Université IBM i 2018

16 et 17 mai

IBM Client Center Paris

Session S12
IBM i & Data Science: Introduction à 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   MarolleauBenoit

IBM Client Center Montpellier

Session S12

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

Université IBM i 2018 - 16 et 17 mai

IBM Client Center Paris

Benoit MAROLLEAU - Cloud Architect
IBM Client Center Montpellier, France

@ : Benoit.marolleau@fr.ibm.com
linkedin.com/in/benoitmarolleau
MarolleauBenoit

for Business

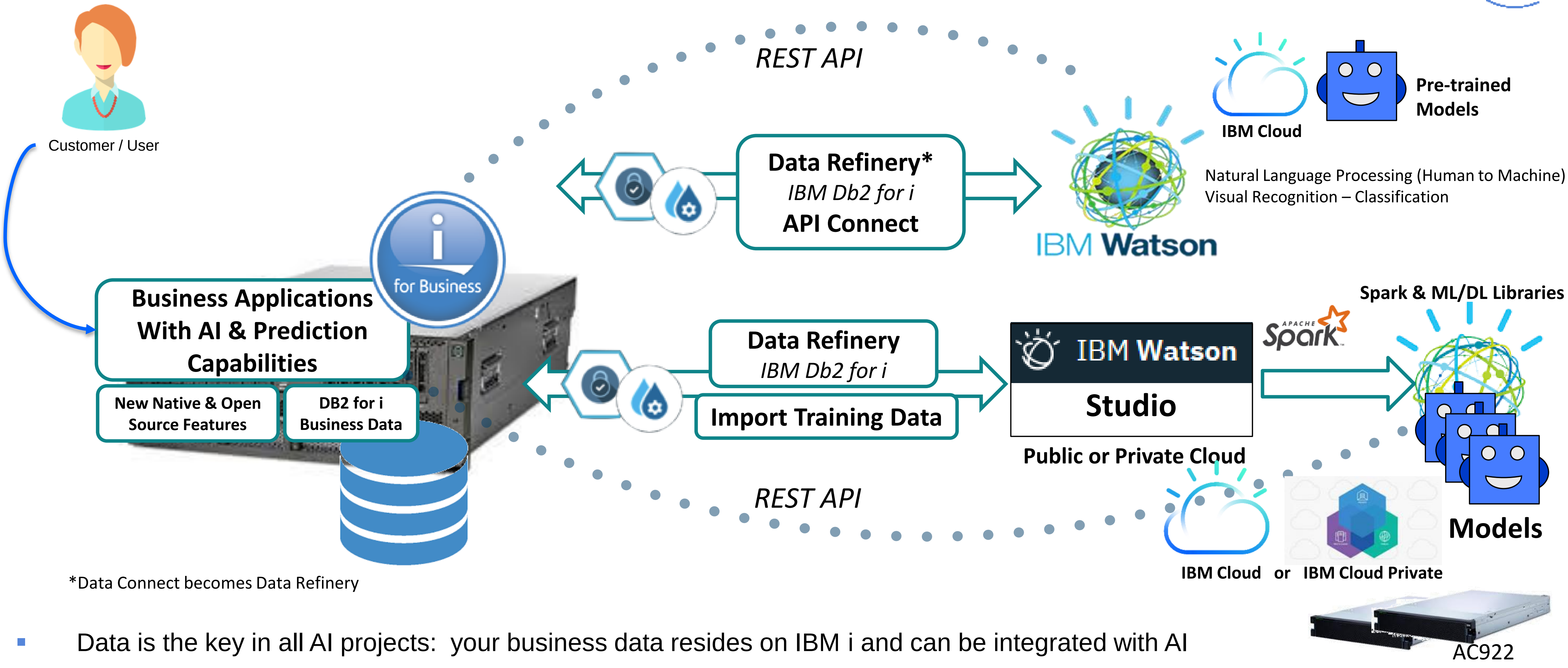
IBM Systems

Plan de la présentation

- Introduction: Pourquoi l’intelligence artificielle? Quelle bénéfices?
- Solutions AI d’entreprise: IBM Watson, Watson Studio, Data Science Experience
- AI & IBM i
  - Exemple d’intégration dans un environnement IBM i avec démonstration
- Comment démarrer sur un projet AI - Questions / Réponses

IBM i & Artificial Intelligence

Approximate (AI) & precise (Transactional) computing together



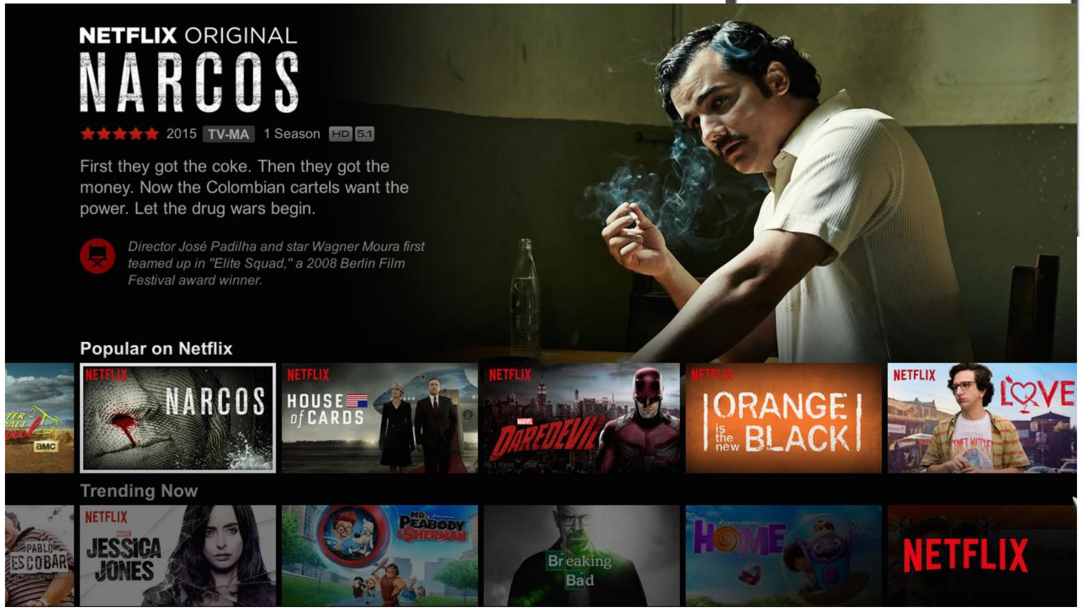
*Data Connect becomes Data Refinery

- 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 specifics 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

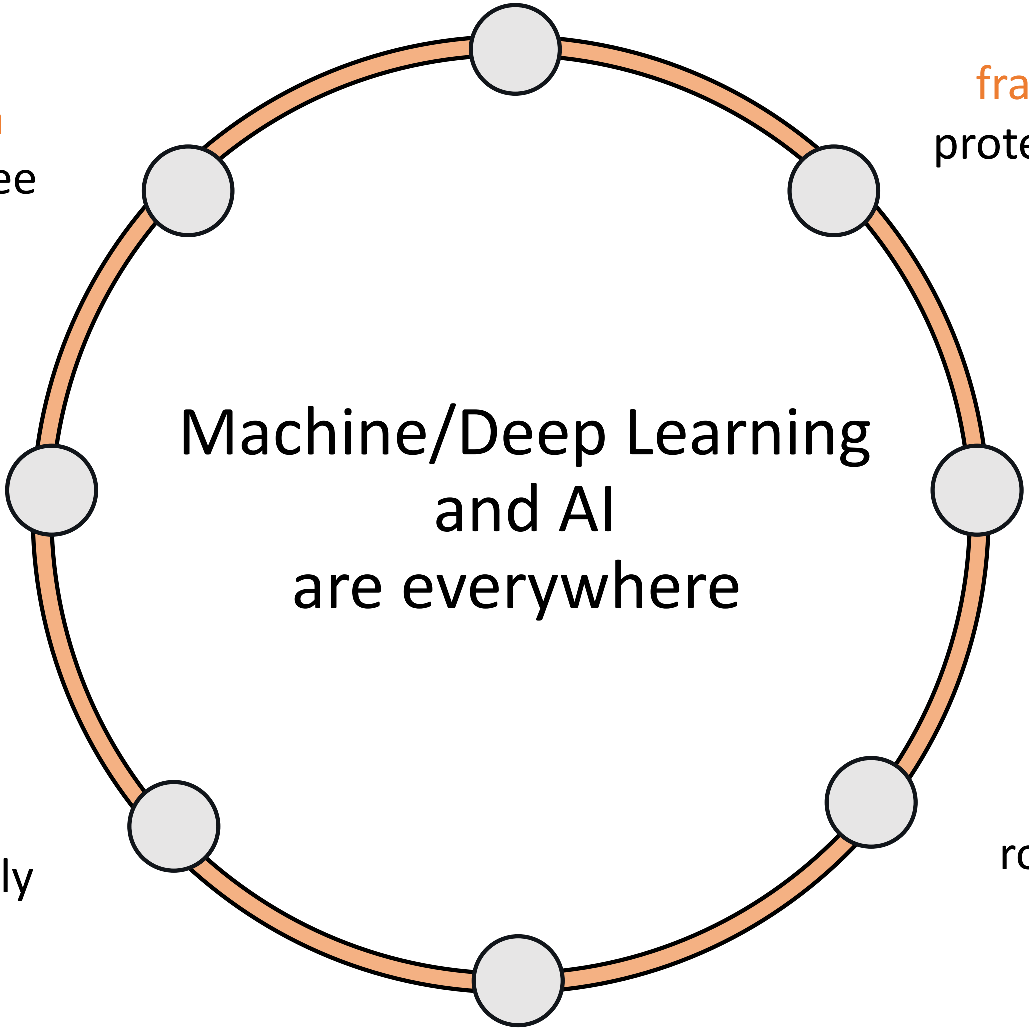
The future is now

facial recognition unlocks your phone

speech recognition lets you go hands-free

fraud detection protects your credit

autonomous vehicles detect pedestrians



recommendations help you shop faster

machine vision detects cancer early

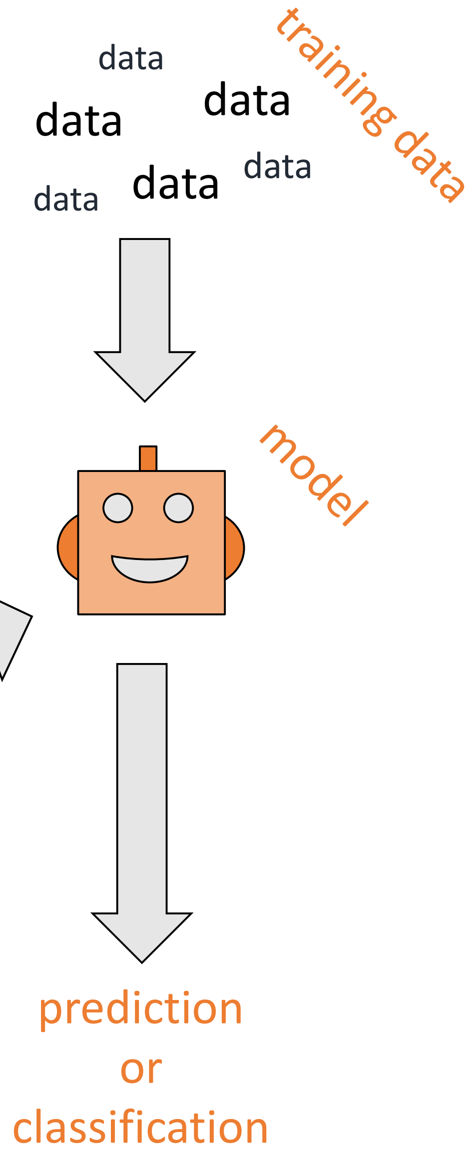
chat bots route calls quicker

spam detection unclogs your Inbox

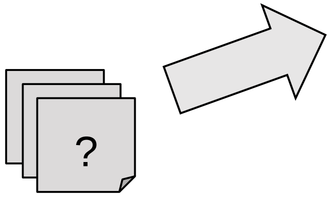
How does machine learning work?

Machine learning requires TONS OF DATA

1. A machine learning model is trained to recognize patterns in historical data

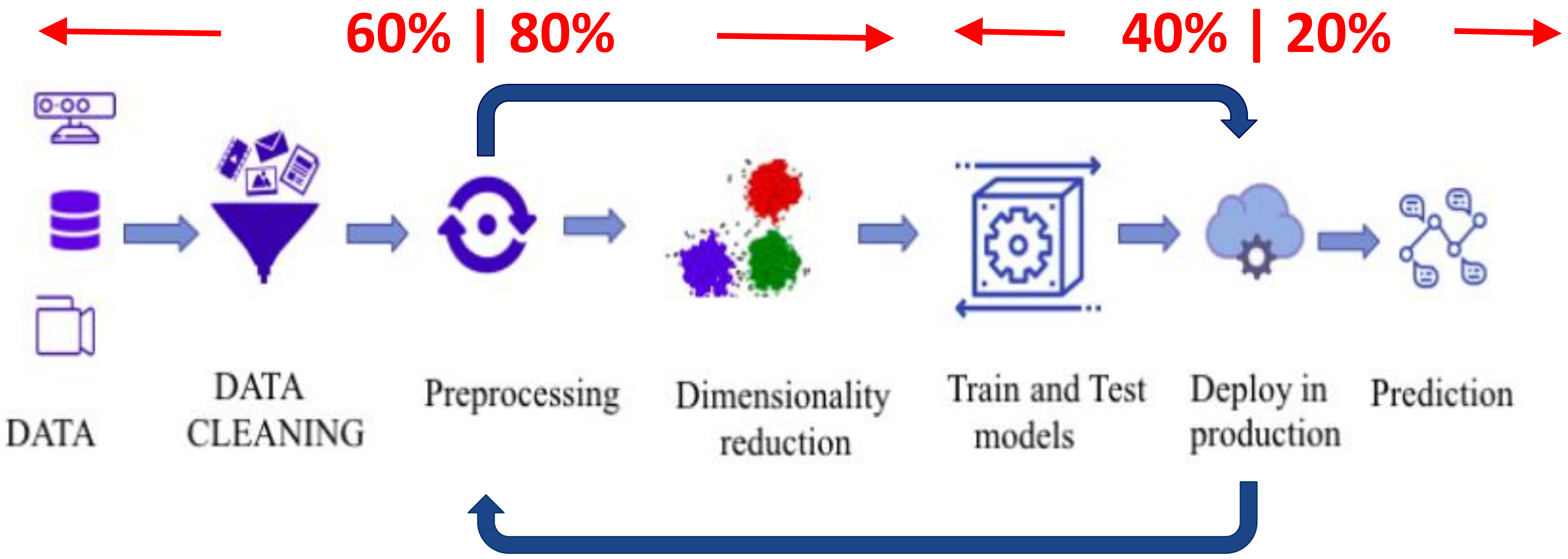


2. The model is then shown new data and asked to predict or classify it.



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

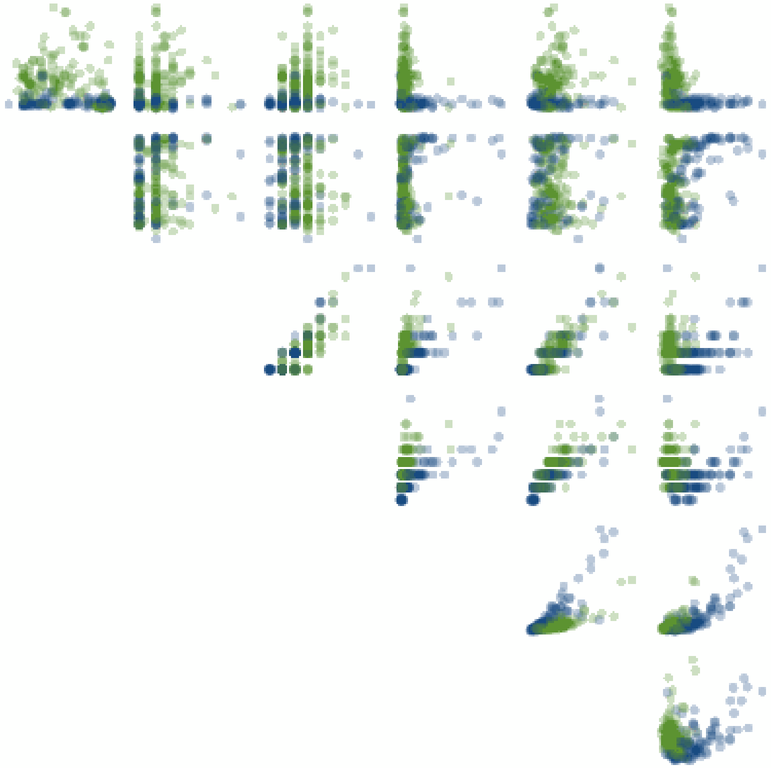
What is Machine Learning?

R2
D3

A Visual Introduction to Machine Learning

In machine learning, computers apply **statistical learning** techniques to automatically identify patterns in data. These techniques can be used to make highly accurate predictions.

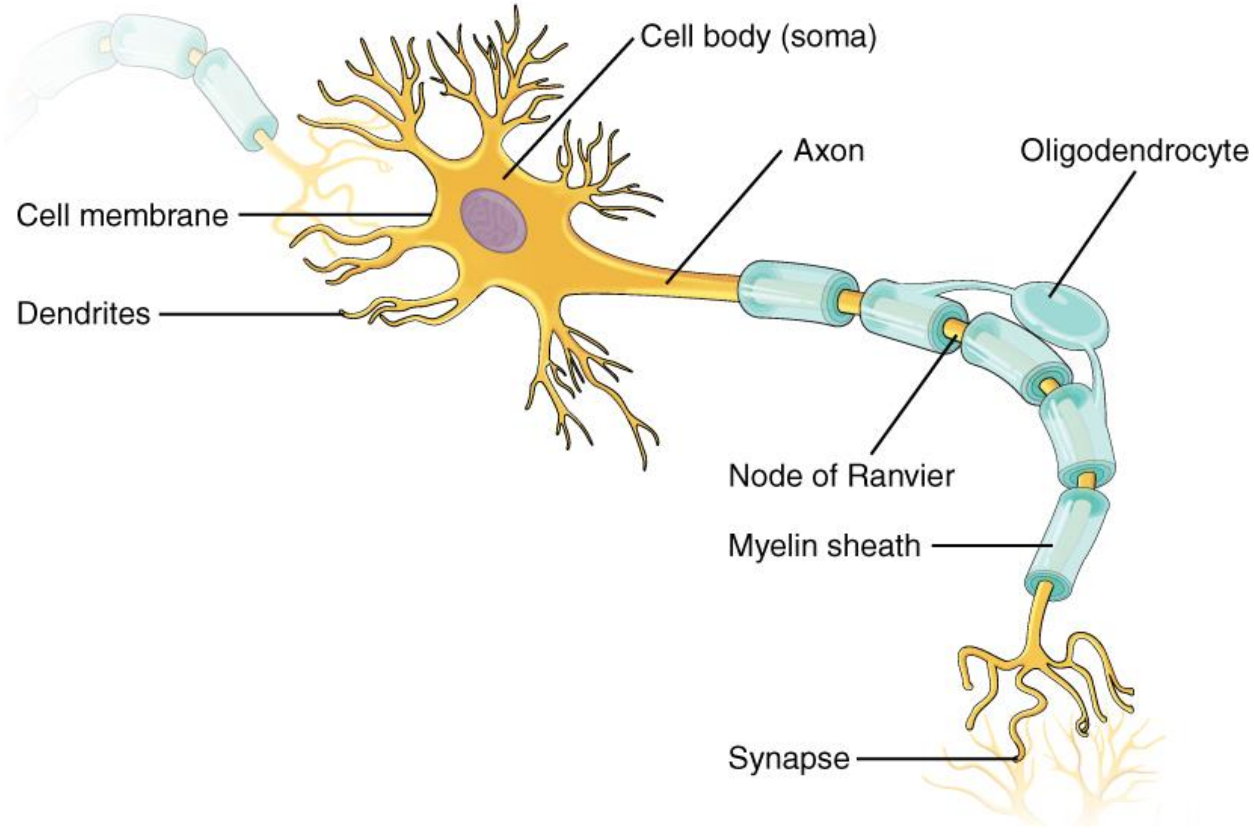
Keep scrolling. Using a data set about homes, we will create a machine learning model to distinguish homes in New York from homes in San Francisco.



