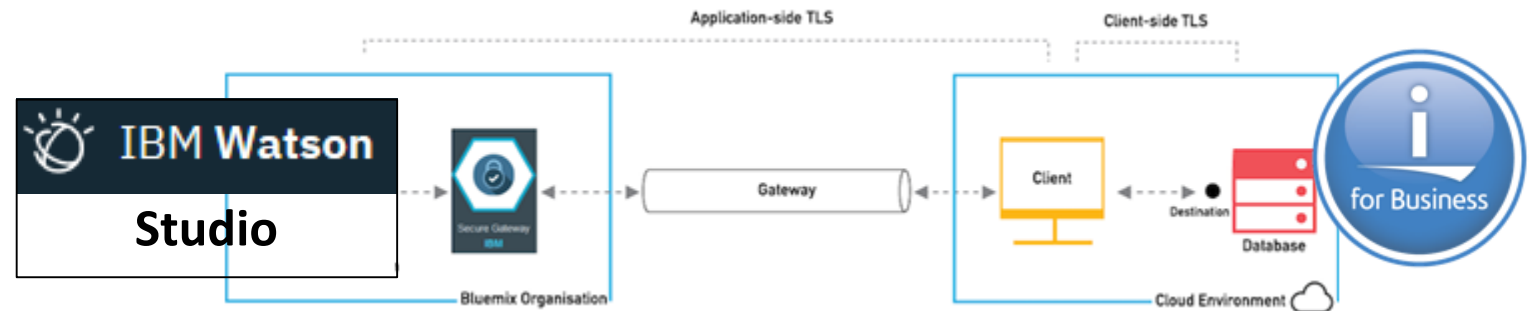
The results are displayed in a table with the following columns: ID, PRODUCT_LINE, GENDER, AGE, MARITAL_STATUS, and PROFESSION. The table contains 30 rows of data.

ID	PRODUCT_LINE	GENDER	AGE	MARITAL_STATUS	PROFESSION
60257	Personal Accessories	M	27	Single	Professional
60258	Personal Accessories	F	39	Married	Other
60259	Mountaineering Equipment	F	39	Married	Other
60260	Personal Accessories	F	56	Unspecified	Hospitality
60261	Golf Equipment	M	45	Married	Retired
60262	Golf Equipment	M	45	Married	Retired
60263	Camping Equipment	F	39	Married	Other
60264	Camping Equipment	F	49	Married	Other
60265	Outdoor Protection	F	49	Married	Other
60266	Golf Equipment	M	47	Married	Retired
60267	Golf Equipment	M	47	Married	Retired
60268	Mountaineering Equipment	M	21	Single	Retail
60269	Personal Accessories	F	66	Married	Other
60270	Camping Equipment	F	35	Married	Professional
60271	Mountaineering Equipment	M	20	Single	Sales
60272	Mountaineering Equipment	M	20	Single	Sales
60273	Mountaineering Equipment	M	20	Single	Sales
60274	Personal Accessories	F	37	Single	Other
60275	Camping Equipment	M	42	Married	Other
60276	Camping Equipment	F	24	Married	Retail
60277	Personal Accessories	F	24	Married	Retail
60278	Mountaineering Equipment	M	29	Single	Retail
60279	Camping Equipment	M	29	Single	Retail
60280	Camping Equipment	M	43	Married	Trades
60281	Camping Equipment	M	43	Married	Trades
60282	Camping Equipment	F	43	Unspecified	Hospitality
60283	Camping Equipment	F	43	Unspecified	Hospitality
60284	Camping Equipment	F	43	Unspecified	Hospitality
60285	Camping Equipment	F	43	Unspecified	Hospitality
60286	Mountaineering Equipment	M	29	Single	Other
60287	Personal Accessories	M	41	Married	Professional
60288	Camping Equipment	M	32	Married	Other
60289	Camping Equipment	M	34	Married	Sales
60290	Personal Accessories	F	18	Single	Student
60291	Camping Equipment	F	31	Married	Executive
60292	Camping Equipment	M	28	Single	Trades
60293	Mountaineering Equipment	F	24	Single	Student
60294	Camping Equipment	F	29	Single	Other



Alternative: Direct Connection to IBM i

- ❑ Create a Connection for direct access to your IBM i database.
- ❑ Optionally use a **Secure Gateway** Service for a secure tunneled connection to your system



Alternative: Direct Connection to IBM i



← New connection (APPDB - Db2 for i)

Connection overview

Name

APPDB

Description

IBM Db2 database for i

Connection details

Hostname or IP Address *

cap-sg-prd-1.integration.ibmcloud.com

Location *

T01FF3B4

Secure Gateway

☐ Use a secure gateway

Port *

16987

DRDA port

Username *

benoit

Password *

.....

ProjectsToolsCommunityServices

Data Refinery

Project IBM i	APPDB	GOSALES
Data assets	Schemas (109)	Tables (20)
APPDB	#LIBRARY	SALES
IBM COS Connection	ACMEAIR	SYSCHKCST
157913b6-7fba-4e95-97bb...	ASN	SYSCOLUMNS
32059d22-bb5e-428b-960...	BENOIT	SYSCST
c85f7031-1cc4-4308-a16b...	CLALEVEE	SYSCSTCOL
	CLPADMIN	SYSCSTDEP
	CO407	SYSFIELDS
	GOSALES	SYSINDEXES
	NULLID	SYSKEYCST
	QADVSEC	SYSKEYS
	QCAEXP	SYSPACKAGE
	QCA400W	SYSREFCST
	QCBLL	SYSTABLEDEP
	QCBLLP	SYSTABLES

Data Asset (from Db2 for i)

 IBM Watson

ProjectsToolsCommunityServices

My Projects / Project IBM i / GoSales_Tx_NaiveBayes_IBMi.csv

Refine

PreviewProfile

Schema: 5 Columns
Preview (1000 rows)

PRODUCT_LINE <i>Type: String</i>	GENDER <i>Type: String</i>	AGE <i>Type: String</i>	MARITAL_STATUS <i>Type: String</i>	PROFESSION <i>Type: String</i>
Personal Accessories	M	27	Single	Professional
Personal Accessories	F	39	Married	Other
Mountaineering Equipment	F	39	Married	Other
Personal Accessories	F	56	Unspecified	Hospitality
Golf Equipment	M	45	Married	Retired
Golf Equipment	M	45	Married	Retired
Camping Equipment	F	39	Married	Other
Camping Equipment	F	49	Married	Other
Outdoor Protection	F	49	Married	Other
Golf Equipment	M	47	Married	Retired
Golf Equipment	M	47	Married	Retired
Mountaineering Equipment	M	21	Single	Retail
Personal Accessories	F	66	Married	Other

×

1010

Data Asset

GoSales_Tx_NaiveBayes_IBMi.csv

Description

No description available for this asset

Tags

No tags available for this asset

Creator: benoit.marolleau@fr.ibm.com

Added: 10:49 AM UTC, 2018/04/09

Size: 2.993 MB



Data Asset (from Db2 for i)



IBM Watson

ProjectsToolsCommunityServices

My Projects / Project IBM i / GoSales_Tx_NaiveBayes_IBMi.csv

Refine

Preview

Profile

Creating data profile

Feel free to continue working or stay here and refresh the page occasionally to get an update on the profile's status. When the profile's ready, you'll be able to view it.