Read the speaker notes, Appendix, and check out
"A Visual Introduction to Machine Learning" **– http://bit.ly/1LRTlSi**

Deep Learning = Training Artificial Neural Networks

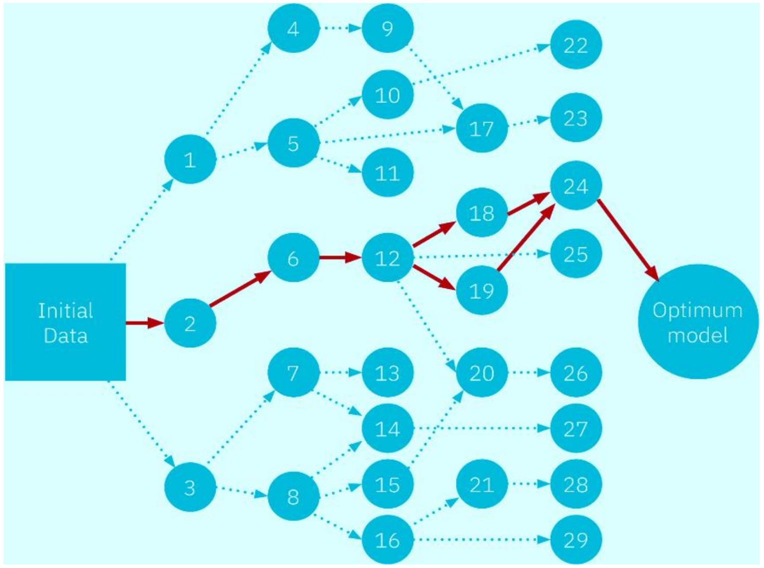
Based on biological neurons. Artificial neurons learn by recognizing patterns in data.



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 *predicted most likely to buy* 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

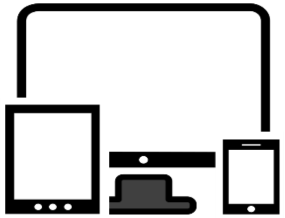
Machine learning

*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

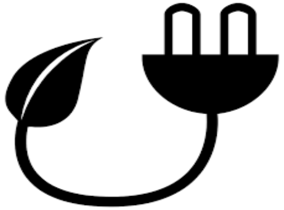
Retail
Market Basket Analysis, Next Best Offer, Customer Churn, propensity to buy



Marketing
Discount targeting, email optimization, lifetime client value

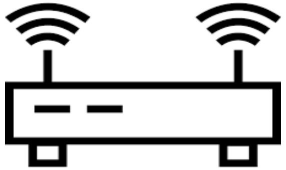
Healthcare
Medicare fraud, AI-assisted diagnosis, drug demand forecast

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

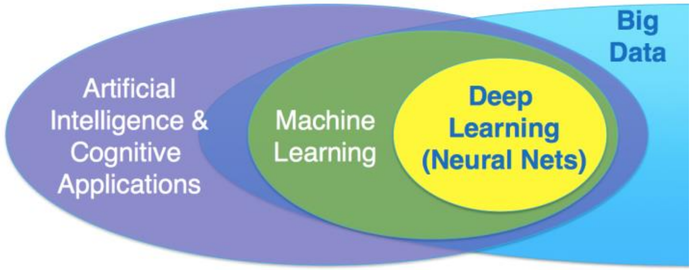


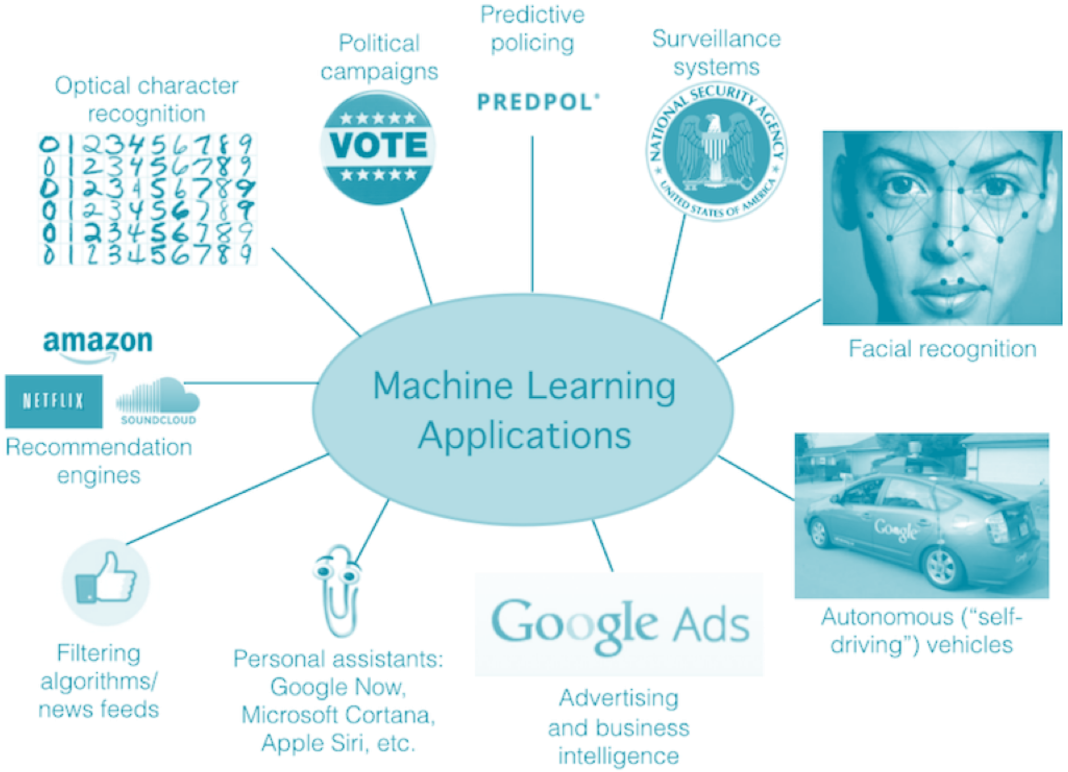
Security
Malicious activity detection, logs analysis

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

Big Data: Machine Learning techniques

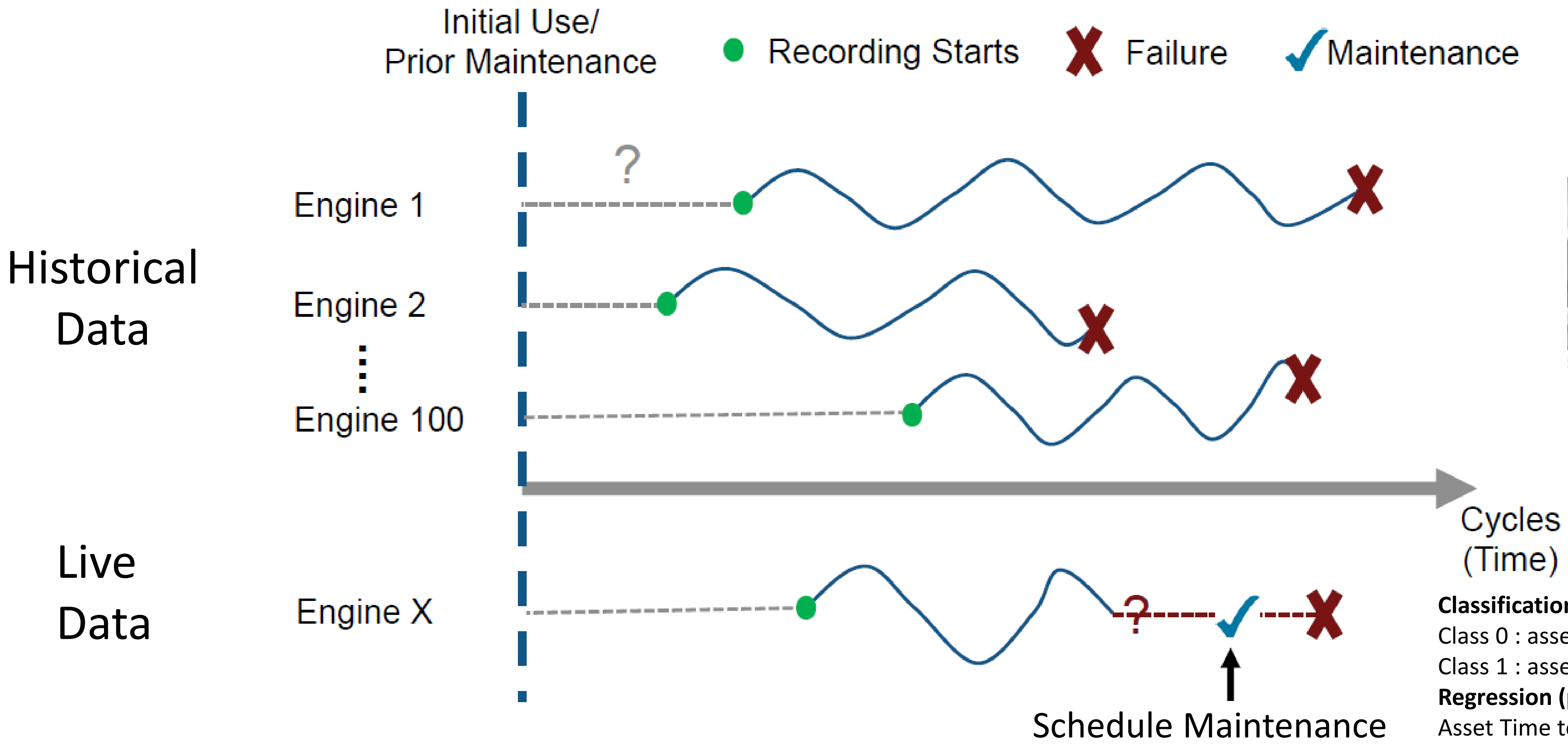
- Classification: predict class from observations
  - - E.g. Spam Email Detection
- Clustering: group observations into “meaningful” groups
  - - E.g. Amazon Recommendations
- Regression (prediction): predict value from observations
  - - E.g. Energy consumption





and many different technologies and libraries are available:

Example: Predictive Maintenance



Classification:
Class 0 : asset will fail within the next 15 cycles
Class 1 : asset won‘t fail within the next 15 cycles
Regression (prediction):
Asset Time to failure = 18 cycles

DSX Demo available https://ibm.ent.box.com/v/power-iot-dsx-video-mp4
Historical Data from NASA https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/

Enterprises generate TONS OF DATA

Data that requires governance

Which must be cleaned and shaped for training

...then models must be designed

scikit learn

TensorFlow™

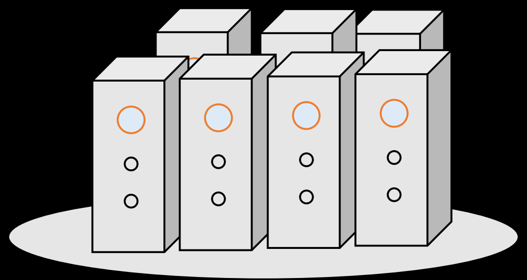
PYTORCH

SPSS

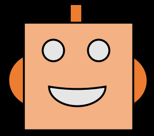
Keras

Caffe

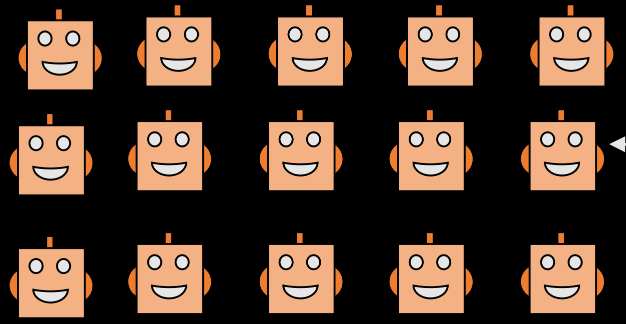
...and trained on high performance compute



To select an optimal model...



...that must be hosted and monitored



Why are enterprises struggling to capture the value of AI?

- Data
- Data resides in silos & difficult to access
- Unstructured and external data wasn't considered

- Governance
- If the data isn't secure, self-service isn't a reality
- Challenge understanding data lineage and getting to a system of truth

- Skills
- Data Science skills are in low supply and high demand
- Nurturing new data professionals is challenging

- Tools & Infrastructure
- Need an environment that enables a "fail fast" approach
- Discrete tools present barriers to productivity

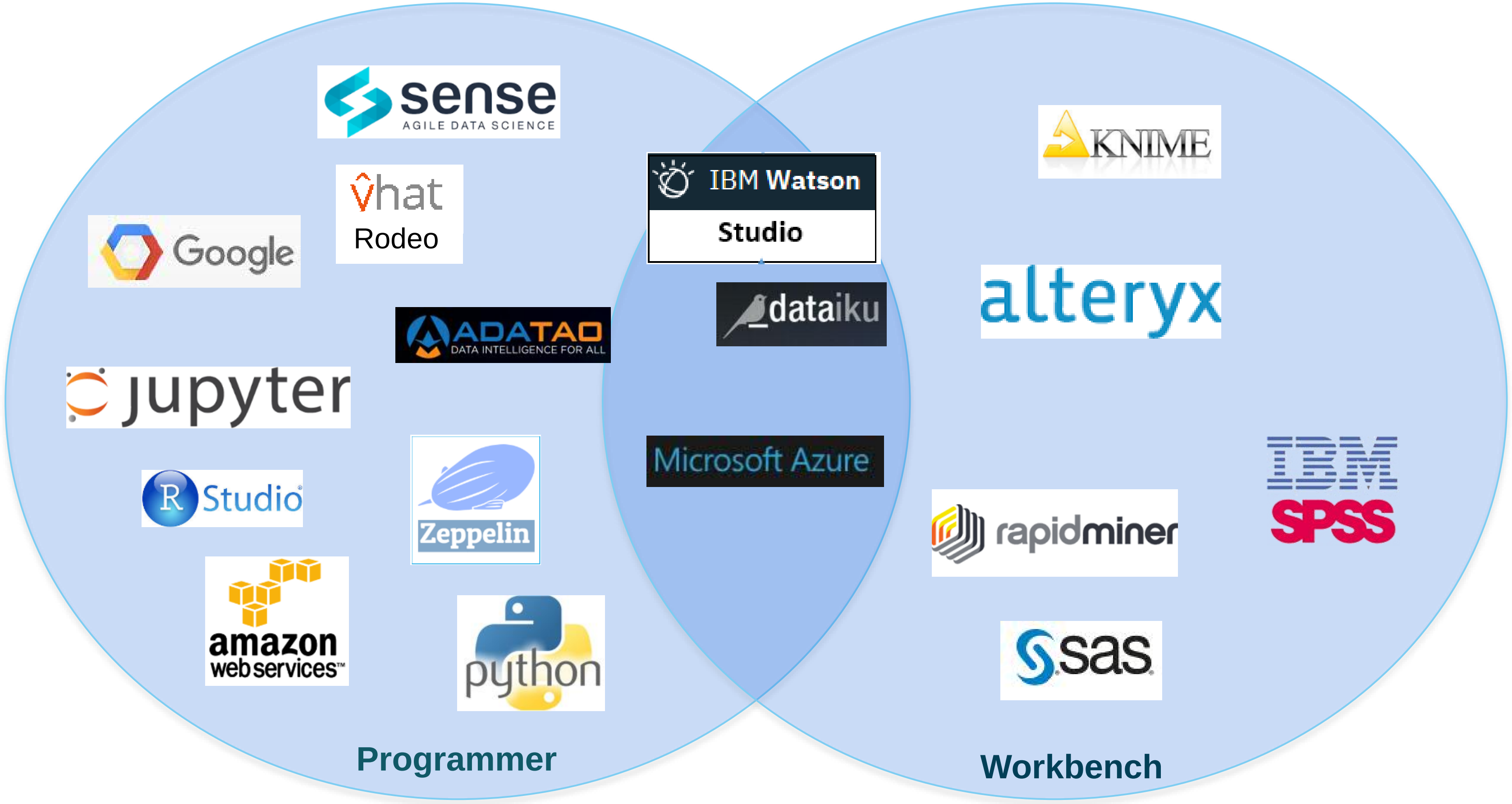
Artificial Intelligence: IBM Solutions

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

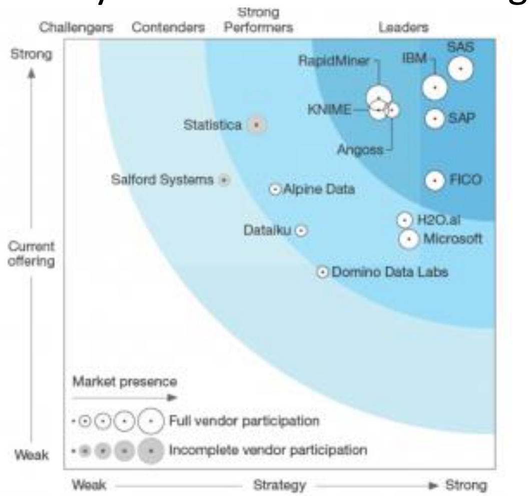


We’ve been recognized for our vision

Gartner Magic Quadrant 2017
Data Science Platforms

DeveloperWeek 2017
Devie

Forrester Wave 2017
Predictive Analytics & Machine Learning



Source: https://www.gartner.com/doc/reprints?id=1-3TKD8OH&ct=170215&st=sb
http://www.developerweek.com/awards/2017-devies-award-winners/

Watson: AI for Smarter Business

Watson Business Solutions			
Compliance Assist	Customer Care	Expert Assist	Voice of the Customer

Watson Applications		
Watson Assistant	Watson Cybersecurity	Compare & Comply

ISV & Third Party Applications		
…	…	…

Watson Studio

Watson APIs

Connect & Access Data › **Search & Find Relevant Data** › **Prepare Data (Ingest, Curate, & Enrich)** › **Build & Train AI Models** › **Deploy AI Models** › **Monitor, Analyze, Manage** ›

Continuous Learning

Watson Machine Learning and Deep Learning as a Service

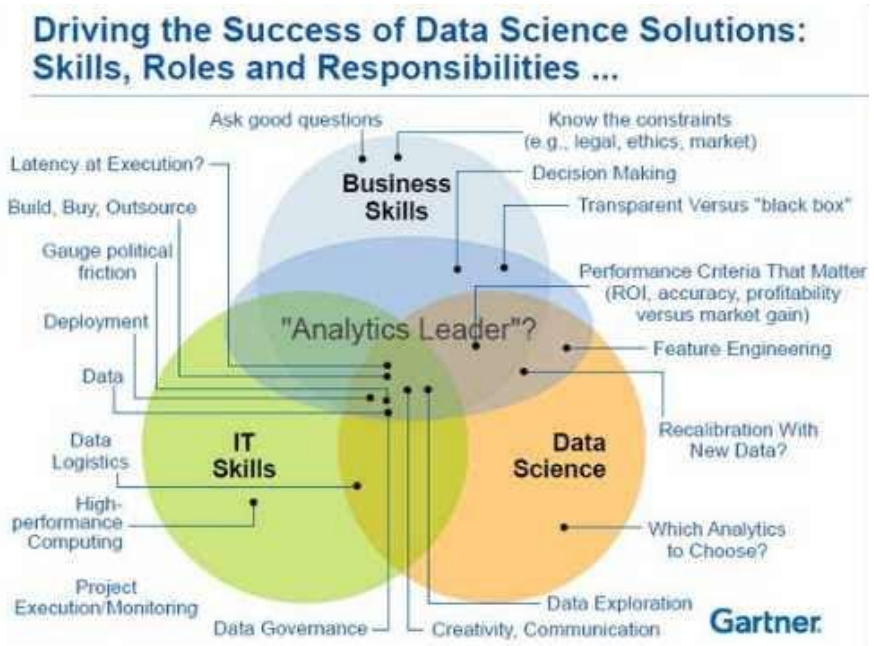
Watson Knowledge Catalog

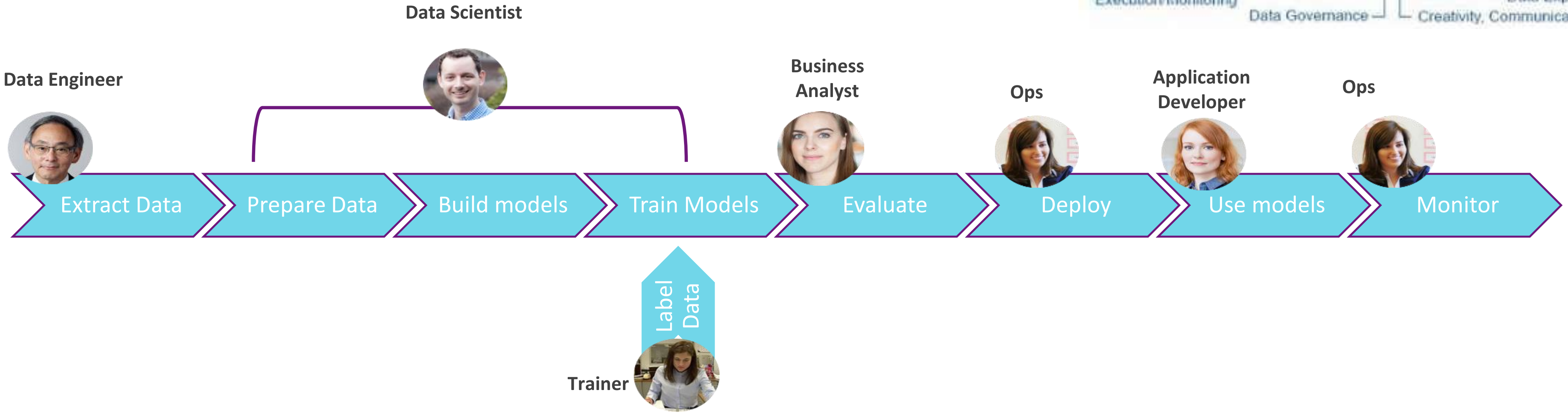
Watson Enriched Data & Analytical Assets	Watson Powered Search & Social Collaboration	Active Policy Enforcement	Model Governance, Traceability & Lineage	Data Kits

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)





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 a 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

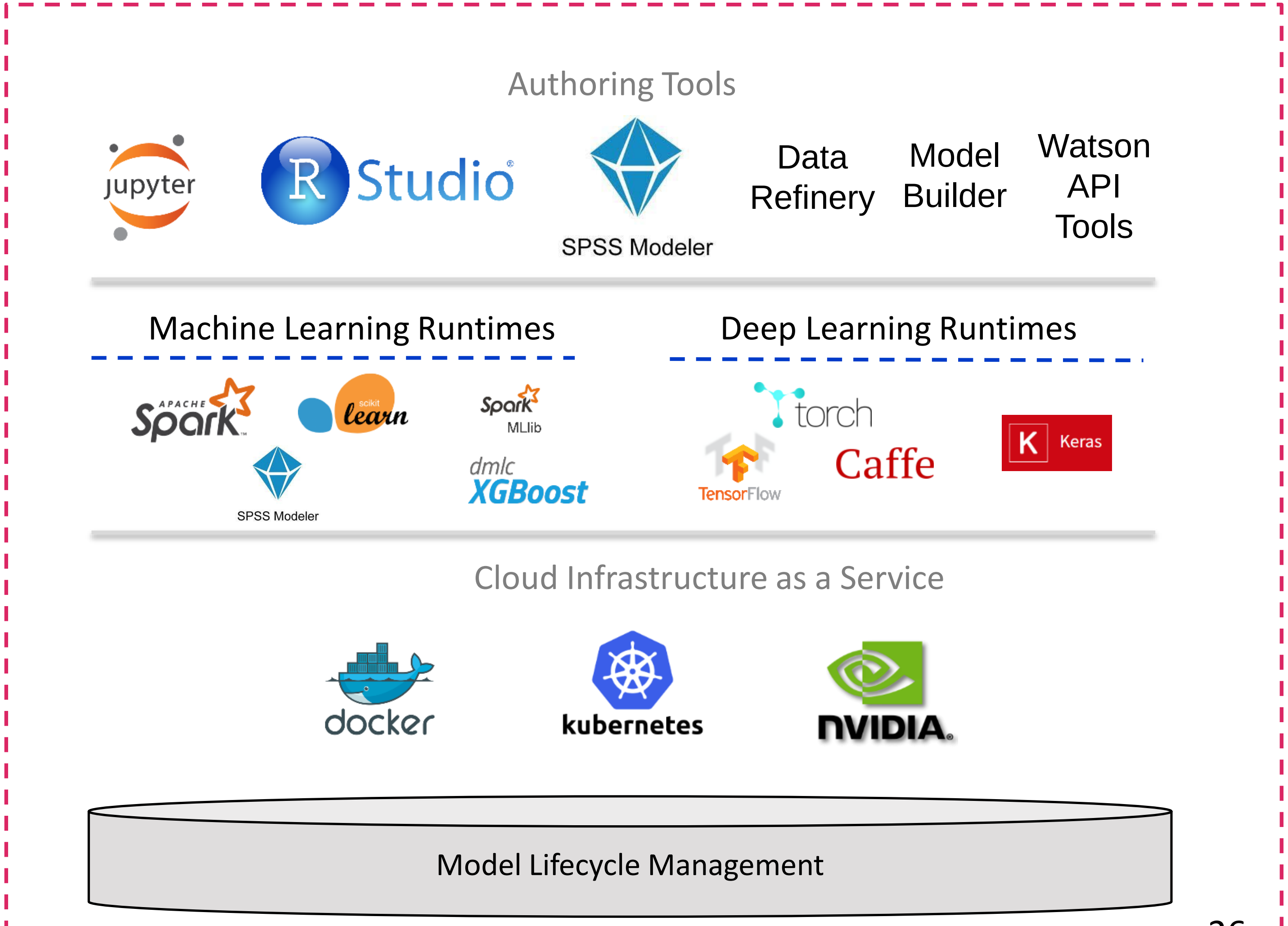
Connect & Access Data	Search and Find Relevant Data	Prepare Data for Analysis	Build and Train ML/DL Models	Deploy Models	Monitor, Analyze and Manage
Connect and discover content from multiple data sources in the cloud or on premises. Bring **structured** and **unstructured** data to one toolkit.	**Find** 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.	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.	**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 your models easily and have them **scale automatically** for online, batch or streaming use cases	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



Watson Studio
Differentiating Capabilities

Integrated Collaboration Environment

- Data Scientists, Subject Matter experts, Business Analysts & Developers all in one environment to accelerate innovation, collaboration and productivity
- Built-in learning to get started or go the distance with advanced tutorials

Choice of Tools for the full AI lifecycle

- Best in-breed open source and IBM tools that support the end-to-end AI lifecycle
- Choice of code or no-code tools to build and train your own ML/DL models or easily train and customize pre-trained Watson APIs

Support for all levels of expertise

- Use Watson smarts and recommendations for the best algorithms to use given your data, OR
- Use the rich capabilities and controls to fine tune your models

Experiment centric DL workflow

- Monitor batch training experiments then compare cross-model performance without worrying about log transfers and scripts to visualize results.
- You focus on designing your neural networks. We'll manage and track your assets.

Model lifecycle & management

- Deploy models into production then monitor them to evaluate performance.
- Capture new data for continuous learning and retrain models so they continually adapt to changing conditions.

Integrated with Knowledge Catalog

- Intelligent discovery of data and AI assets that enables reuse & improves productivity
- Seamlessly integrated for productive use with Machine Learning and Data science
- Powerful governance tools to control and protect access to data

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.

Existing Customer

Geo: Japan
Sector: Human Resource Management/ IT Professional Services

Background

- Forum Engineering (FE) is a **leading human resource management company** specialized in engineering in Japan. They provide engineer staffing services to their clients across different sectors, including Automotive, Industrial Machinery, Electricity, Electronics, Precision Equipment, and Information and Communications sectors. Their clients are leading manufacturers in Japan, and they introduce human resources to their clients and job opportunities to engineers.
- They conduct thorough research on client's technical needs, corporate culture and requirement then pair them with engineers' technical capabilities, personal preferences and personality to ensure the best match and experience for both parties.

Business Problem

- The organization spent a majority of their time and resources dispatching sales people to conduct lengthy interviews and evaluations with their clients and engineer candidates.
- With such a wide range of clients' technical demands, and large variation in engineers' skills and specialties, the manual background research, ensuing analyses, and matching performed were extremely time-consuming. They find the limitation of growth because of their labor-intensive process.
- Forum Engineering needed help to increase growth rate by automating manual process and shifting to knowledge-based business, and provide their quality engineer staffing services and accurate matching program.

Solution

- Powered by IBM Watson and Watson Studio, Forum Engineering released 2 new initiatives- **"Insight Matching"** and **"Cognitive Staffing"** to fundamentally change the way they run their business. The IBM-powered solution allows FE to gather & analyze a massive amount of internal and external data, and quickly calculate matching ratios based on its analysis and reasoning. They also use the solution to build matching dictionaries (seeding keywords/content to look for), and to run matching accuracy tests. In particular, Watson Studio provides not only the data analysis, but Spark data processing and Machine Learning capabilities that FE was seeking.
- The solution is expected to **dramatically improve their matching accuracy, eliminate human bias from the process, and increase their sales/staffing efficiency. They expect further growth by shifting surplus resources to growth area and new business development.**

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

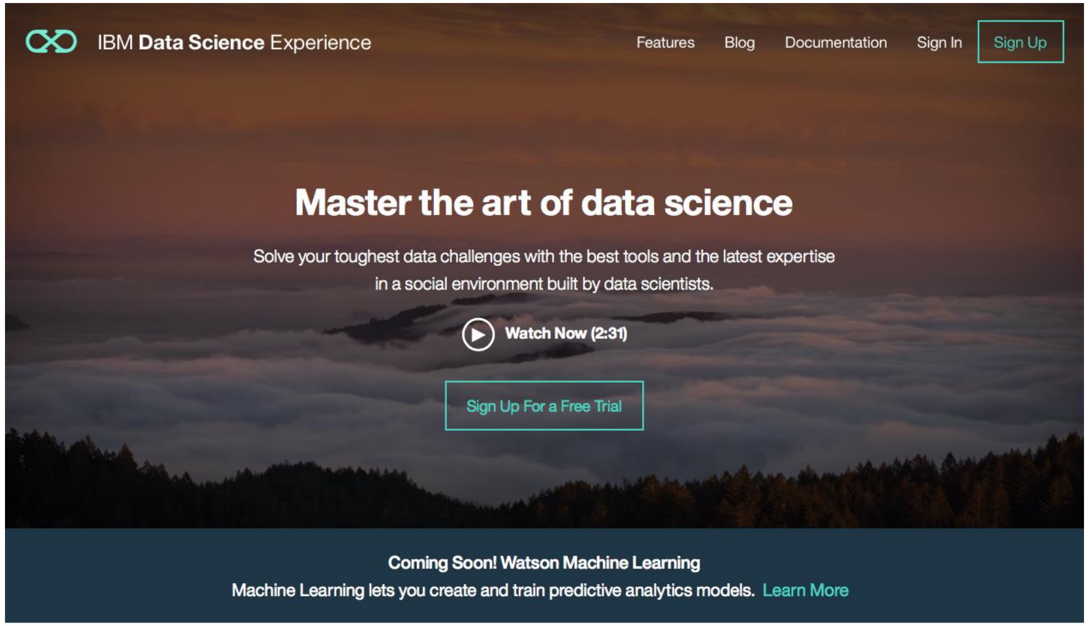
- IBM Machine Learning*
- SPSS Modeler Canvas*
- Prescriptive Analytics - DOcplexcloud
- Projects and Version Control
- Managed Spark Service

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

* beta

Core Attributes of Watson Studio/ Data Science Experience

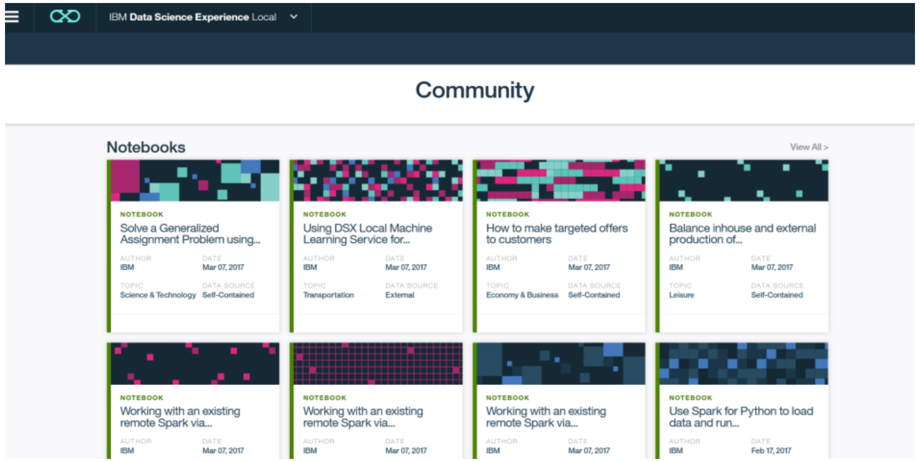
- **Data Science as a Team Sport**
  Lets *data scientists/engineers*, *analysts*, stakeholders
  collaborate to collect, share, explore*, analyze data*
  in order to *derive insights and train models,*
  and share or deploy resulting assets
- **Projects** - collaborate as team or work individually
- **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**

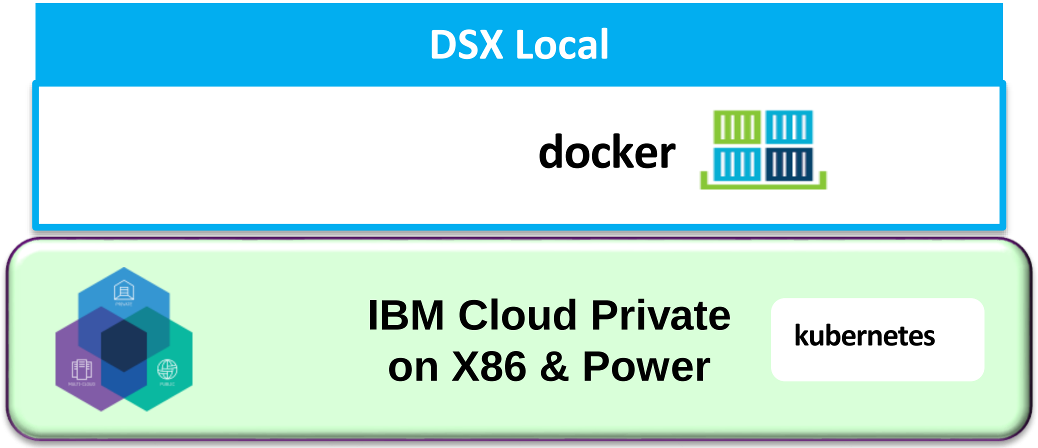


Try it yourself at https://datascience.ibm.com

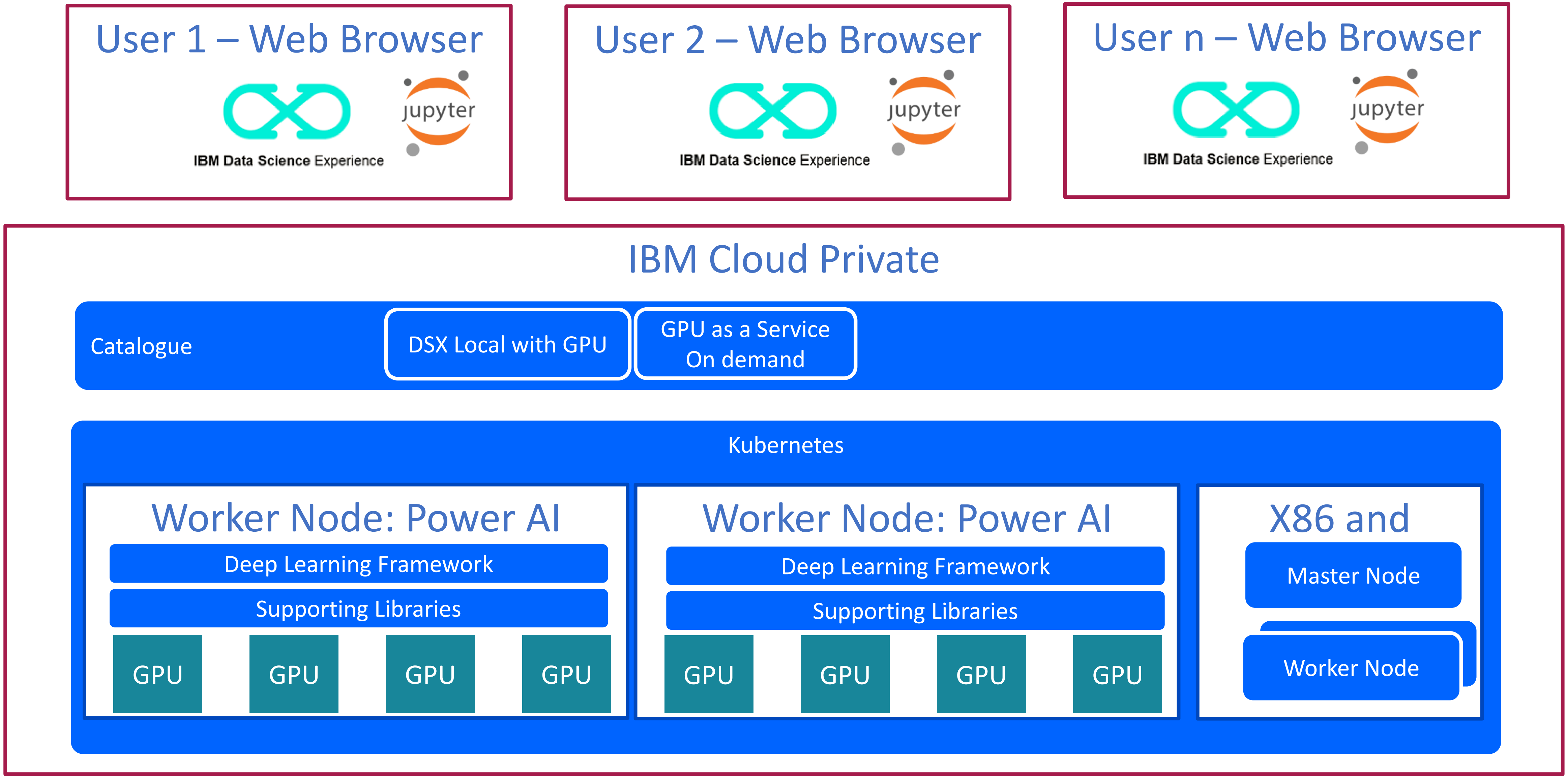
Core Attributes of Watson Studio/ Data Science Experience

- 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





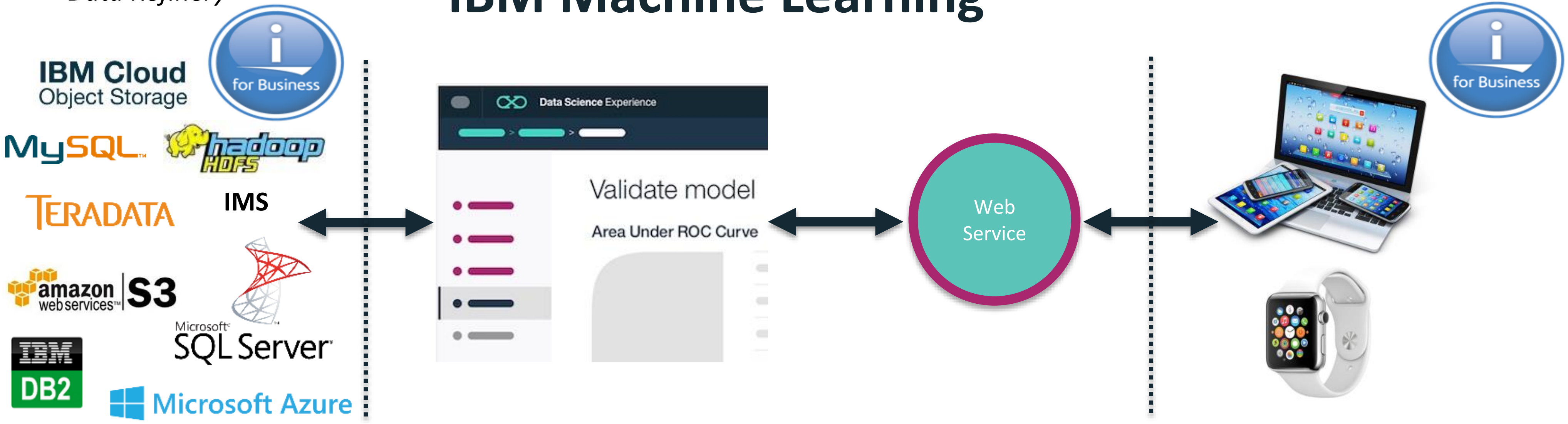
Deployment Example: IBM Private Cloud & DSX Local