Data Asset

GoSales_Tx_NaiveBayes_IBMi.csv

Description

No description available for this asset

Tags

No tags available for this asset

Creator: benoit.marolleau@fr.ibm.com

Added: 10:49 AM UTC, 2018/04/09

Size: 2.993 MB

Data Asset (from Db2 for i)



IBM Watson

ProjectsToolsCommunityServices

My Projects / Project IBM i / GoSales_Tx_NaiveBayes_IBMi.csv

Refine

PreviewProfile

Current profile

163 classifiers

Last profile

9 Apr 2018 - 12:50 pm

View Log

Columns

5

Rows

5 000

Delete

Update Profile

PRODUCT_LINE

Type: Varchar

GENDER

Type: Varchar

AGE

Type: Varchar

MARITAL_

Type: Varchar

CLASSIFIER DISTRIBUTION

Text

Inferred

100%

0%

0%

Account Number

Address Line 1

Address Line 2

Address Line 3

Airport Code

Showing 5 of 162

View All

CLASSIFIER DISTRIBUTION

Gender

Inferred

100%

0%

0%

Organization Name

Address Line 2

Use Address Line 2

0%

100%

0%

Showing 5 of 162

View All

CLASSIFIER DISTRIBUTION

Code

Inferred

100%

0%

0%

Account Number

Address Line 1

Address Line 2

Address Line 3

Airport Code

Showing 5 of 162

View All

CLASSIFIER DISTRIBUTION

Legal Mari

Inferred

100%

0%

0%

Account N

Address Li

Address Li

Address Li

Airport Co

Showing 5

FREQUENCY

Camping Equipment

Personal Accessories

Mountaineering Equipment

FREQUENCY

M

FREQUENCY

29

28

26

25

27

37

33

FREQUENCY

Married

Single

GoSales_Tx_NaiveBayes_IBMi.csv

Description

No description available for this asset

Tags

No tags available for this asset

Creator: benoit.marolleau@fr.ibm.com

Added: 10:49 AM UTC, 2018/04/09

Size: 3.028 MB

Work on your data: Data Refinery (Data Flow for Cleaning, Labeling...)



- ☐ Work on your data **graphically** using Data Flows & Data Refinery
- ☐ From a CSV data asset or from a Connection to Db2 for i

The screenshot displays the IBM Watson Data Refinery web interface. At the top, a dark navigation bar contains the IBM Watson logo and links for Projects, Tools, Community, and Services. Below this, a breadcrumb trail shows the path: My Projects / Project IBM i / GoSales_Tx_NaiveBayes_IBMi.csv_. To the right of the breadcrumb is a 'Refine' button with a play icon, followed by icons for information, undo, redo, and a grid view. The main content area is titled 'Summary' and shows a data flow from 'GoSales_Tx_NaiveBayes_IBMi.csv' (Source) to 'GoSales_Tx_NaiveBayes_IBMi.csv_shaped.csv' (Output) via '9 Steps' (Data flow). Below the summary, the 'Runs' section is visible, with tabs for 'History' and 'Schedule'. The 'History' tab is active, showing a table of recent runs.

TIMESTAMP	STATUS	DURATION	ROWS READ / WRITTEN	SIZE	INITIATED BY
9 Apr 2018 - 01:18 pm	✓ Completed	21 min 40 sec	60252 / 60252	3.01 MB	benoit.marolleau@fr.ibm....
9 Apr 2018 - 01:01 pm	✓ Completed	13 sec	60252 / 60252	2.16 MB	benoit.marolleau@fr.ibm....



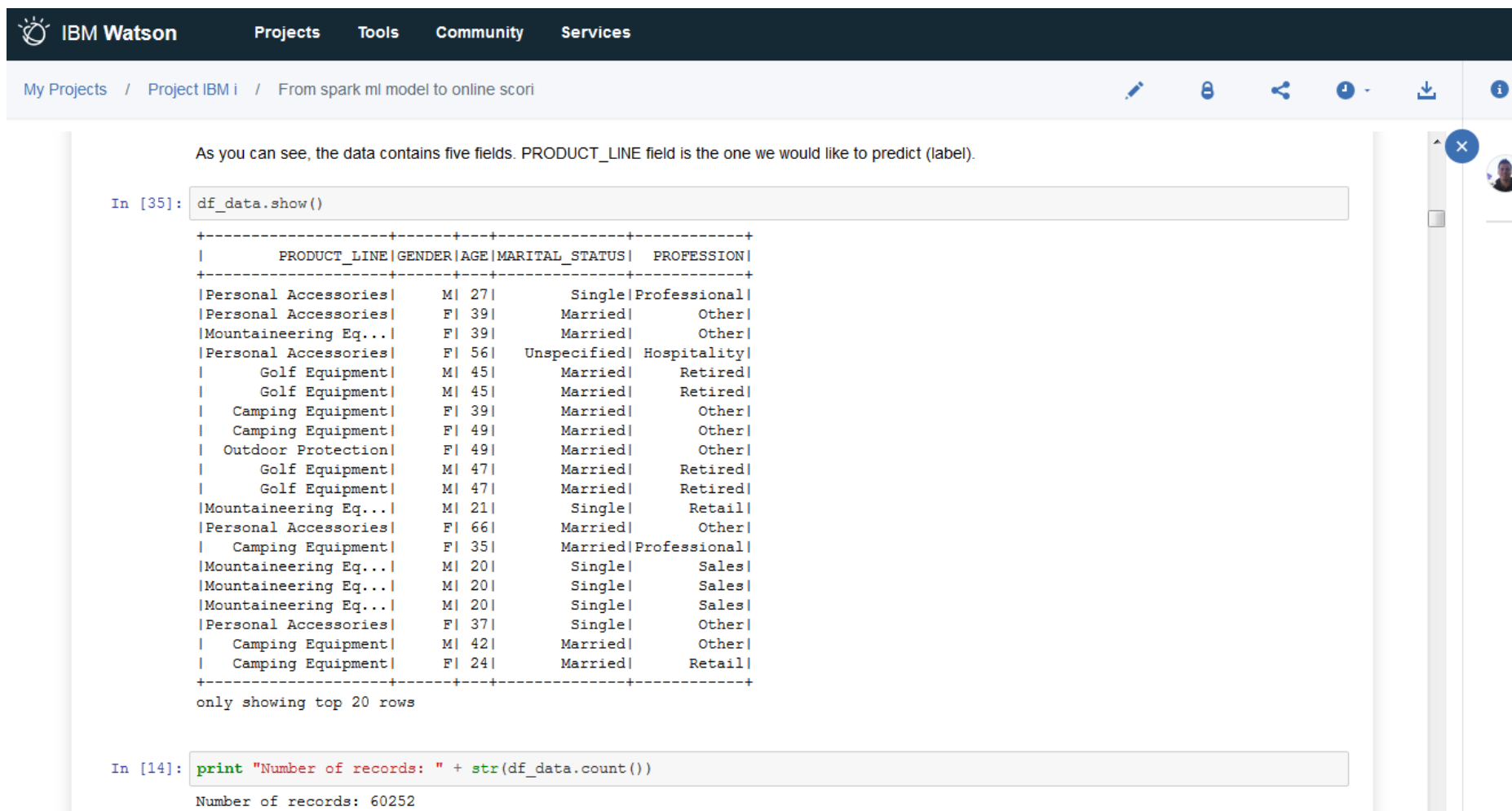
Work on your data: Data Refinery (Data Flow for Cleaning, Labeling...)



SALES						
ID	PRODUCT_LINE	GENDER	AGE	MARITAL_STATUS	PROFESSION	
Type: Integer	Type: String	Type: Char	Type: Smallint	Type: String	Type: String	
60257	Personal Accessories	M	27	Single	Professional	
60258	Personal Accessories	F	39	Married	Other	
60259	Mountaineering Equipment	F	39	Married	Other	
60260	Personal Accessories	F	56	Unspecified	Hospitality	
60261	Golf Equipment	M	45	Married	Retired	
60262	Golf Equipment	M	45	Married	Retired	
60263	Camping Equipment	F	39	Married	Other	
60264	Camping Equipment	F	49	Married	Other	
60265	Outdoor Protection	F	49	Married	Other	
60266	Golf Equipment	M	47	Married	Retired	
60267	Golf Equipment	M	47	Married	Retired	
60268	Mountaineering Equipment	M	21	Single	Retail	

Work on your data: Jupyter Notebook

- ❑ AND/OR : Work on your data using your favorite language & ML/DL libraries & frameworks



IBM Watson Projects Tools Community Services

My Projects / Project IBM i / From spark ml model to online scori

As you can see, the data contains five fields. PRODUCT_LINE field is the one we would like to predict (label).

```
In [35]: df_data.show()
```