Core Attributes of Watson Studio/ Data Science Experience

IBM Data Connect
"Data Refinery"

IBM Machine Learning



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

IBM i & Watson Studio / DSX
+ Demonstration

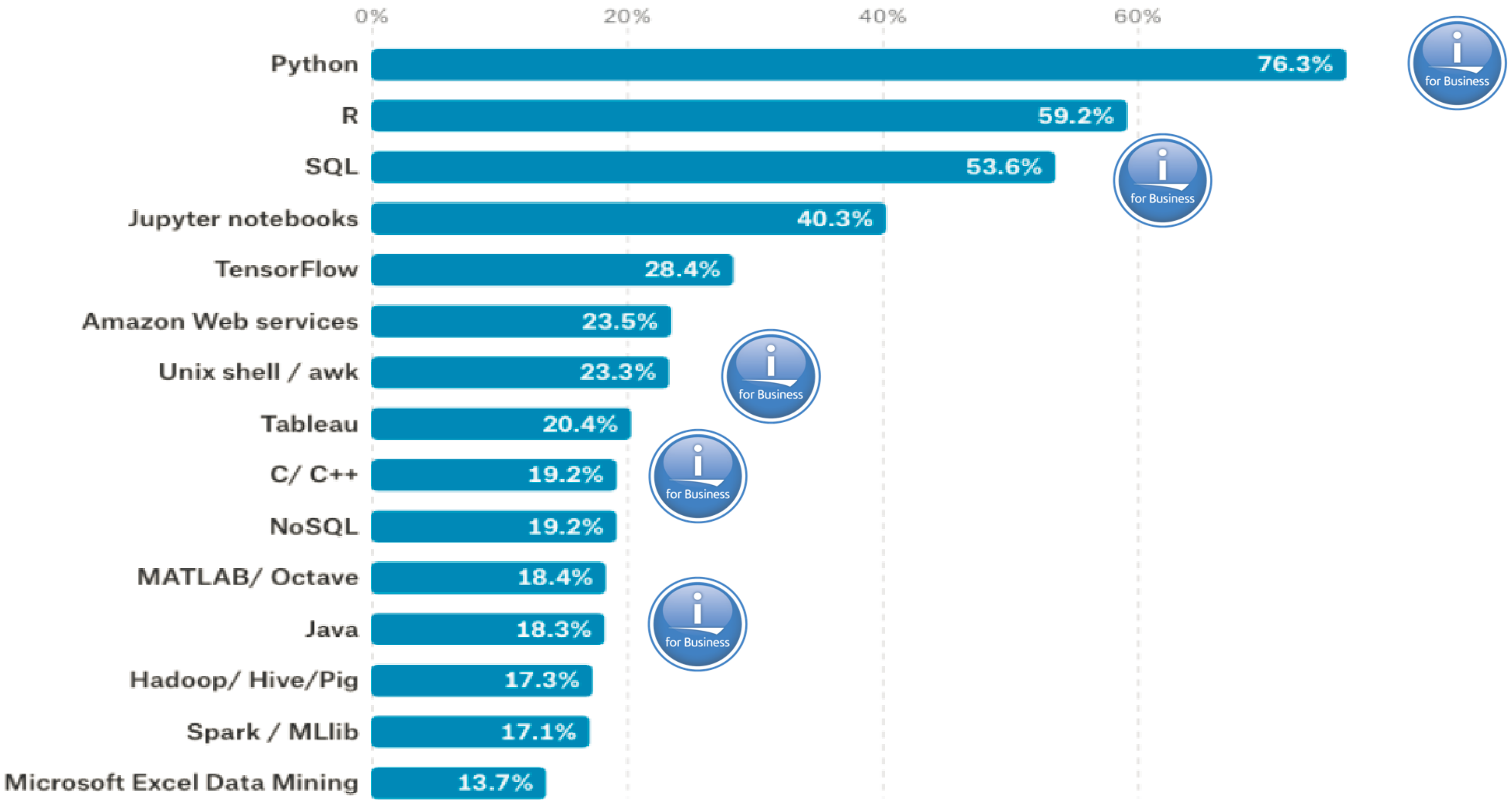
IBM i , AI & Data Science

- AI = Algorithms + Data (including Data in Db2 for i)
- Native (RPG) & New Db2 features & Languages available on i facilitates the IBM i ←→ 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

AI & Predictive Analytics,
Mobile, IoT, API Economy…

Data Science tools & technologies

Kaggle 2017 Data Science Tools Survey



Ready for AI: Connecting your business to the future

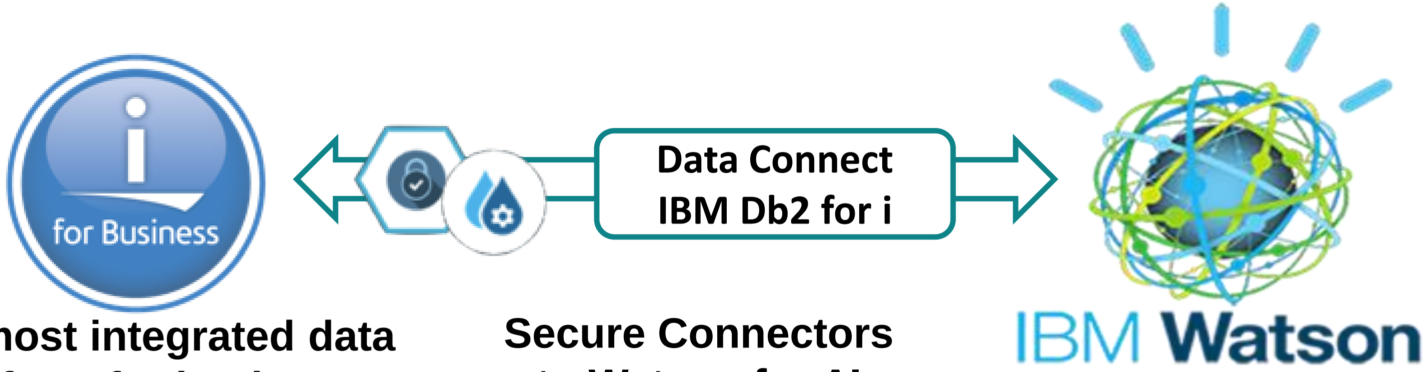
JORI The art of fine seating

Ready for AI: Connecting your business to the future

thirty years

JORI The art of fine seating

- 100+ 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% 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

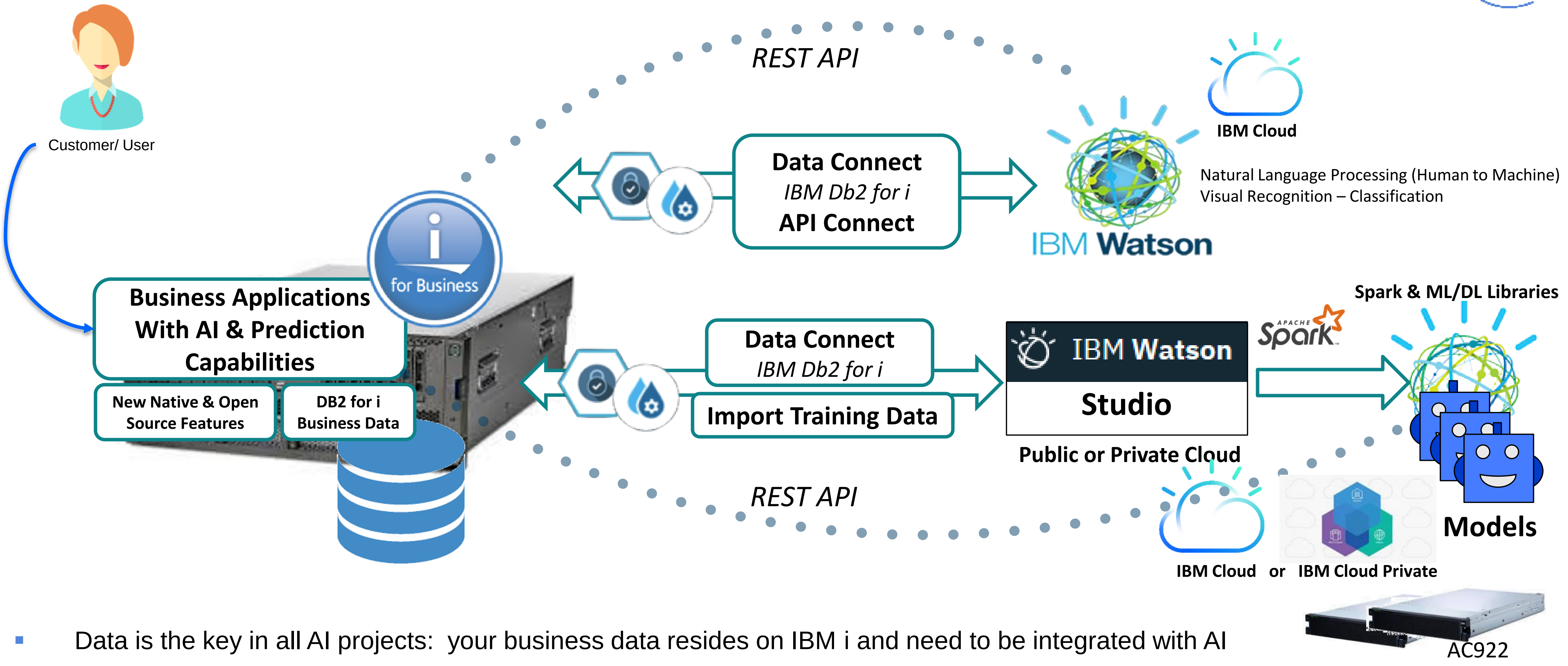
Modern, Open Source applications development

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

Secure Connectors to Watson for AI

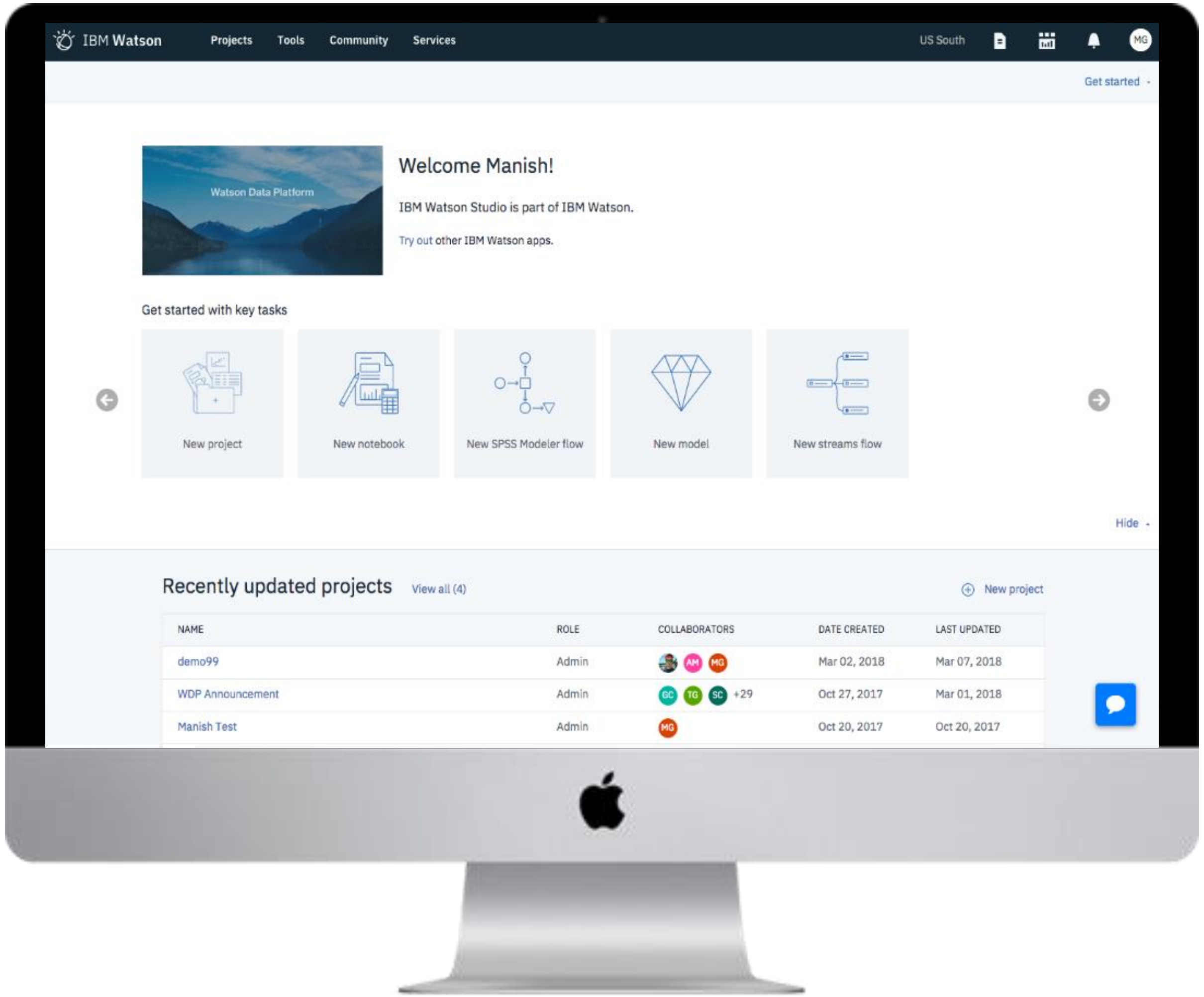
- Data Connect for Db2 for i
- SQL
- Python, Node.JS
- Free form RPG

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 premises (DSX Local w/ Cloud Private)

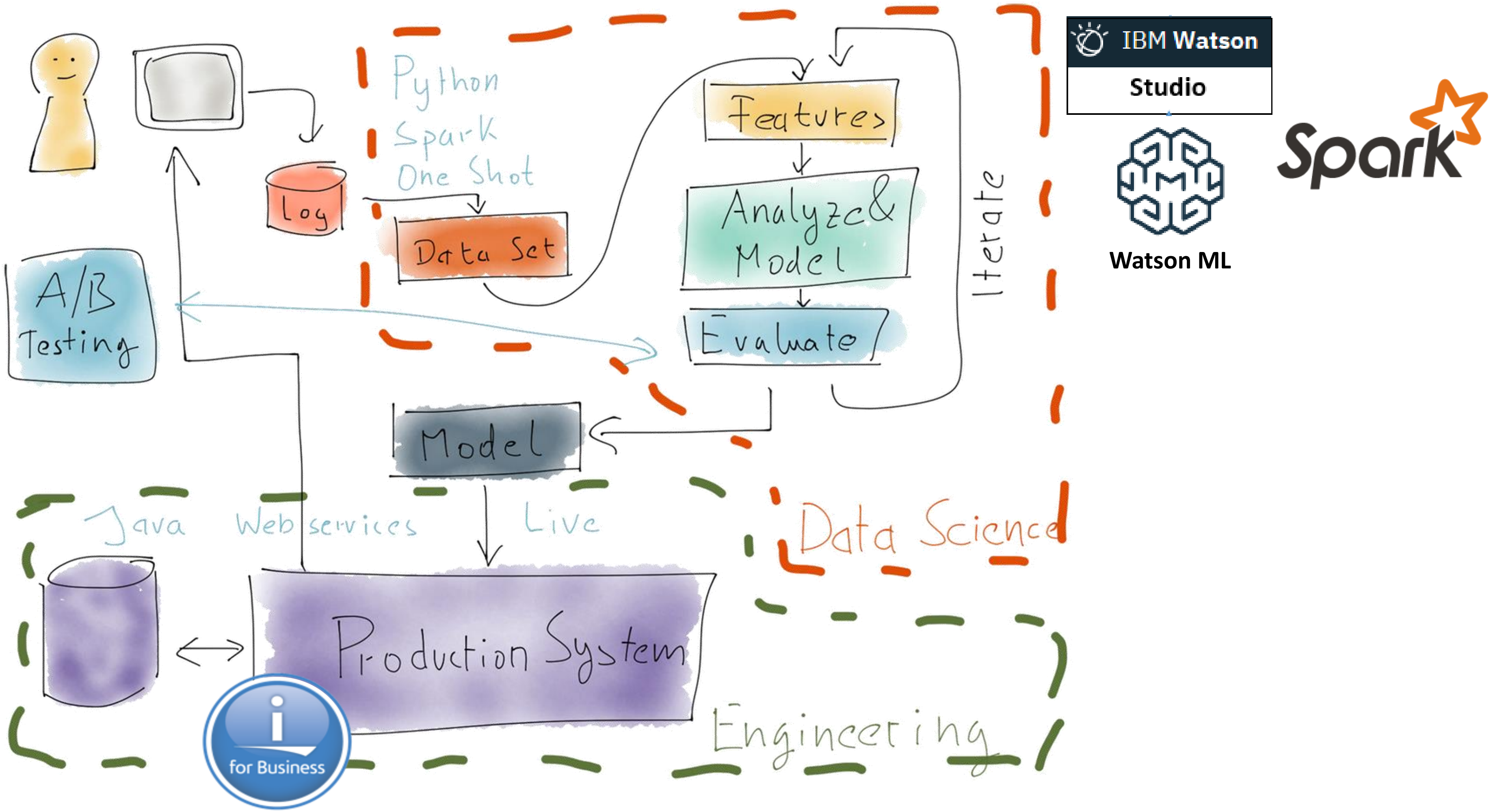
Demo



IBM Watson
Projects
Tools
Community
Services
US South
Get started
Welcome Manish!
IBM Watson Studio is part of IBM Watson.
Try out other IBM Watson apps.
Watson Data Platform
Get started with key tasks
New project
New notebook
New SPSS Modeler flow
New model
New streams flow
Hide
Recently updated projects
View all (4)
New project
NAME
ROLE
COLLABORATORS
DATE CREATED
LAST UPDATED
demo99 Admin Mar 02, 2018 Mar 07, 2018
WDP Announcement Admin +29 Oct 27, 2017 Mar 01, 2018
Manish Test Admin Oct 20, 2017 Oct 20, 2017

Machine Learning & IBM i Demo:

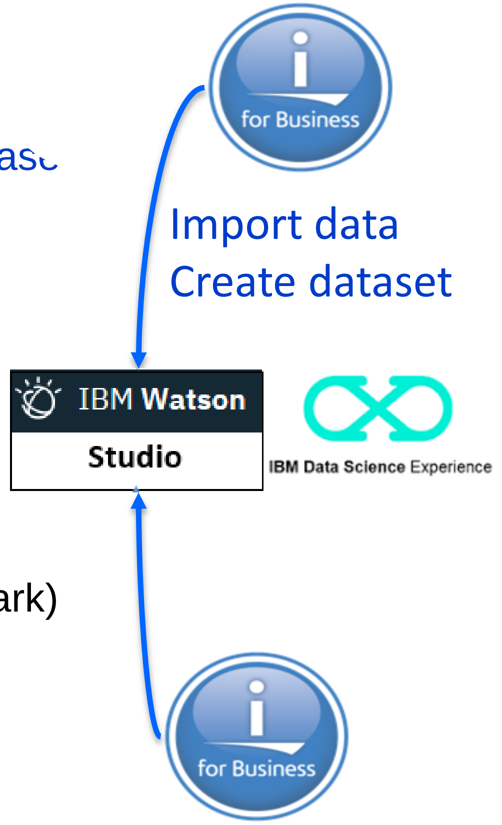
Predict outdoor equipment purchase



Machine Learning & IBM i Demo:

Predict outdoor equipment purchase

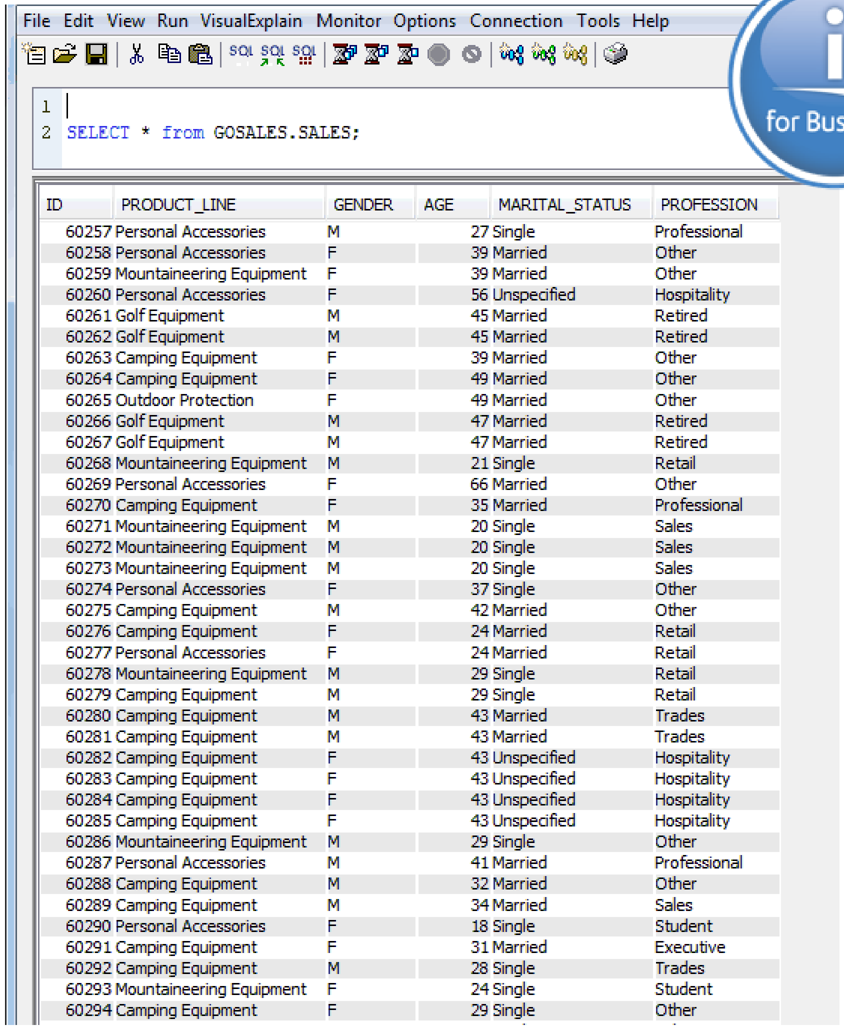
1. Export or connect your data on IBM i
   - Use Access Client Solutions (ACS) for CSV Export or create a jdbc Connector to your Db2 for i database
   - In that demonstration, one table (GOSALES/SALES) containing historical data - outdoor equipment purchases -- Alternative: Direct connection to your 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.



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

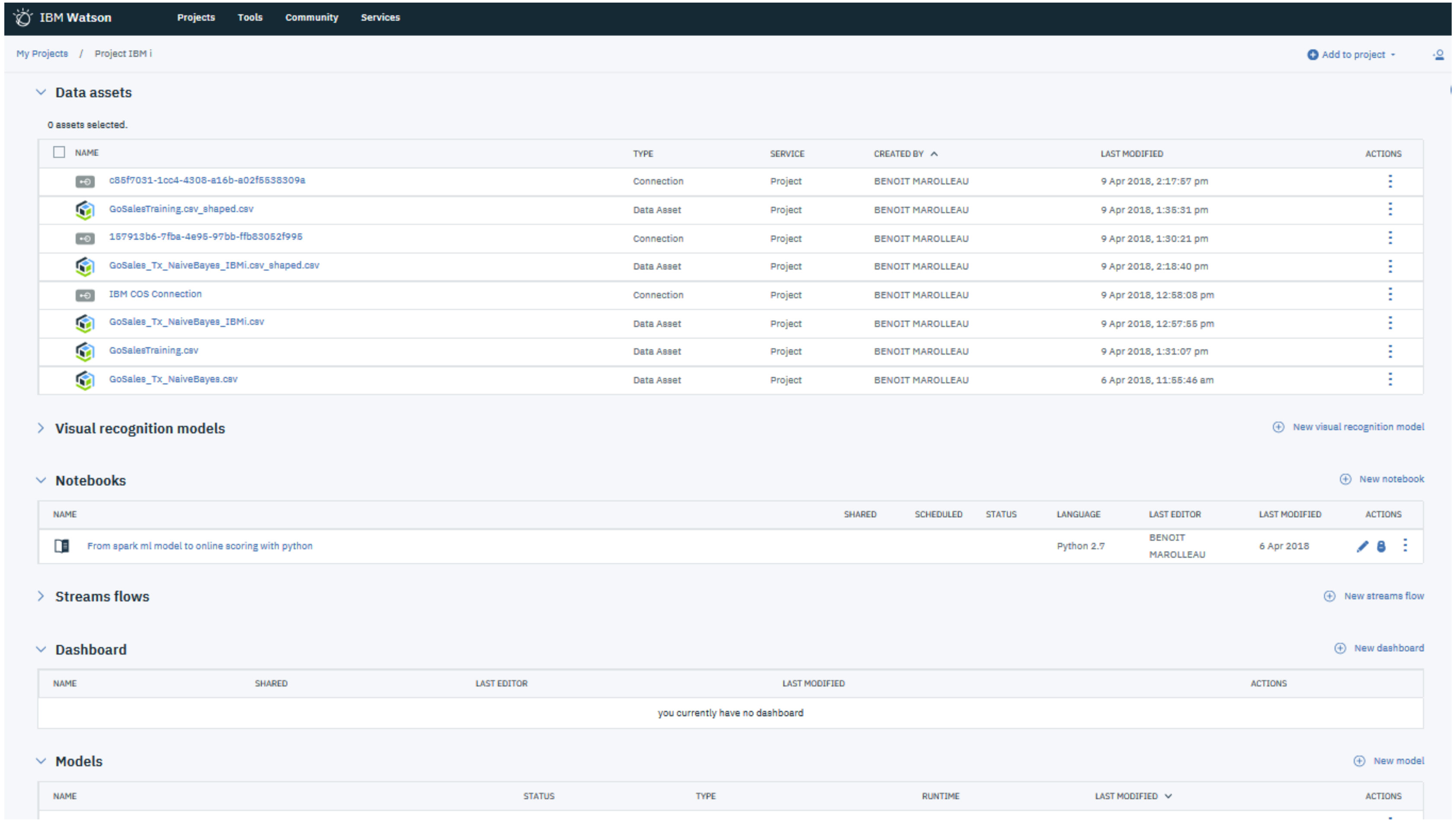
Projects

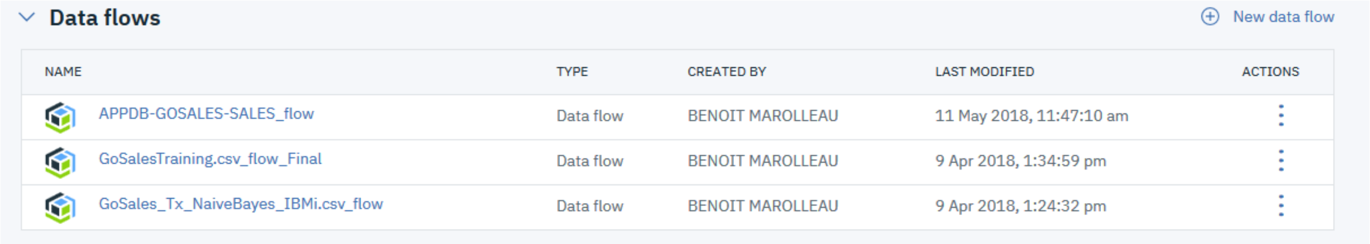
Data Assets
from import or
Direct datasource
connection

Code (Notebooks)
Work on the data
programmatically

Models
(Machine Learning)

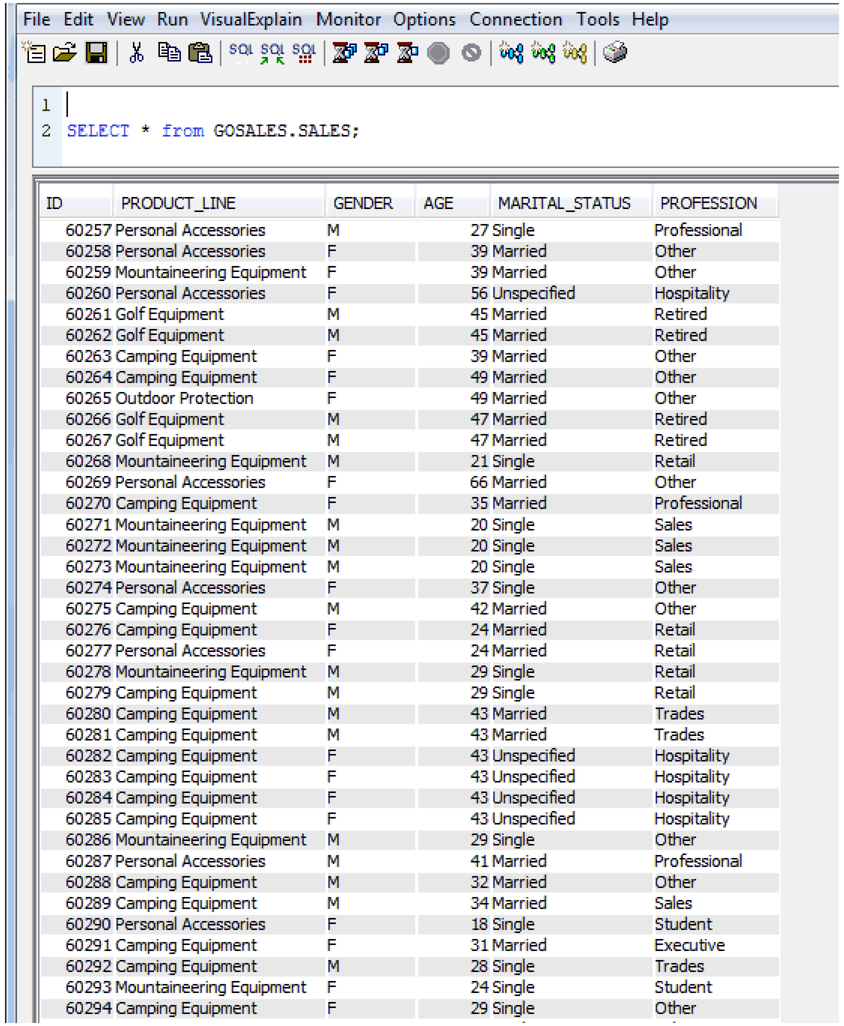
Data Flows
(Data Refinery)
from CSV
or direct Connections



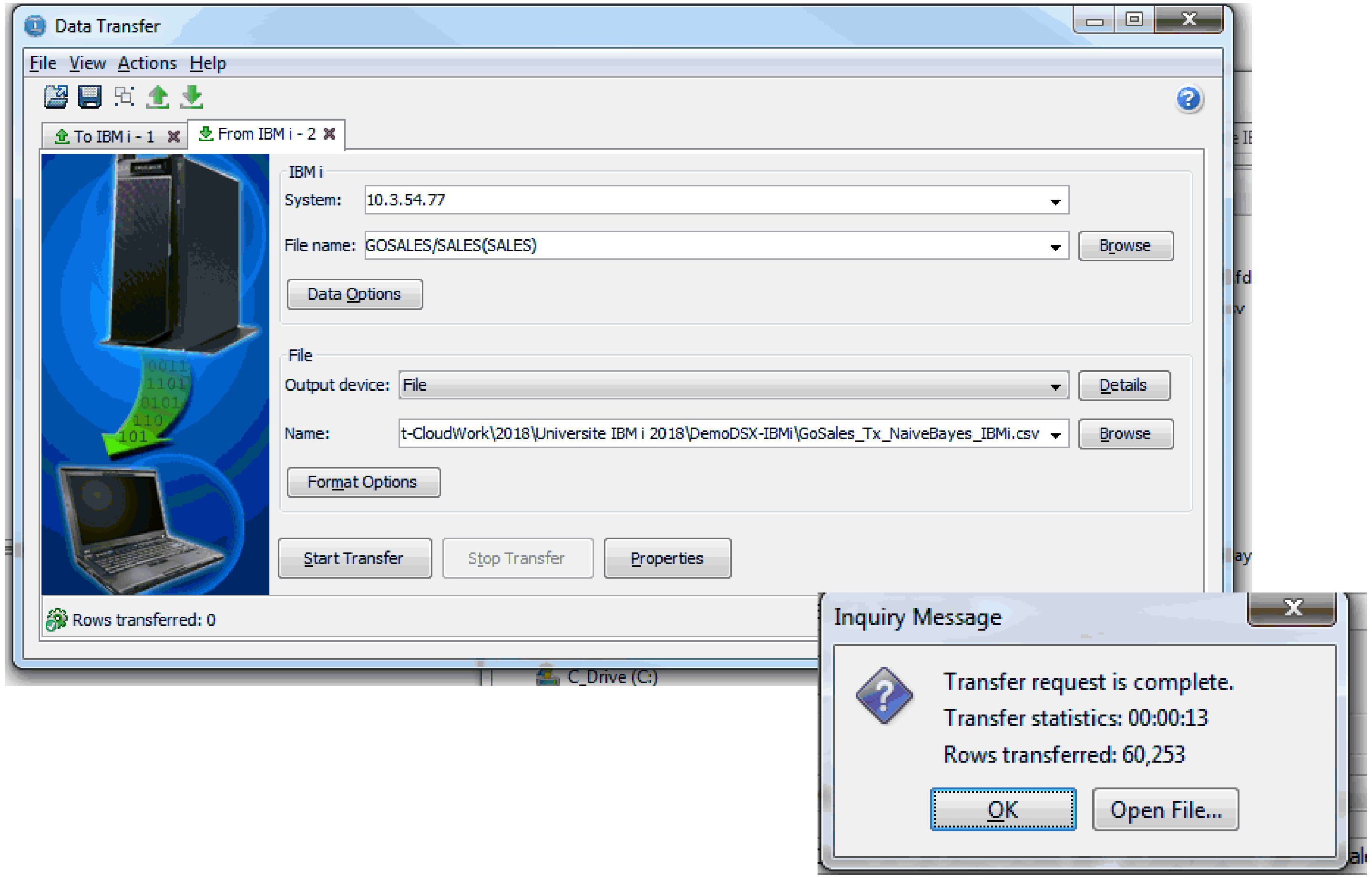


Extract Data

- 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

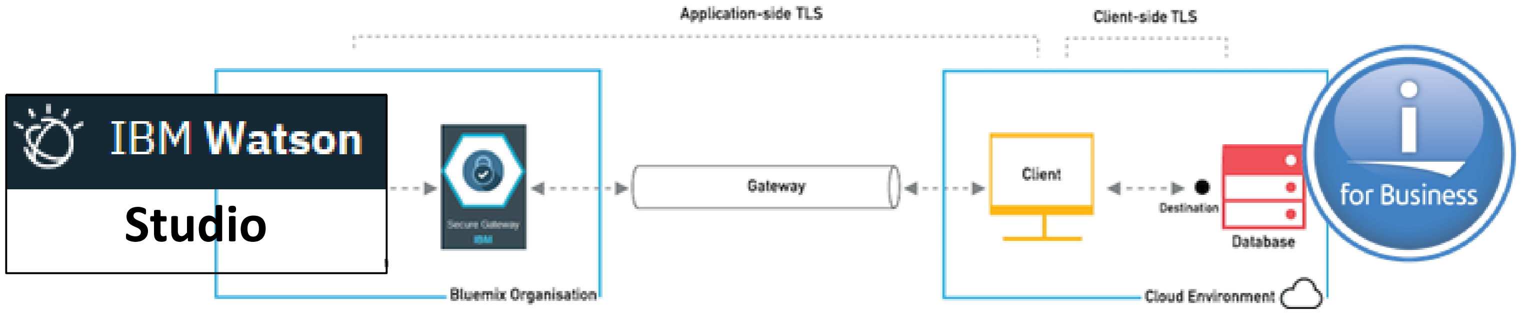


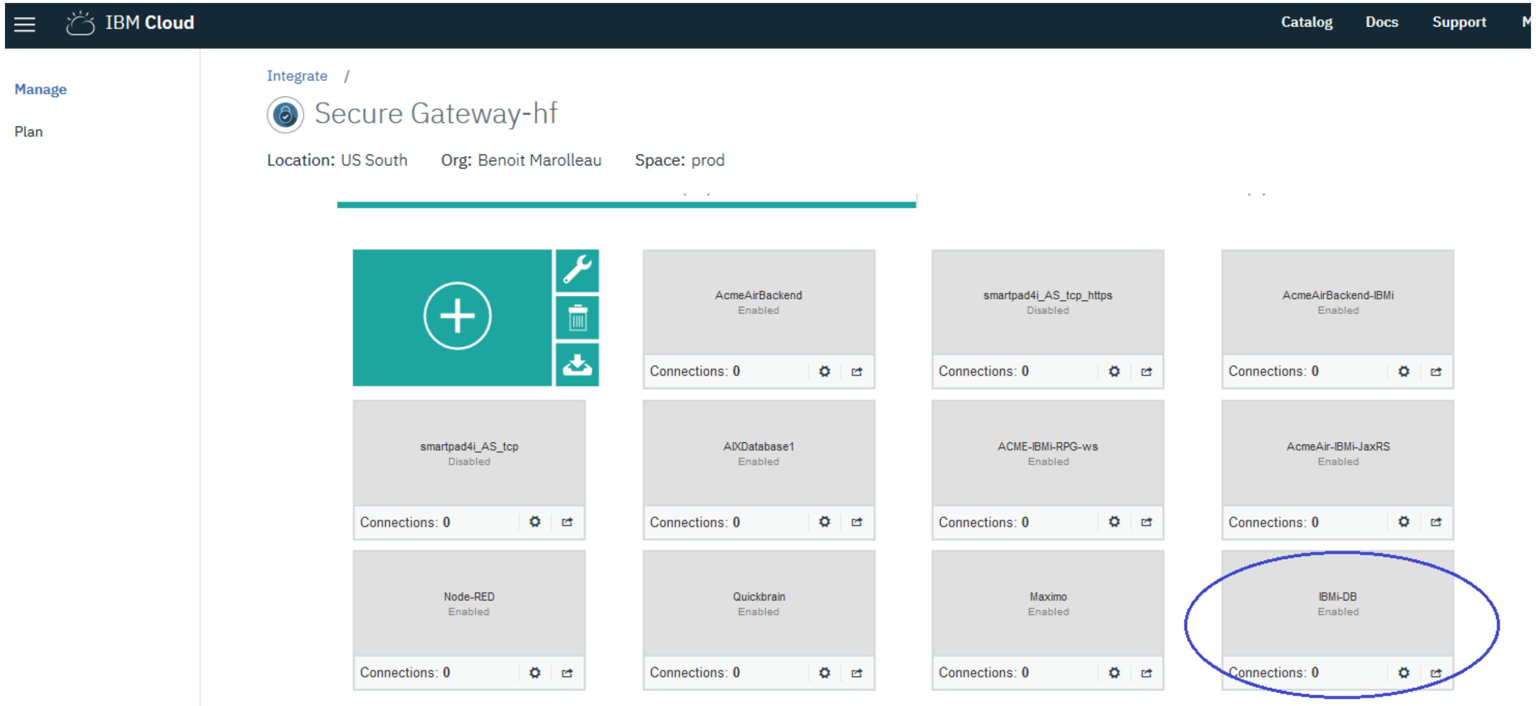
ID	PRODUCT_LINE	GENDER	AGE	MARITAL_STATUS	PROFESSION
60257	Personal Accessories	M	27	Single	Professional
60258	Personal Accessories	F	39	Married	Other
60259	Mountaineering Equipment	F	39	Married	Other
60260	Personal Accessories	F	56	Unspecified	Hospitality
60261	Golf Equipment	M	45	Married	Retired
60262	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