PRODUCT_LINE	GENDER	AGE	MARITAL_STATUS	PROFESSION
Personal Accessories	M	27	Single	Professional
Personal Accessories	F	39	Married	Other
Mountaineering Eq...	F	39	Married	Other
Personal Accessories	F	56	Unspecified	Hospitality
Golf Equipment	M	45	Married	Retired
Golf Equipment	M	45	Married	Retired
Camping Equipment	F	39	Married	Other
Camping Equipment	F	49	Married	Other
Outdoor Protection	F	49	Married	Other
Golf Equipment	M	47	Married	Retired
Golf Equipment	M	47	Married	Retired
Mountaineering Eq...	M	21	Single	Retail
Personal Accessories	F	66	Married	Other
Camping Equipment	F	35	Married	Professional
Mountaineering Eq...	M	20	Single	Sales
Mountaineering Eq...	M	20	Single	Sales
Mountaineering Eq...	M	20	Single	Sales
Personal Accessories	F	37	Single	Other
Camping Equipment	M	42	Married	Other
Camping Equipment	F	24	Married	Retail

only showing top 20 rows

```
In [14]: print "Number of records: " + str(df_data.count())
```

Number of records: 60252

Data Preparation / Model Training & Deployment

using a Python Jupyter Notebook



My Projects / Project IBM i / Online Scoring with Spark ML

File Edit View Insert Cell Kernel Help Trusted | Python 2 with Spa

4. Persist model

In this section you will learn how to store your pipeline and model in Watson Machine Learning repository by using python client libraries.

First, you must import client libraries.

Note: Python 2 and Apache® Spark 2.0 or higher is required.

```
In [63]: from repository_v3.mlrepositoryclient import MLRepositoryClient
         from repository_v3.mlrepositoryartifact import MLRepositoryArtifact
```

Authenticate to Watson Machine Learning service on Bluemix.

Action: Put authentication information from your instance of Watson Machine Learning service here.

```
In [64]: wml_credentials={
         "url": "https://ibm-watson-ml.mybluemix.net",
         "access_key": "mN3bwIsi/xpOh08Bd8D8Zah6UF++y/ZqlqTlJDmJQr6dqC83AFnAyPemGJw9js0JHxGxQ3pIogjgEOjN0TGDTCcL0h32gVzPkwMbmHX",
         "username": "75693a81-d133-4c15-8570-3dc0f0f092c1",
         "password": "13cb03fd-2b6b-4b24-a17f-49eb2efaf7b9",
         "instance_id": "366f855c-745d-4820-9b10-975dac91d5f7"
         }
```

Tip: `wml_service_path`, `user` and `wml_password` can be found on **Service Credentials** tab of service instance created in Bluemix. If you cannot : in **Service Credentials** generate new credentials by pressing **New credential (+)** button.

Model Training & Deployment

using the Model Building Wizard



IBM Watson

ProjectsToolsCommunityServices

US South

My Projects / Project IBM i / Product Line Prediction

Select Data

Train

Evaluate

Select a technique

Column value to predict (Label Col)

PRODUCT_LINE (String)

Feature columns

GENDER (String), AGE (Integer), MARITAL_STATUS (String), PROFESSION (String)

GENDER (String)

AGE (Integer)

MARITAL_STATUS (String)

PROFESSION (String)

categories.

number of categories.

values.

Validation Split

Train: 60

Test: 20

Holdout: 20

Select the features
(input columns)
And Label (column to
predict)

Add Estimators

Configured estimators

Select the
Classification/Regression
algorithms (Estimators)

Data Split (Training – Test)

Model Training & Evaluation



ProjectsToolsCommunityServices

My Projects / Project IBM i / Product Line Prediction Refined

Select Data

Train

Evaluate

Select model

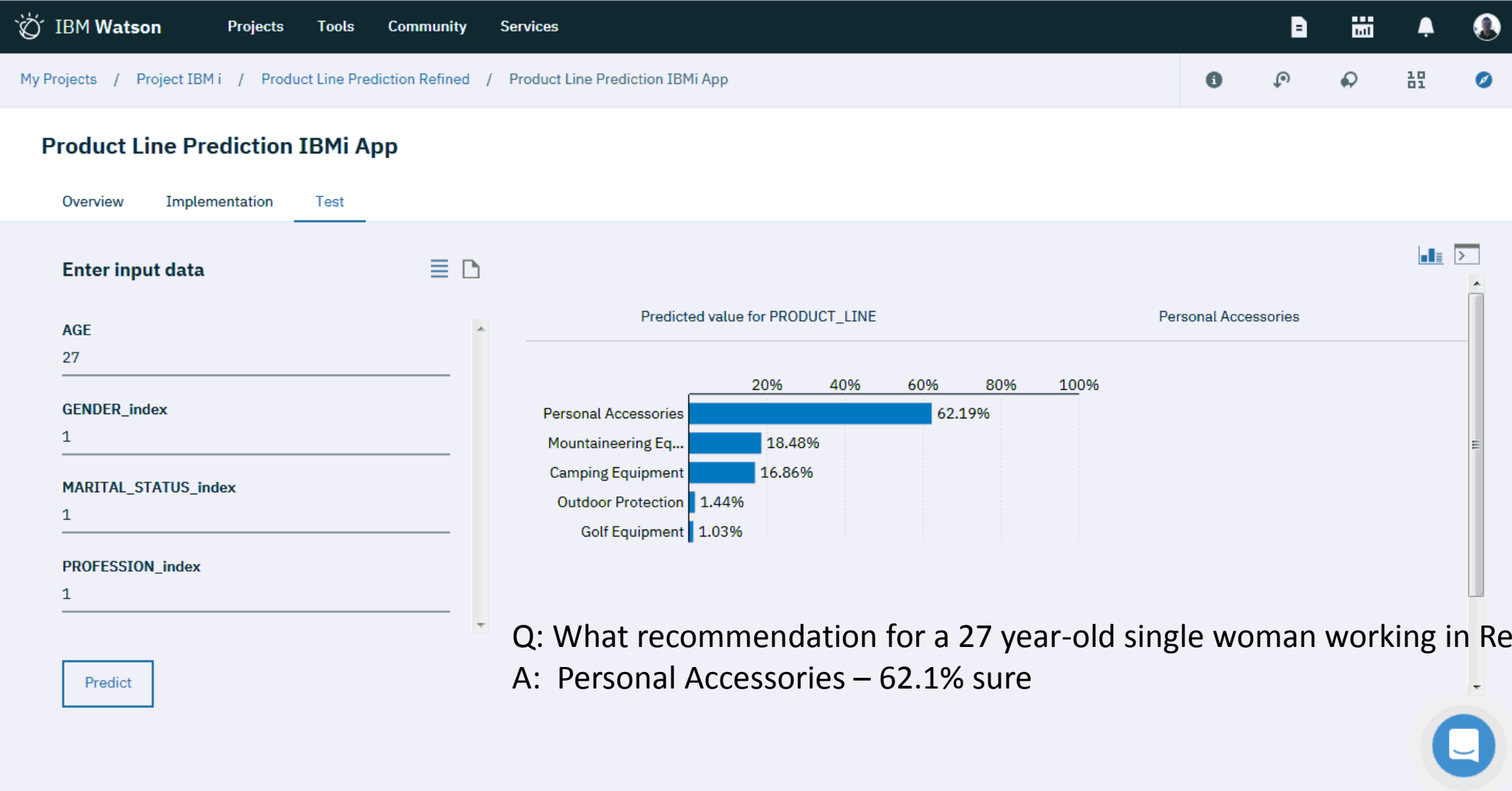
	ESTIMATOR TYPE	STATUS	PERFORMANCE	WEIGHTED TRUE POSITIVE RATE	WEIGHTED FALSE POSITIVE RATE	WEIGHTED PRECISION	WEIGHTED F MEASURE	WEIGHTED RECALL	LAST EVALUATION	ACTIONS
	RandomForestClassifier	Trained & Evaluated	Poor	0.5838	0.22583	0.56539	0.54983	0.5838	9 Apr 2018, 1:43 PM	
	DecisionTreeClassifier	Trained & Evaluated	Poor	0.56219	0.23524	0.52019	0.5128	0.56219	9 Apr 2018, 1:43 PM	

Close

Previous

Save

Model deployment, test, publication



Model deployment, test, publication



IBM Watson

ProjectsToolsCommunityServices

My Projects / Project IBM i / Product Line Prediction Refined / Product Line Prediction IBMi App

Product Line Prediction IBMi App

OverviewImplementationTest

Enter input data

AGE

77

GENDER_index

0

MARITAL_STATUS_index

0

PROFESSION_index

0

Predict

Predicted value for PRODUCT_LINE

Golf Equipment

20%40%60%80%100%

Golf Equipment67.43%

Personal Accessories15.09%

Camping Equipment12.98%

Mountaineering Eq...3.73%

Outdoor Protection0.76%

Q: What recommendation for a 77 year-old married & retired man ?