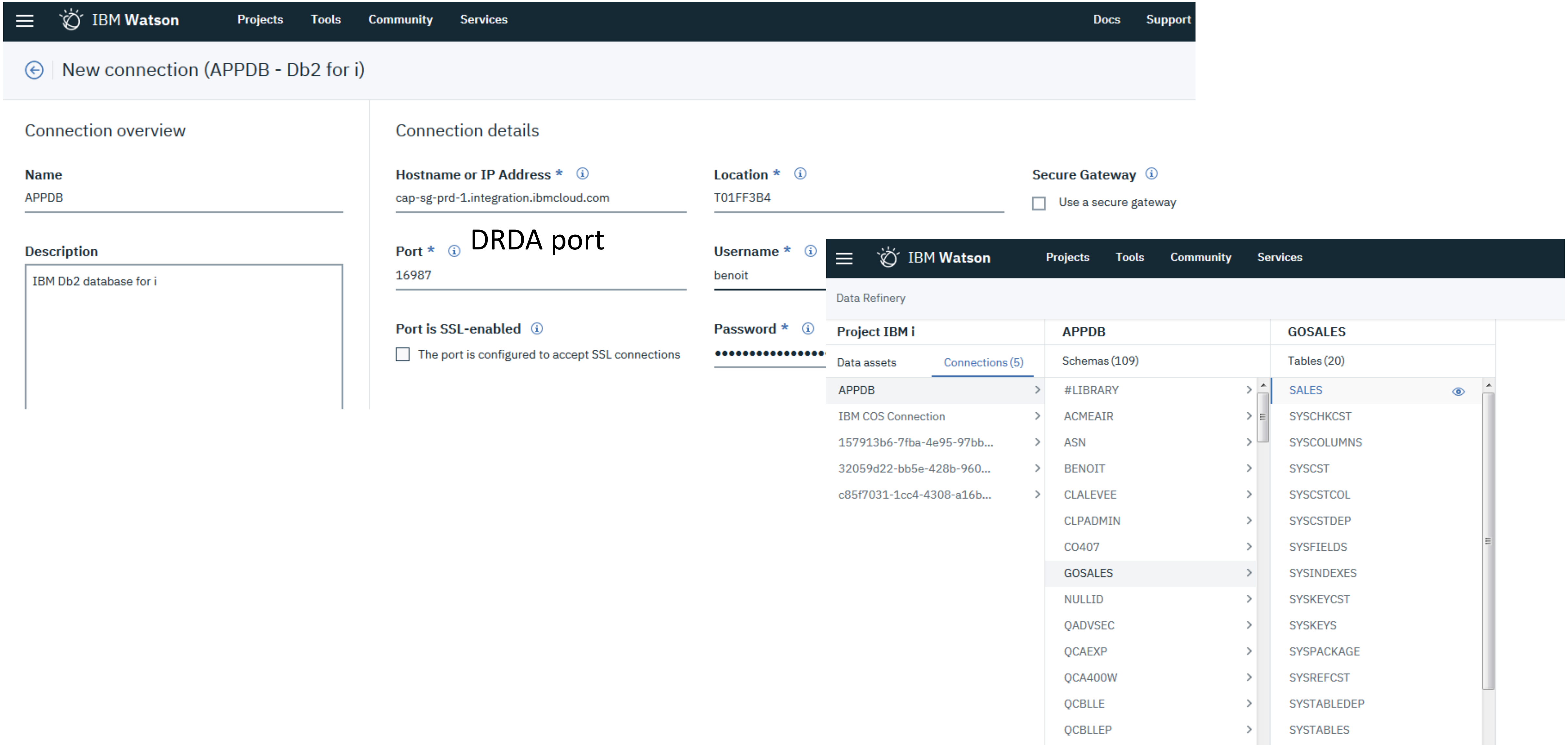
Alternative: Direct Connection to IBM i

- ❑ 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

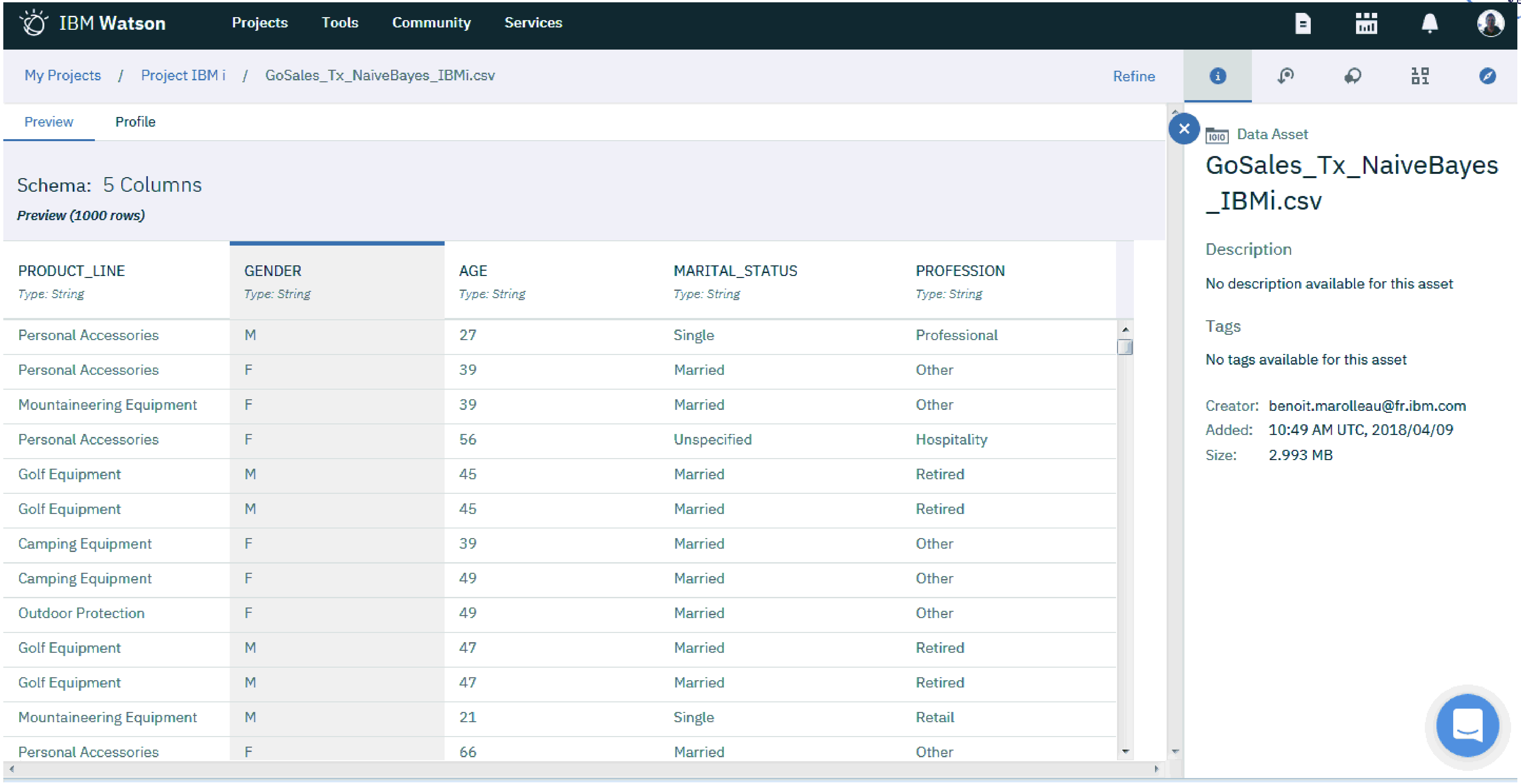




Alternative: Direct Connection to IBM i

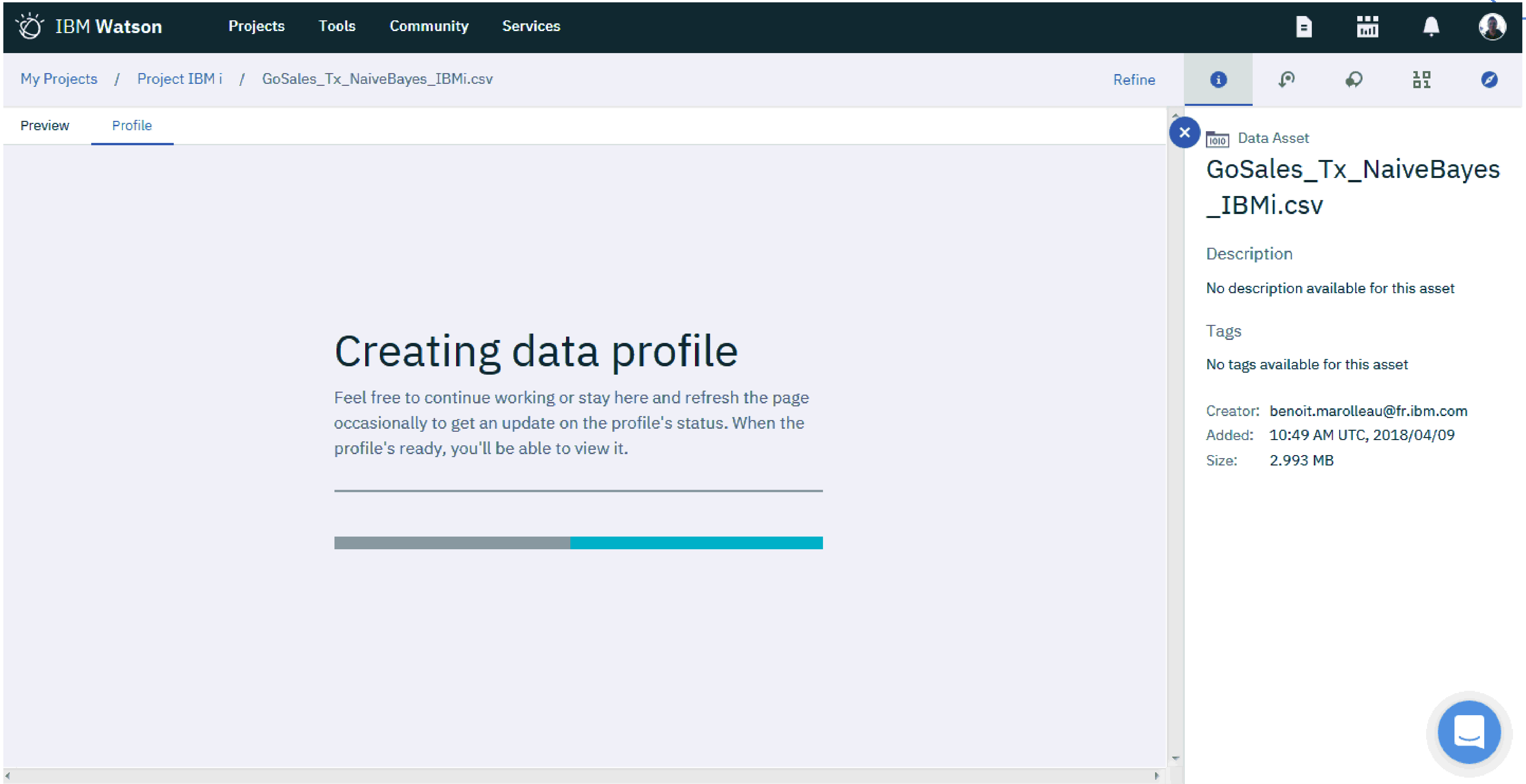


Data Asset (from Db2 for i)

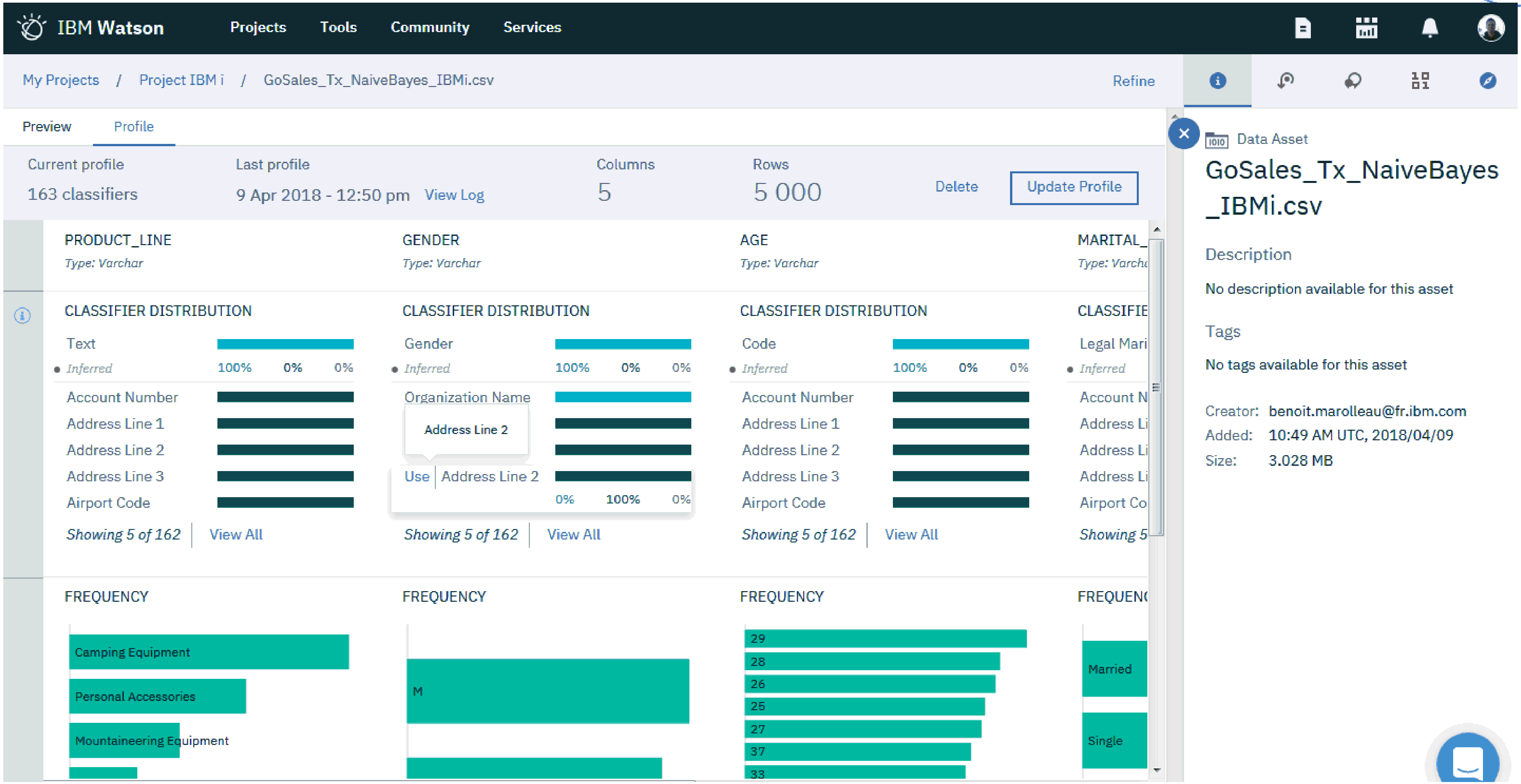


PRODUCT_LINE *Type: String*	GENDER *Type: String*	AGE *Type: String*	MARITAL_STATUS *Type: String*	PROFESSION *Type: String*
Personal Accessories	M	27	Single	Professional
Personal Accessories	F	39	Married	Other
Mountaineering Equipment	F	39	Married	Other
Personal Accessories	F	56	Unspecified	Hospitality
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

Data Asset (from Db2 for i)

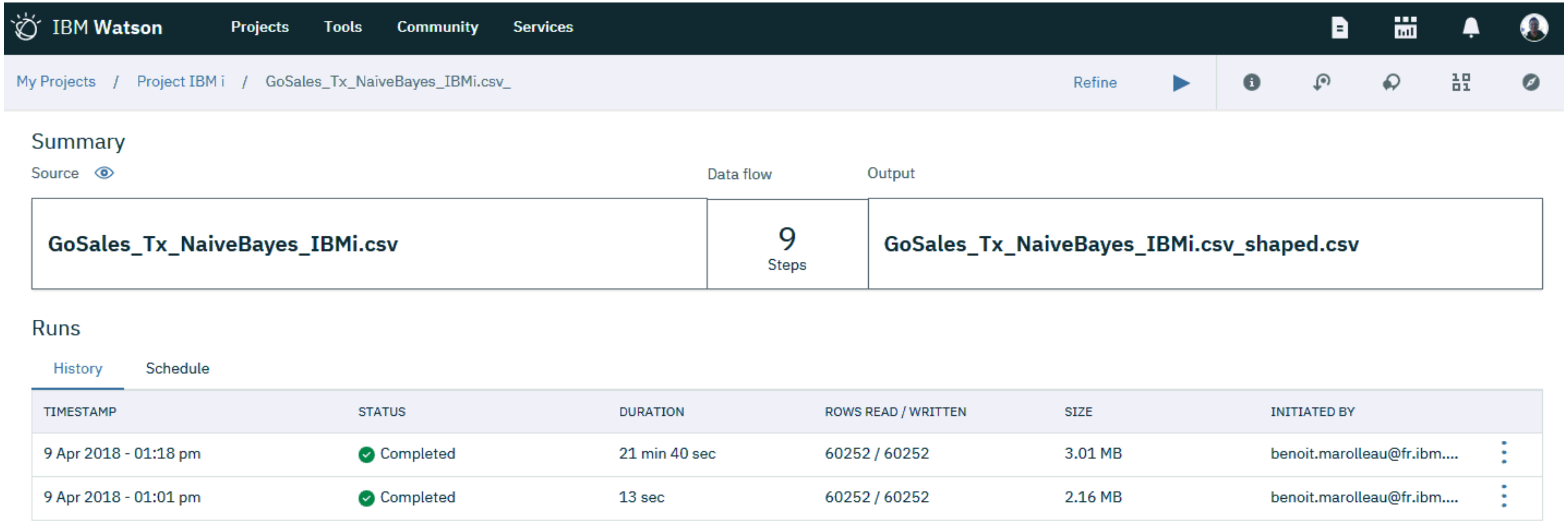


Data Asset (from Db2 for i)



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



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

IBM Watson | Projects | Tools | Community | Services | Docs | Support

Data Refinery

Project IBM i

Data assets | Connec

APPDB

IBM COS Connection

157913b6-7fba-4e95-97bb

32059d22-bb5e-428b-960.

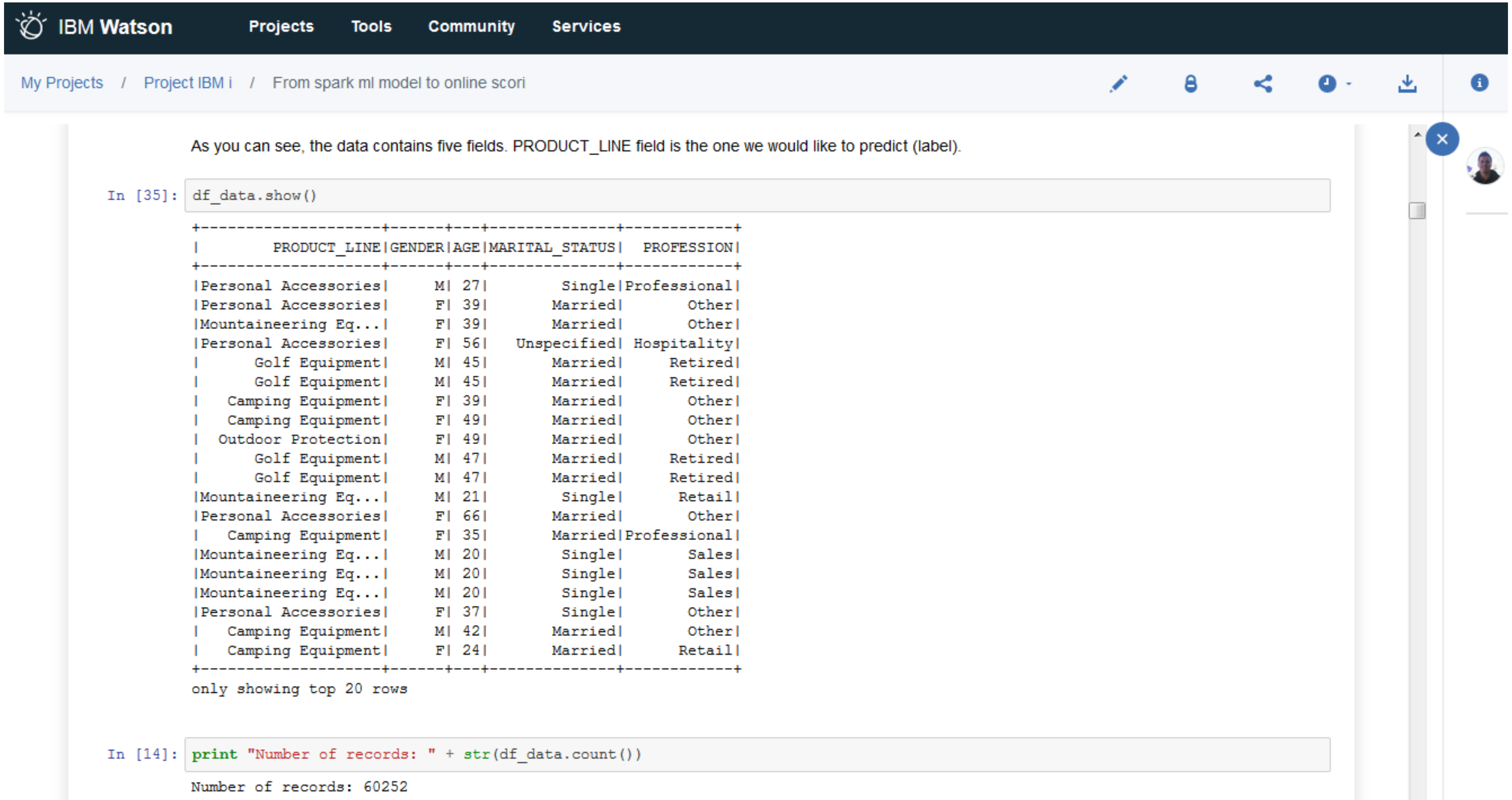
c85f7031-1cc4-4308-a16b.