A: Golf Equipment – 67.43% sure

Model deployment, test, publication

API Specs (Swagger)



IBM Watson Machine Learning API

Service Instance URL or ID

Authorize

Authorization

Step by step instruction how to use Watson Machine Learning service can be found [here](#)

IBM Watson Machine Learning Credentials

To start working with API one needs to generate an `access token` using the `username` and `password` available on the Service Credentials tab of the IBM Watson Machine Learning service instance or also available in the VCAP environment variable.

Example of the Service Credentials:

```
{
  "url": "https://ibm-watson-ml.mybluemix.net",
  "access_key": "ERY9vcBfE4sE+F4g8hcotF9L+j81wXWeZv",
  "username": "c1ef4b80-2ee2-458e-ab92-e9ca97ec657d",
  "password": "030528d4-5a3e-4d4c-9258-5d553513be6f",
  "instance_id": "a751c209-954e-dc32-b441-ad56ce7a9f40"
}
```

Example of obtaining `access token` from Token Endpoint using HTTP Basic Auth (for details please refer to Token section below):

```
curl --basic --user username:password https://ibm-watson-ml.mybluemix.net/v3/identity/token
```

The obtained `access token` needs to be prepended with `Bearer` word and it needs to be passed in the `Authorization` header for API calls.

Example of API request with `Bearer access token` :

```
curl https://ibm-watson-ml.mybluemix.net/v3/wml_instances/00fd89e6-8cf2-4712-a068-ade10277b649/published_models -H "Authorization: Bearer eyJhbGciOiJSUzUxMiIsInR5cCI6IkpXVCJ9.eyJ0ZW50bnR3ZCI6ImU4YmQzZGM3LWI5Y2UtNDY1OC1iZ..."
```

Apache Spark Service Credentials

The IBM Watson Machine Learning co-operates with the Apache Spark as a Service to create `batch`, `stream` deployments and for `learning configuration` functionality.

For API methods requiring Apache Spark Service instance a custom header: `X-Spark-Service-Instance` with Service Credentials must be specified. The header value is a base64 encoded string with the JSON data containing Service Credentials and Spark version.

Model Integration with IBM i Apps



☰ IBM Cloud

Catalog Docs Support Manage

Manage

Service credentials

Plan

Connections

WatsonML

Location: US South Org: benoit.marolleau@fr.ibm.com Space: DataSciX

Credentials are provided in JSON format. The JSON snippet lists credentials, such as the API key and secret, as well as connection information for the service. [View More](#)

Service credentials

New credential +

10 Items per page | 1-1 of 1 items

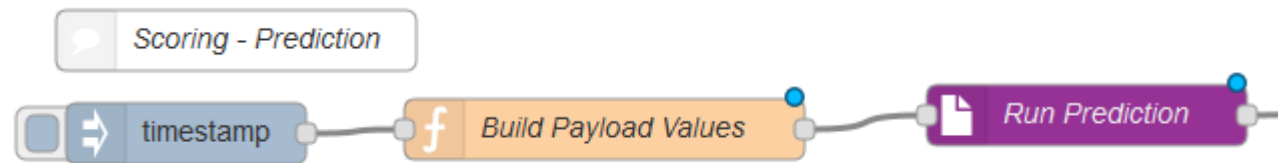
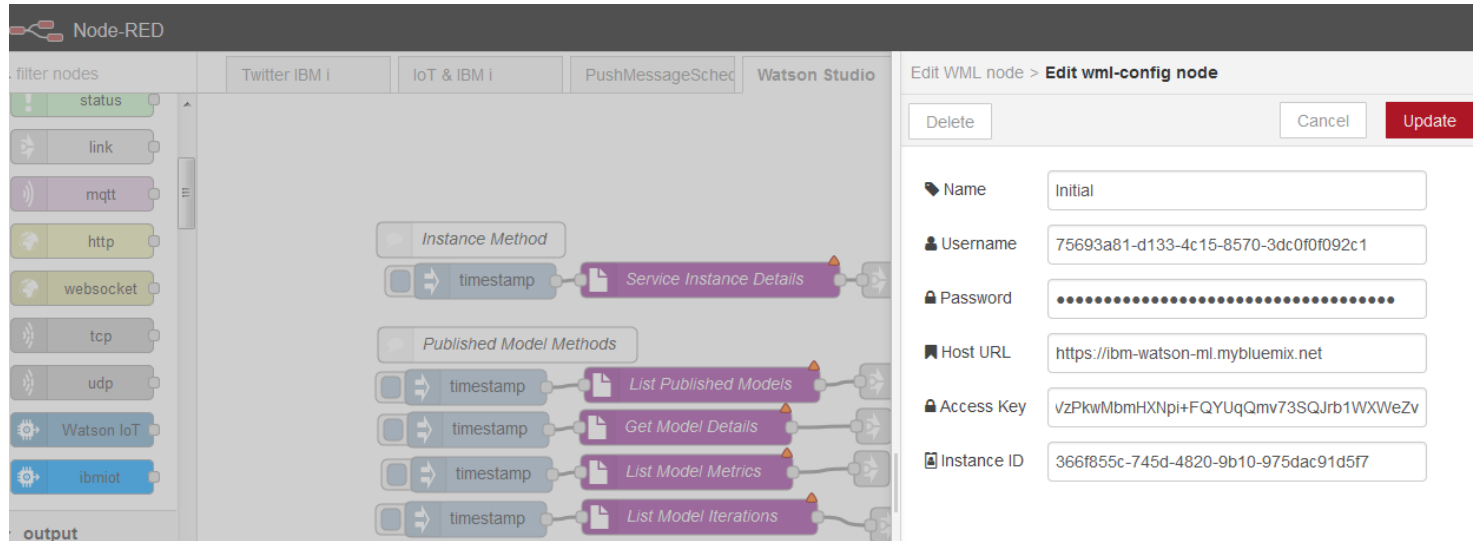
1 of 1 pages < 1 >

☒ KEY NAME	DATE CREATED	ACTIONS
☒ apsx-data	Sep 13, 2017 - 02:40:49	View credentials ▲

```
{
  "url": "https://ibm-watson-ml.mybluemix.net",
  "access_key": "mN3bwIsi/xp0h08Bd8D8Zah6UF++y/ZqlqTlJDMJQr6dqC83AFnAyPemGJw9js0JHxGxQ3pIogjgEOjN0TGDTcL0h32gVzPkwMbmHXNpi+FQYUqQmv73SQJrb1WXWeZv",
  "username": "75693a81-d133-4c15-8570-3dc0f0f092c1",
  "password": "13cb03fd-2b6b-4b24-a17f-49eb2efaf7b9",
  "instance_id": "366f855c-745d-4820-9b10-975dac91d5f7"
}
```

Model Integration with IBM i Apps

Example with Node.js on IBM i : Node-RED Prototype



```
11/05/2018 à 18:41:11 node: d0861edf.f95ad8
msg: Object
  object
    _msgid: "4b7c6d15.f5aae4"
    topic: ""
    payload: object
      fields: array[10]
      values: array[1]
        0: array[10]
          0: 27
          1: 0
          2: 1
          3: 1
          4: array[4]
          5: array[5]
          6: array[5]
          7: 0
          8: "Camping Equipment"
          9: array[5]
      _event: "node:3d4e884c.9f8908"
```

Fig. Graphical coding: let's invoke our model from Node-RED
Left: json input values + Watson ML node for invocation.
Right: API call result in json

Model Integration with IBM i Apps

Example with Node.js on IBM i



```
benoit@10.3.54.77's password:
```

```
##### # # ##### #####
# # # # # # # # # #
# # # # # # # # # #
# ##### # # # # # #
# # # # # # # # # #
# # # # # # # # # #
### ##### # # # # #
```

```
-----*Welcome***DEMO LPAR-IBM Client Center Montpellier-----
```

```
bash-4.4$ node predictProductLine.js 77 0 1 1
```

```
IBM i ** Sending IBM Watson Machine Learning request to Model : PredictProductLineIBMi - IBM Cloud
```

```
{ fields:
```

```
  [ 'AGE',
    'GENDER_index',
    'MARITAL_STATUS_index',
    'PROFESSION_index',
    'features',
    'rawPrediction',
    'probability',
    'prediction',
    'nodeADP_class',
    'nodeADP_classes' ],
```

```
values:
```

```
  [ [ 77,
      0,
      1,
      1,
      [Object],
      [Object],
      [Object],
      3,
```

```
      'Golf Equipment',
```

```
      [Object] ] ] }
```

Product Line purchase prediction for a customer
Ex: 77 years old, Male, Married

Input data from Db2 for i or any programs



How to get started? Q&A

OSDB @ Montpellier Cognitive Systems Lab

Positioning IBM Power Systems at the heart of the Cognitive Era

Power System Linux Center

Engage with Clients, ISVs
& Partners to leverage
the benefits of Linux on
Power, using **Open
Sources** Solutions



Power Acceleration for High Performance

HPC & HPDA
From traditional HPC on
POWER/GPU technologies
to FPGA based
Acceleration with
CAPI/SNAP



Power Acceleration for ISV's

Leverage OpenSource
Databases in the
Competition with Oracle.

Competitive & TCO Eagle Teams

Leverage TCO Eagle Team
Studies & Competitive
Expertise (X86/Oracle)

Software Defined Infrastructure

Leverage Software Define
Infrastructure to expend
Modern Data Platform
Capabilities (Spectrum
Scale, ESS, LSF...)