SALES

ID Type: Integer	PRODUCT_LINE Type: String	GENDER Type: Char	AGE Type: Smallint	MARITAL_STATUS Type: String	PROFESSION 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

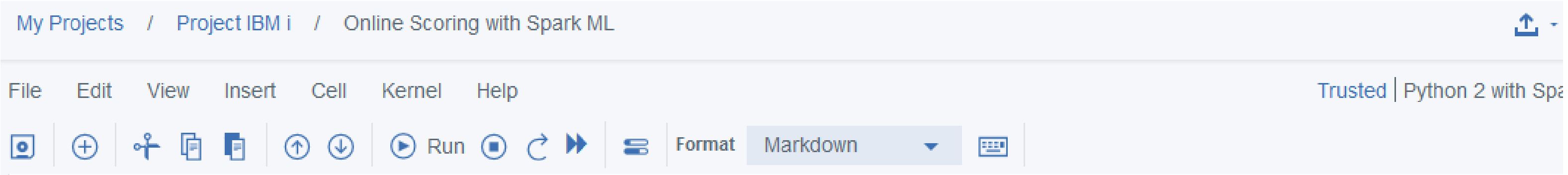
Work on your data: Jupyter Notebook

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



Data Preparation / Model Training & Deployment

using a Python Jupyter Notebook



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.