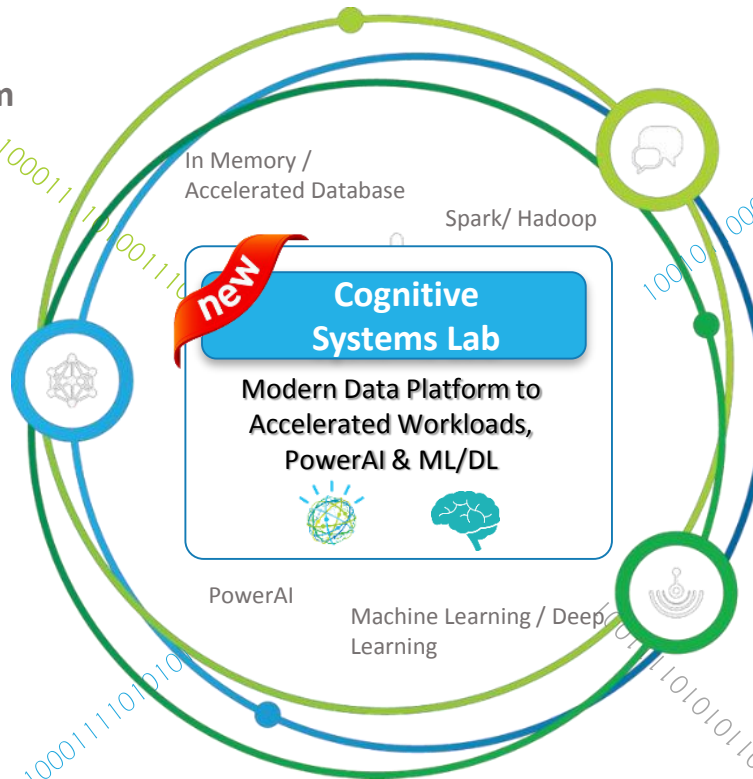


Modern Data Platform

10001111010011101

HPC GPU/NVlink & CAPI Acceleration

10001111010100110011



Competitive Discussions

1001011000111101

Hybrid Cognitive Solutions

01101101010111

What we will deliver

Highly Technical Skills & Hardware (S822LC for HPC & Big Data) available

Talk &
Demonstrate

Exploration & Design
Workshops

PoT/ POC/Benchmarks

Advanced
Technical Support

IBM Client Center Montpellier

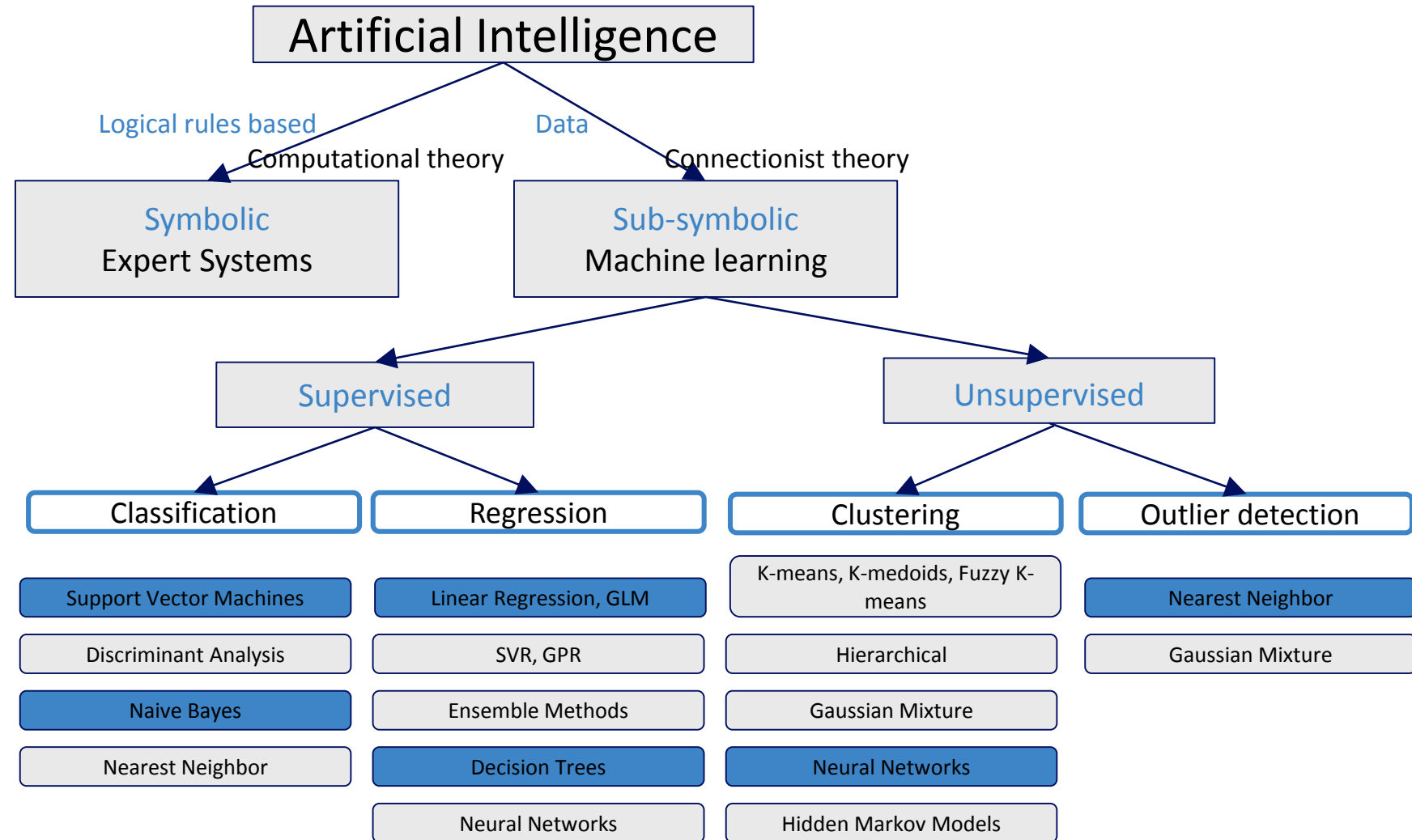


IBM Client Center Montpellier



Backup Slides

The Machine Learning Tasks & algorithms



Supervised learning

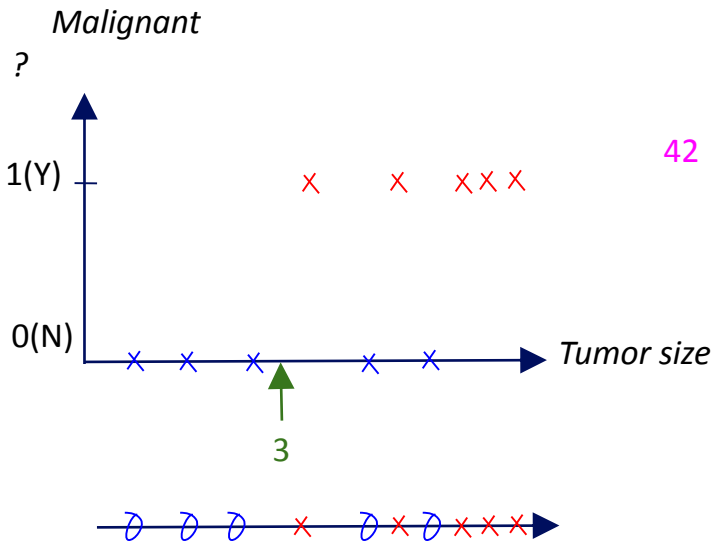
“right answers” given

Classification

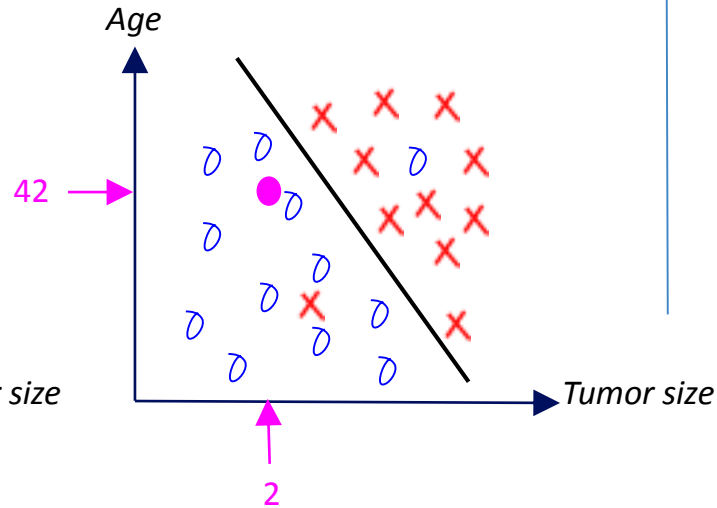
- The output variable takes class labels (discrete valued output)

Breast Cancer (malignant, benign)

1 variable



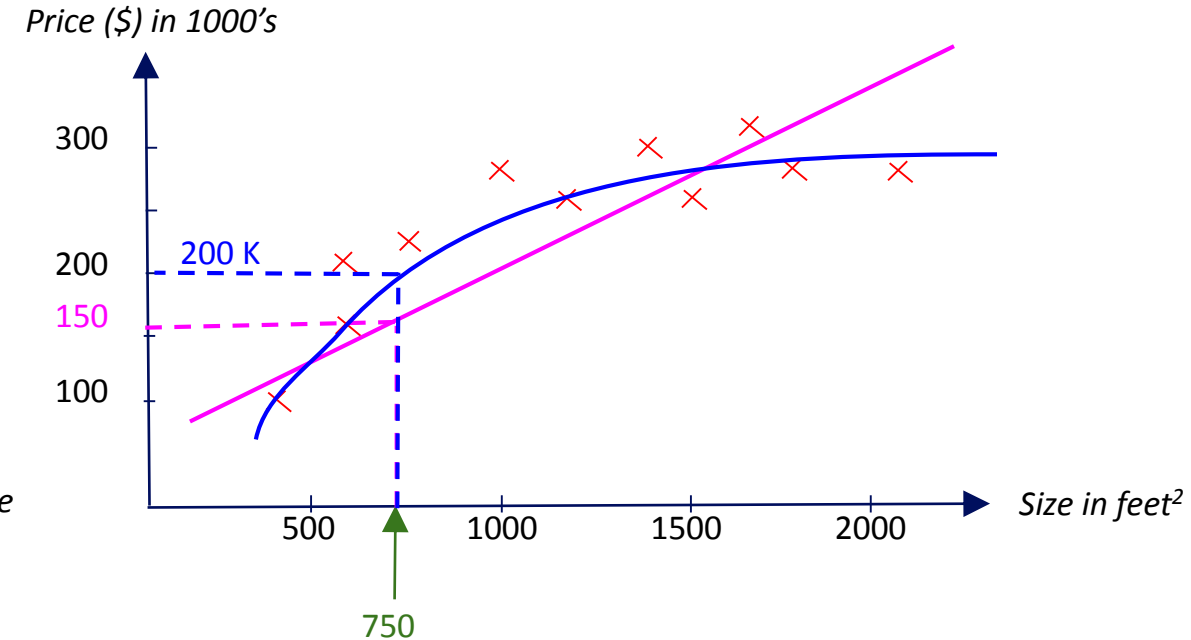
2 variables



Regression

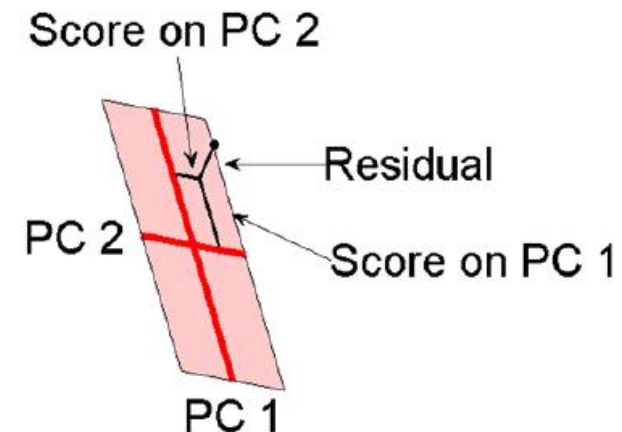
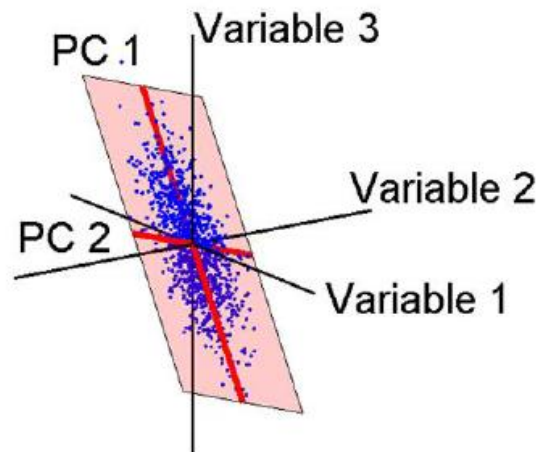
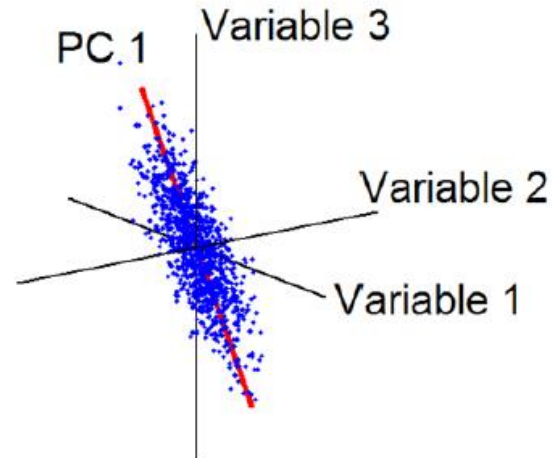
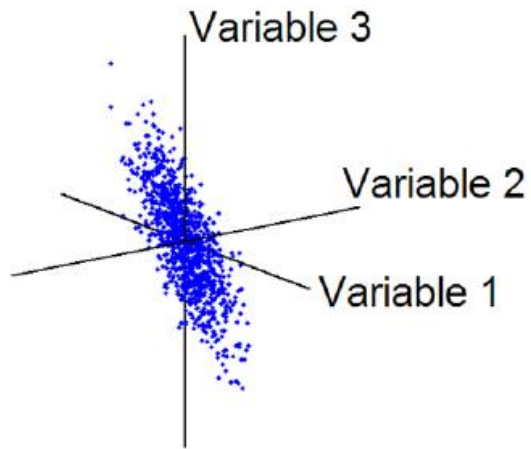
- Predict continuous valued output.

Housing price prediction



PCA for dimensionality reduction

Principal Components Analysis

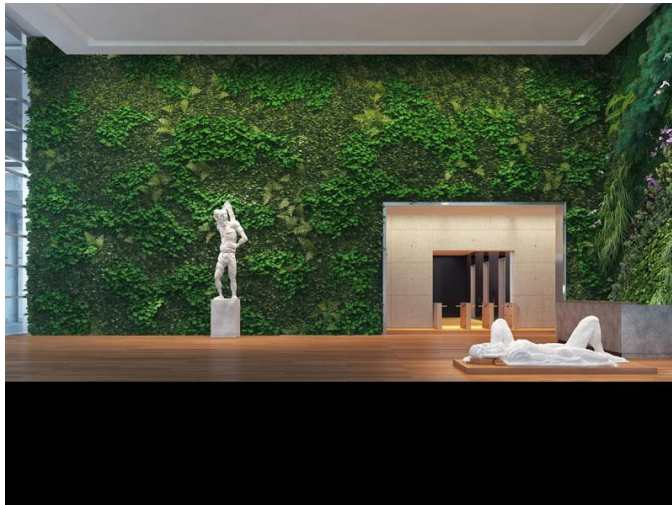


MATLAB EXPO 2017

Spark Technology Center

IBM established Spark Technology Center to contribute to the Apache® Spark™ ecosystem – June 2015

505 Howard Street, San Francisco



IBM Spark Technology Center (STC)
San Francisco, USA

Growing pool of contributors
~50 world wide, and 3 committers

Apache SystemML now an official Apache Incubator project

Founding member of AMPLab (and upcoming RISE Lab)

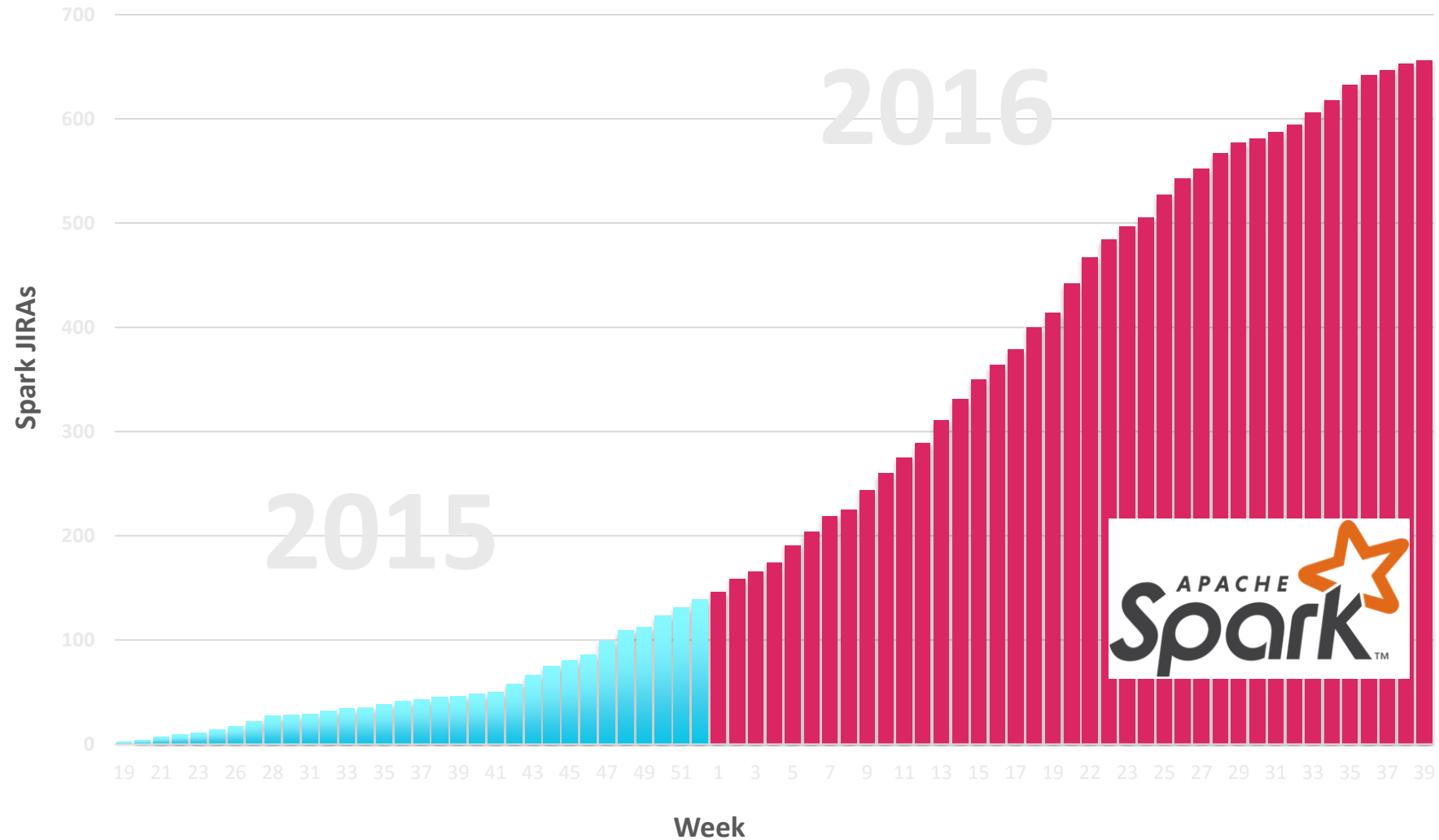
Member of R Consortium

Founding member of Scala Center

Partnerships in the ecosystem

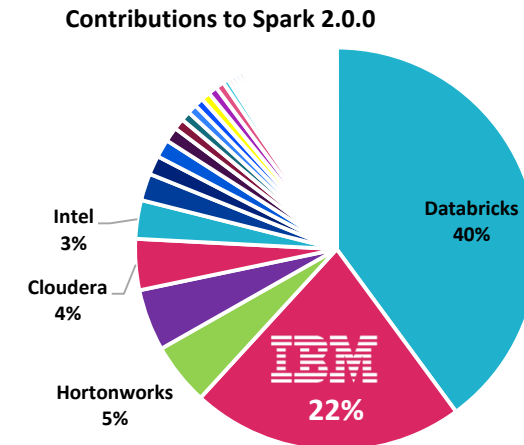
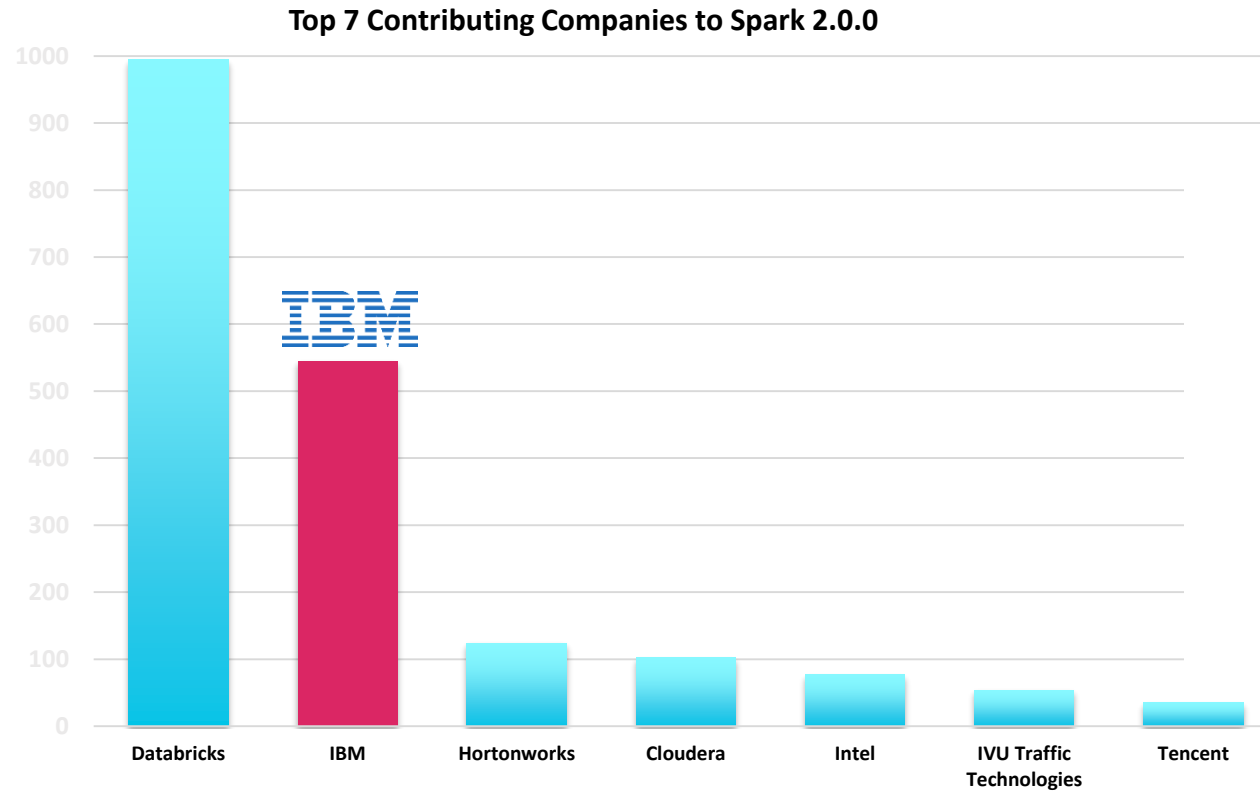
Spark Technology Center contributions have grown over 400% since start in June 2015

STC Spark Contribution Progress



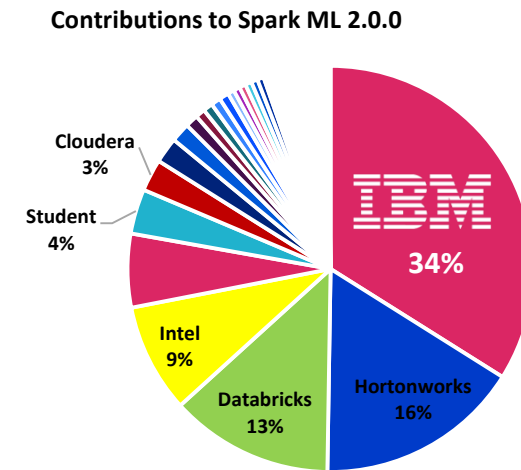
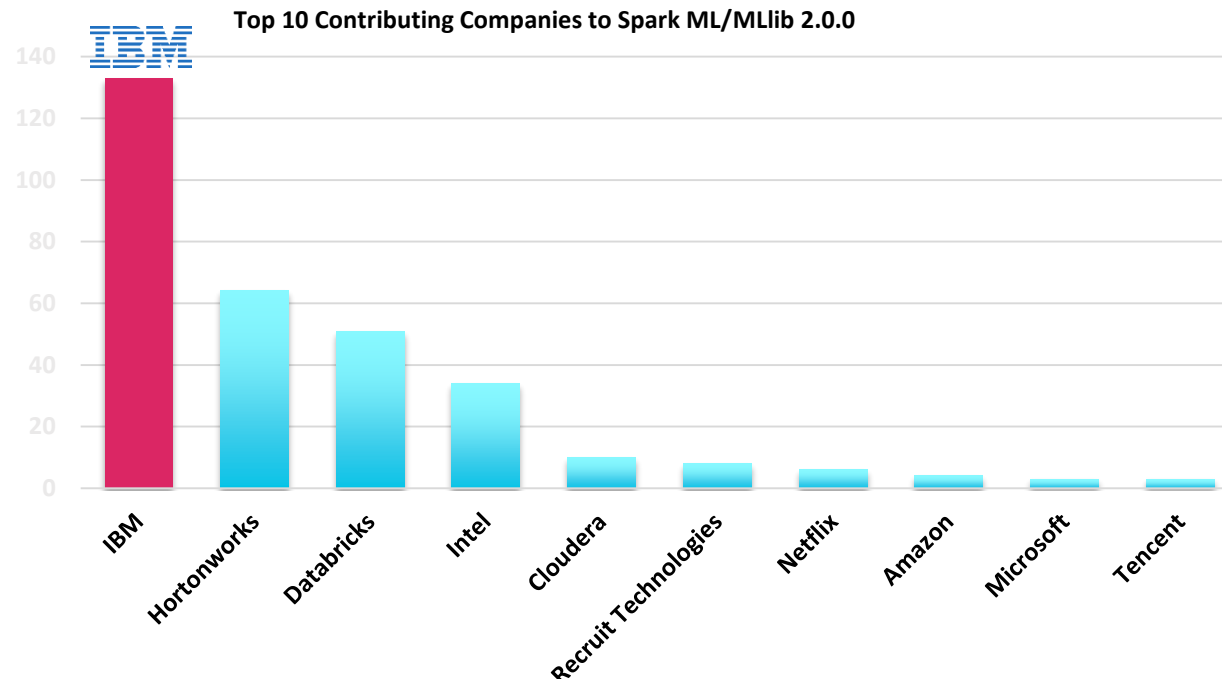
IBM had a significant impact on Spark 2.0

- IBM is **#2 contributor** to Apache Spark
- IBM was the leading contributor in Spark 2.0 to SparkML, PySpark, and SparkR



IBM impact on SparkML / MLlib 2.0

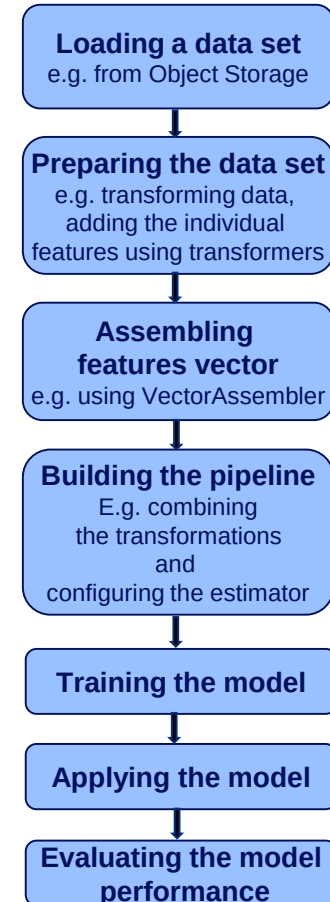
- Spark Machine Learning (ML) provides a toolset to create pipelines of different ML related transformations on your data
- IBM is **#1 contributor** in the Spark (ML)
- Distinction between ML and MLlib:
 - MLlib is based on RDDs; ML is based on data frames.
 - The distinction between both is fading out. In general they usually combine both under the name "Spark ML"



Machine Learning framework in Apache Spark

Pipeline components:

- **Transformers (e.g. indexing, normalization):**
Dataframe -> Dataframe with features
- **Estimators:**
Dataframe -> ML model
- **Models:**
Dataframe -> Dataframe with predictions
- **Pipelines:**
Dataframe -> (chained transformers and estimators) -> ML model
- **Evaluators:**
Dataframe -> ML model



IBM Watson Studio & Machine Learning Pricing



- Watson Studio
 - Per user licensing + Processing Units
 - <https://www.ibm.com/cloud/watson-studio/pricing>
- Watson Machine Learning
 - Per prediction licensing + Processing Units
 - <https://www.ibm.com/cloud/machine-learning/pricing>
- IBM Data Science Experience Local (includes IBM Machine Learning*)
 - Per user licensing

*IBM Machine Learning = on premises Watson Machine Learning