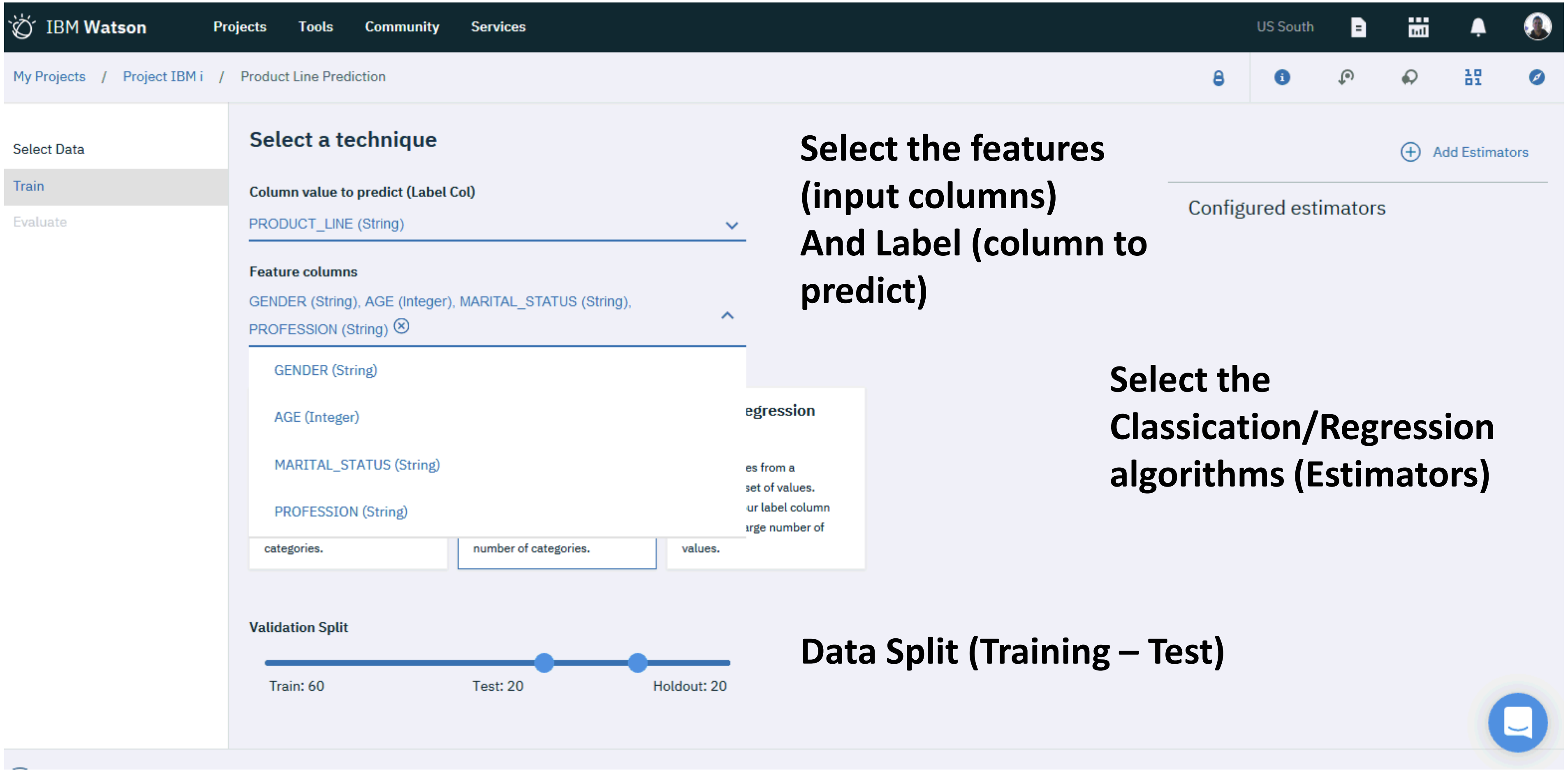
In [64]:
```
wml_credentials={
  "url": "https://ibm-watson-ml.mybluemix.net",
  "access_key": "mN3bwIsi/xpOh08Bd8D8Zah6UF++y/ZqlqTlJDmJQr6dqC83AFnAyPemGJw9js0JHxGxQ3pIogjgEOjN0TGDTcL0h32gVzPkwMbmHX
  "username": "75693a81-d133-4c15-8570-3dc0f0f092c1",
  "password": "13cb03fd-2b6b-4b24-a17f-49eb2efaf7b9",
  "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 **Serice Credentials** generate new credentials by pressing **New credential (+)** button.

Demo Notebook: https://github.com/bmarolleau/gosales-ibmi

Model Training & Deployment

using the Model Building Wizard

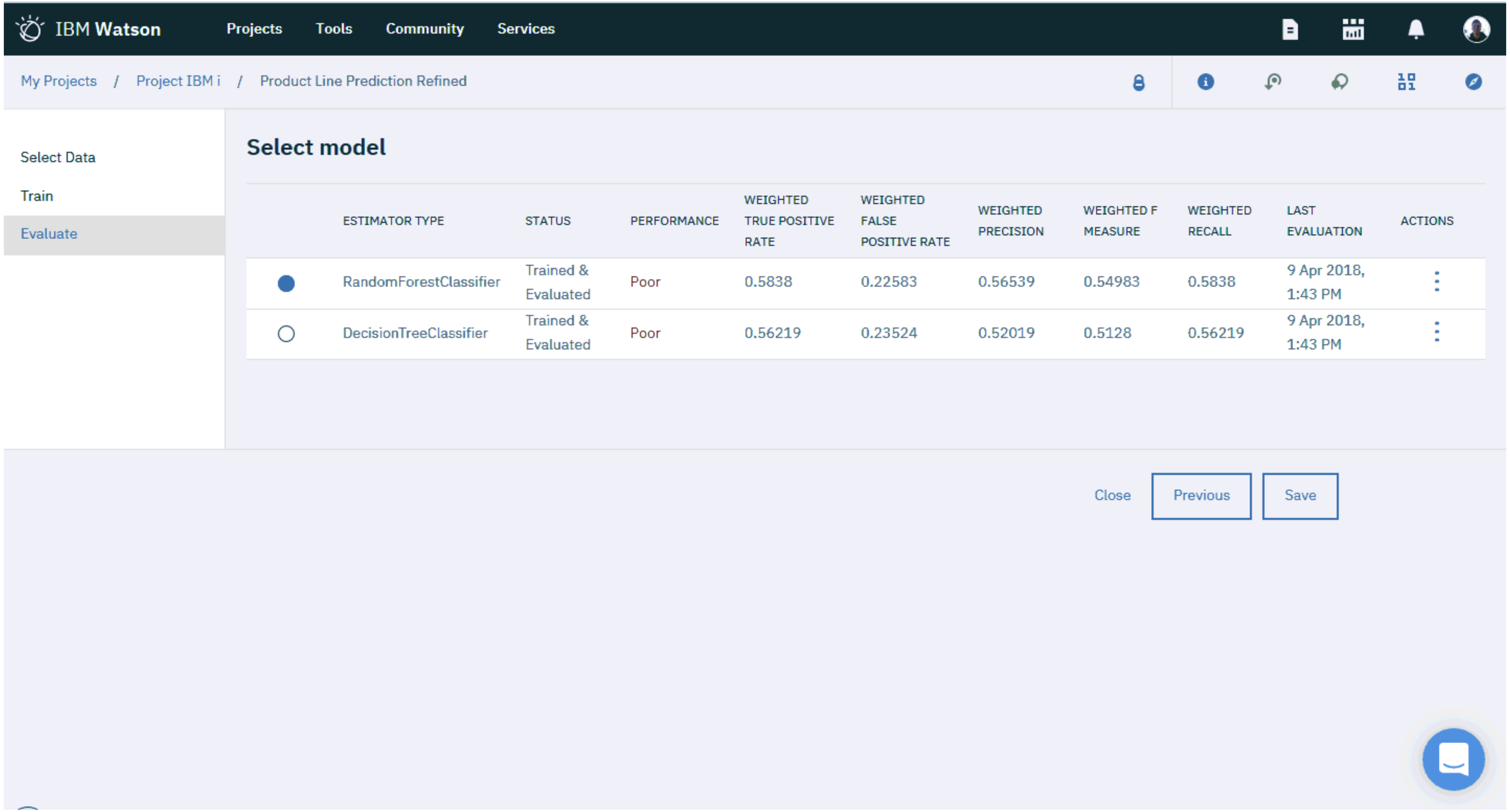


Select the features (input columns)
And Label (column to predict)

Select the Classication/Regression algorithms (Estimators)

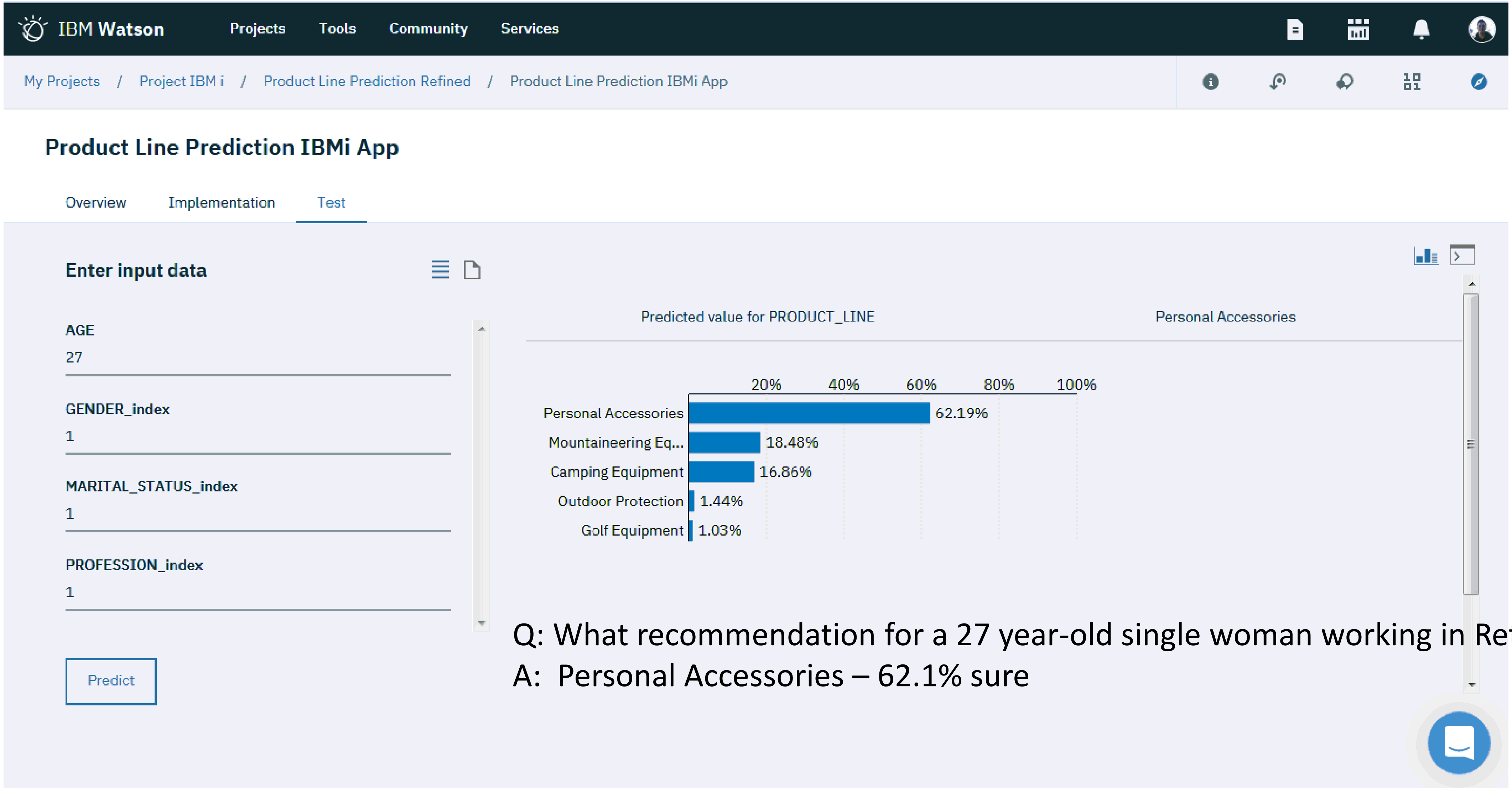
Data Split (Training – Test)

Model Training & Evaluation



	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	⋮

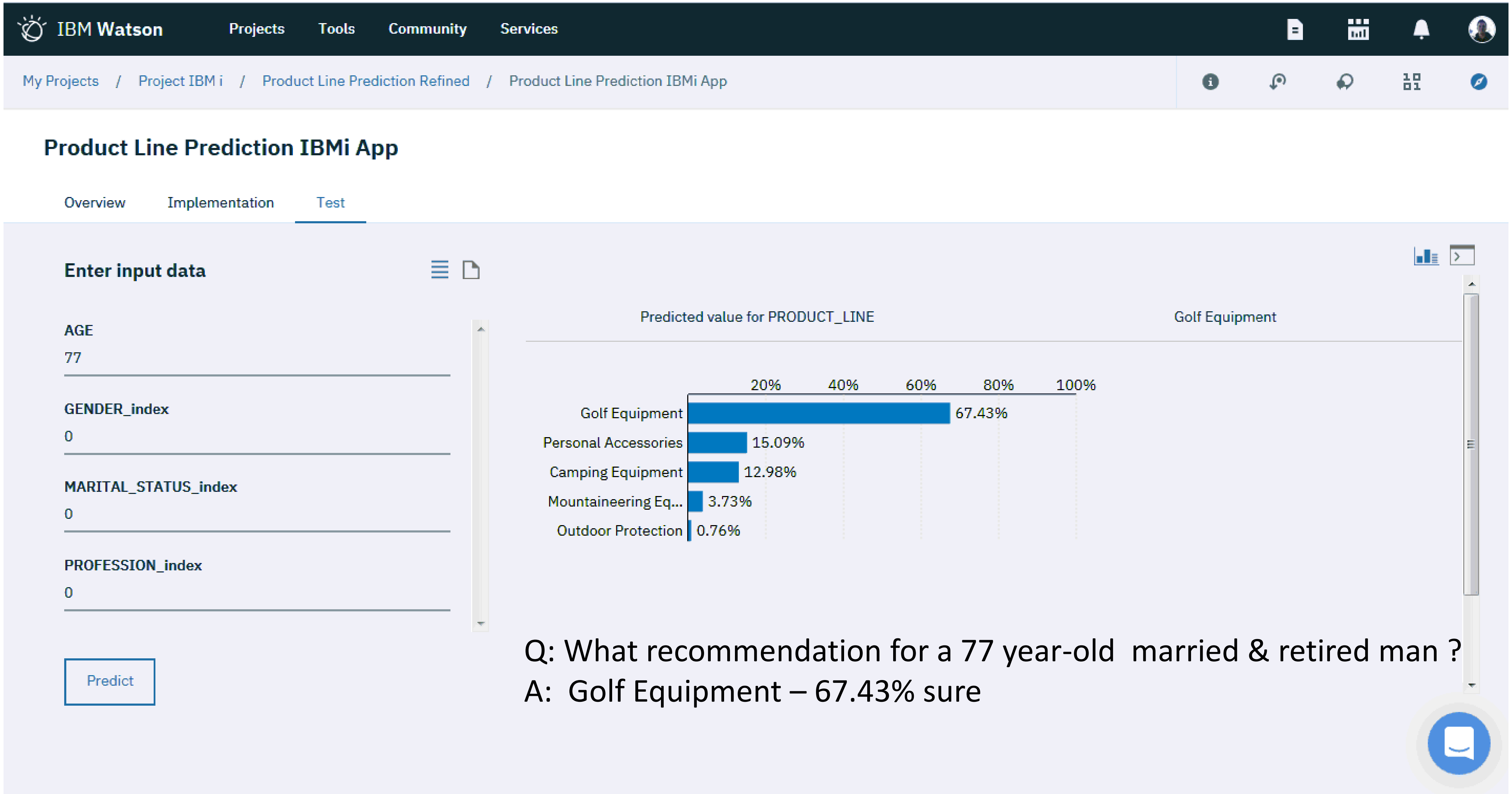
Model deployment, test, publication



Q: What recommendation for a 27 year-old single woman working in Retail?

A: Personal Accessories – 62.1% sure

Model deployment, test, publication



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` :

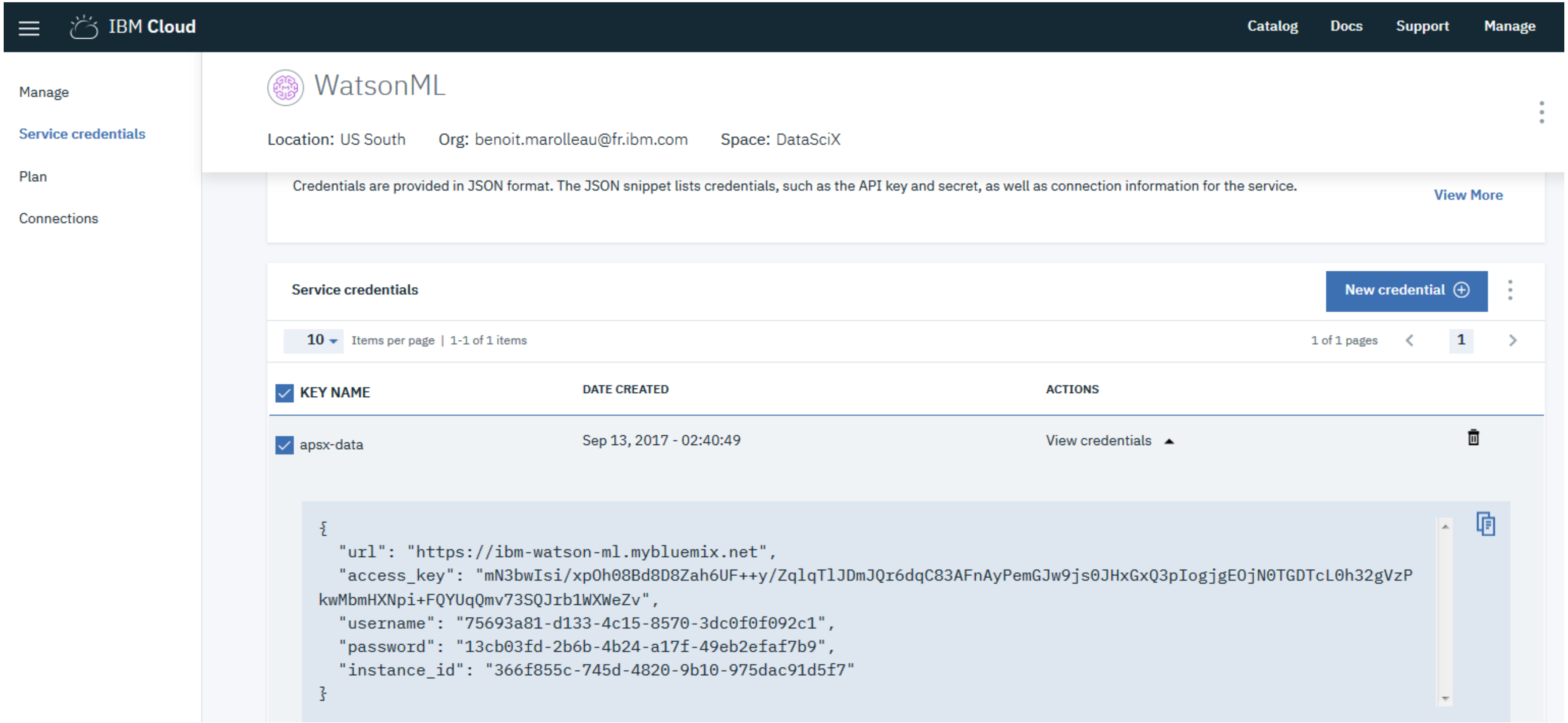
```
curl https://ibm-watson-ml.mybluemix.net/v3/wml_instances/00fd89e6-8cf2-4712-a068-ade10277b649/published_models -H "Authorization: Bearer eyJhbGciOiJSUzUxMiIsInR5cCI6IkpXVCJ9.eyJ0ZW5hbnRJZCI6ImU4YmQzZGM3LWI5Y2UtNDY1OC1iZ..."
```

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.

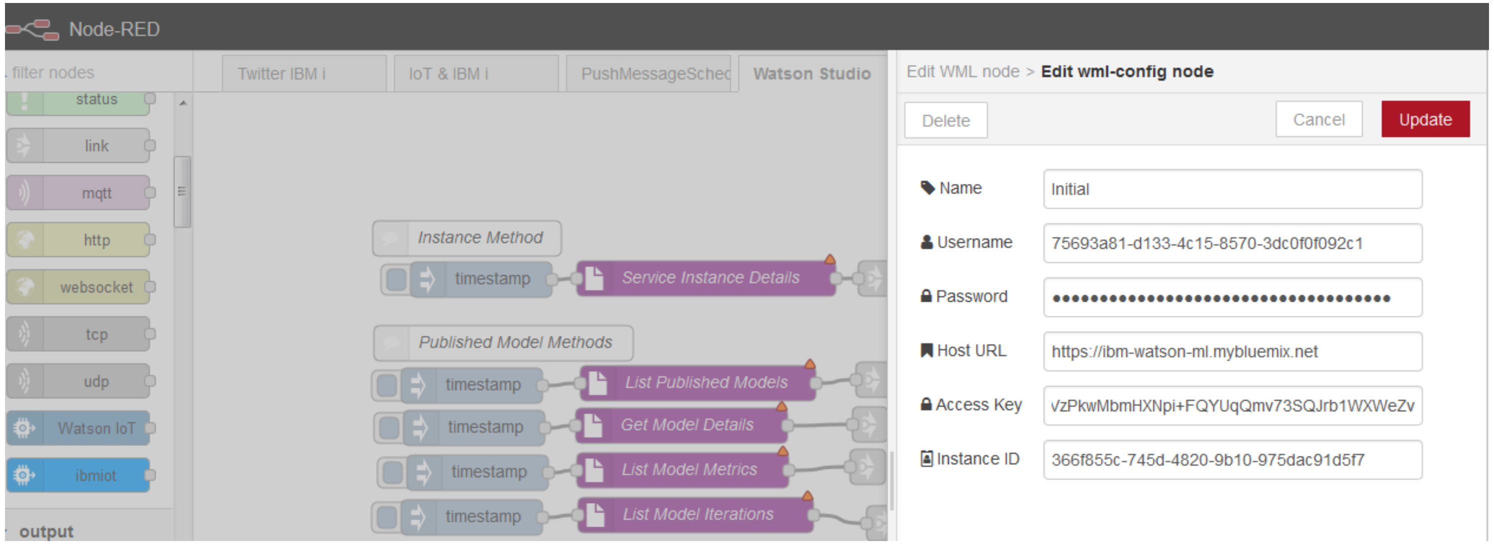
For API methods requiring Apache Spark Service instance a custom header: `X-Spark-Service-Instance` with Service Credentials must be specified. The header value is a base64 encoded string with the JSON data containing Service Credentials and Spark version.

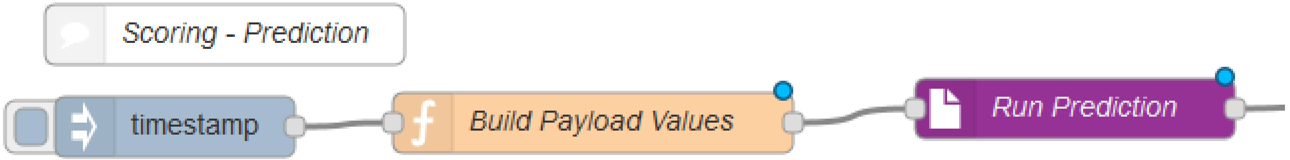
Model Integration with IBM i Apps



Model Integration with IBM i Apps

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





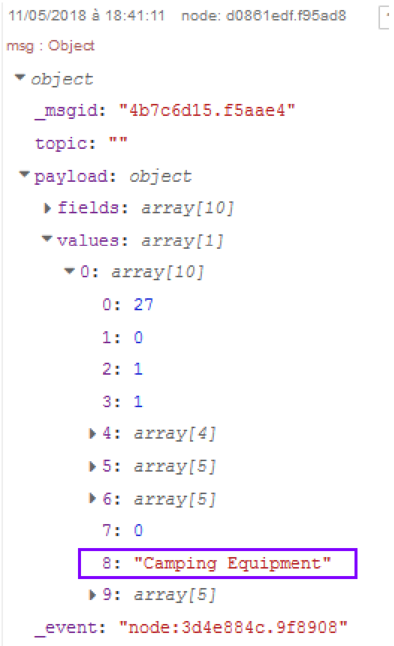


Fig. Graphical coding: let's invoke our model from Node-RED
Left: json input values + Watson ML node for invocation.
Right: API call result in json

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**

Modern Data Platform

10001111010011101 10001110001110

In Memory / Accelerated Database

Spark/ Hadoop

Competitive Discussions

1001011000111101 0001111101 10010

new

Cognitive Systems Lab

Modern Data Platform to Accelerated Workloads, PowerAI & ML/DL

PowerAI

Machine Learning / Deep Learning

HPC GPU/NVlink & CAPI Acceleration

1000111101010011001 1000111101010

Hybrid Cognitive Solutions

1010101011010 01101101010111

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…)

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
200
EXPERTS
IT Specialists, Architects, Designers, Project Managers
1500
ENGAGEMENTS
Think – Industry Showcases
Explore & Co-Create in IBM Studios
Experience – Demo, Benchmark
2500
CLIENTS
From 78 countries
COGNITIVE
BLOCKCHAIN
CLOUD
Z SYSTEM
for Business
AIX
Linux
POWER
STORAGE
IBM Client Center Montpellier

Sessions connexes à venir

- **S28 - IA sur vos données DB2 avec Watson Analytics – Lab**

  par Christophe Lalevee

- **S34 - Comment développer les applications de demain ? IBM Cloud Private, Docker etc.**

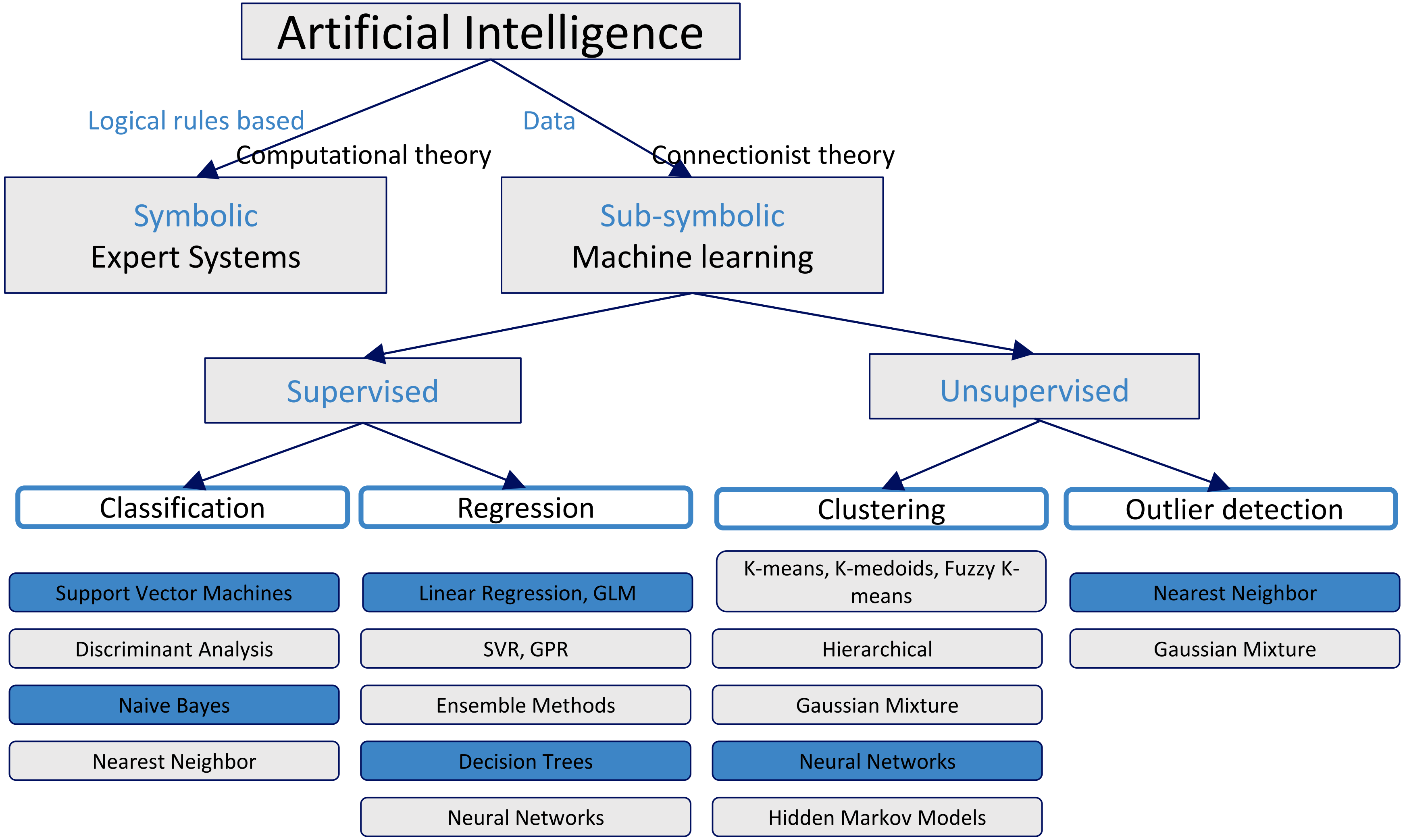
  par Benoit Marolleau

- **S45 – Prototypez un dashboard "social" avec Node-RED, Db2, Watson**

  par Benoit Marolleau

Backup Slides

The Machine Learning Tasks & algorithms



Supervised learning

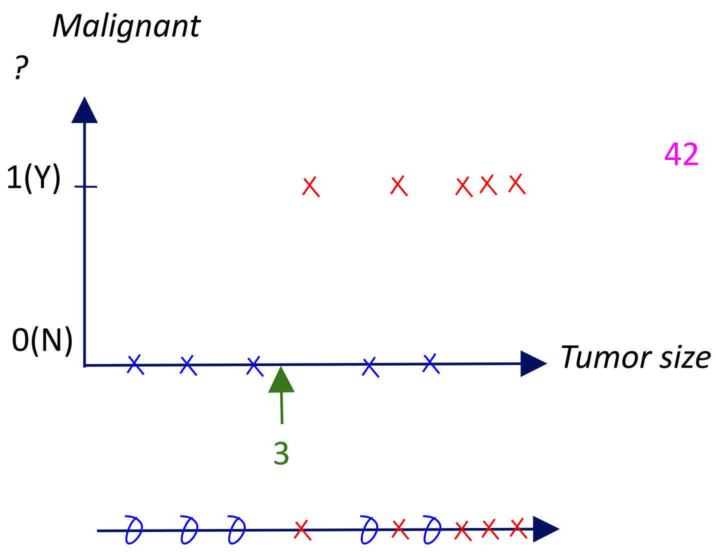
"right answers" given

Classification

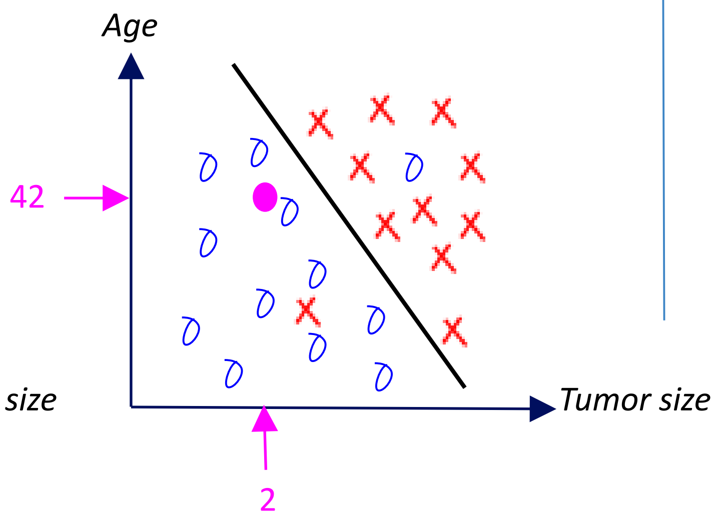
- The output variable takes class labels (discrete valued output)

Breast Cancer (malignant, benign)

1 variable



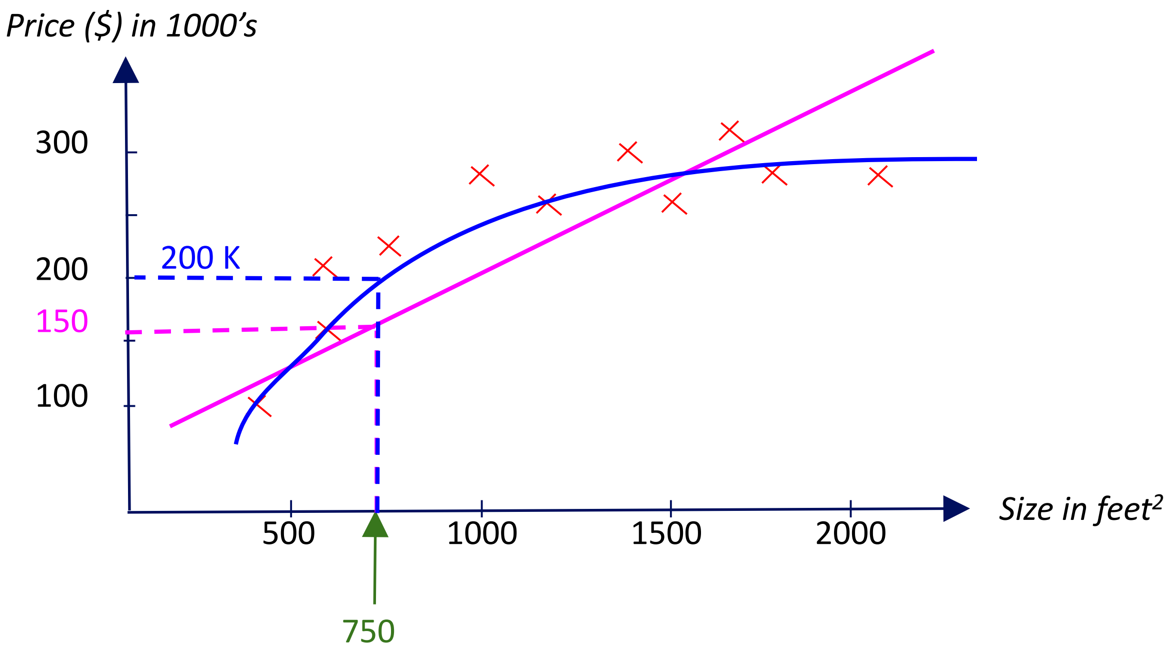
2 variables



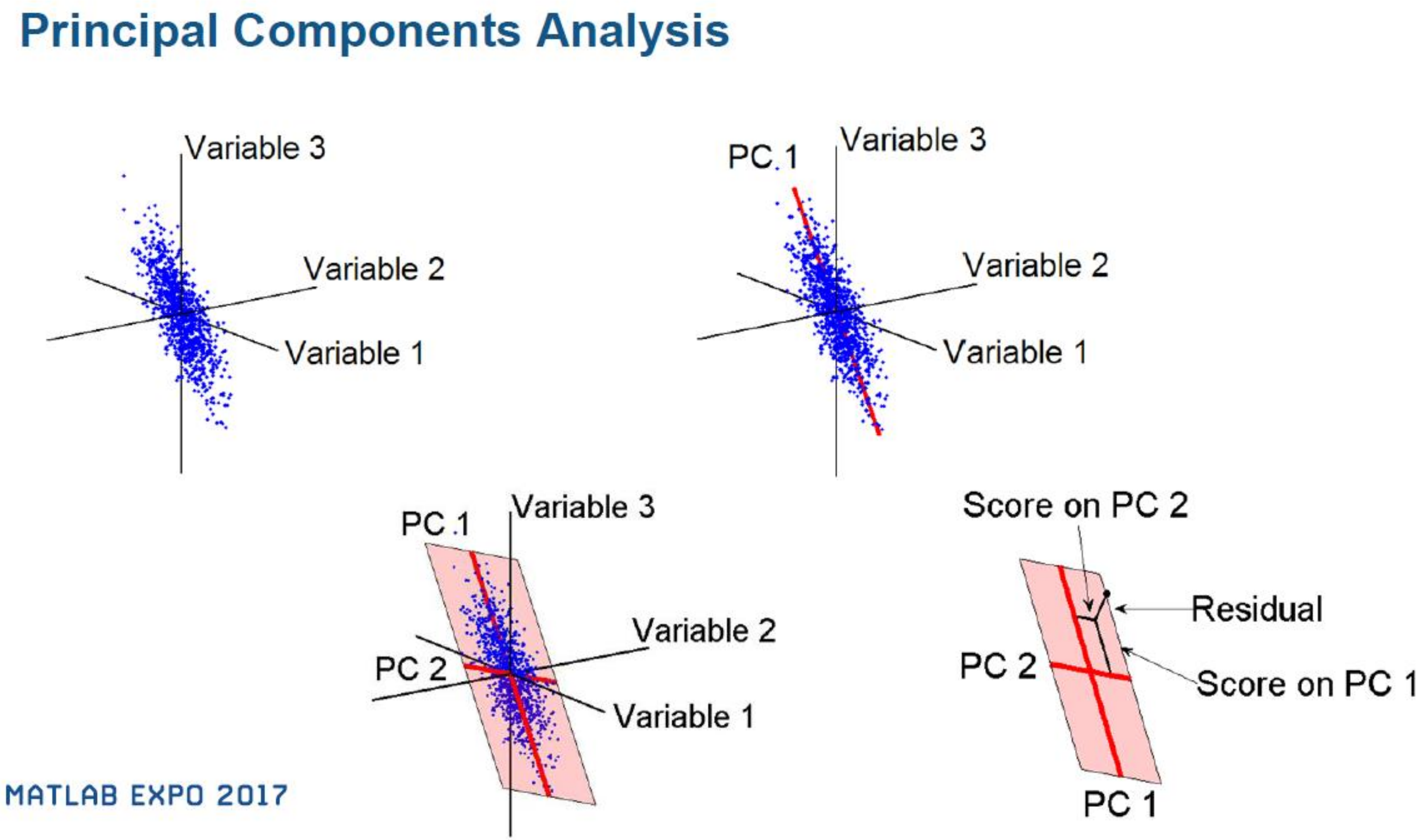
Regression

- Predict continuous valued output.

Housing price prediction



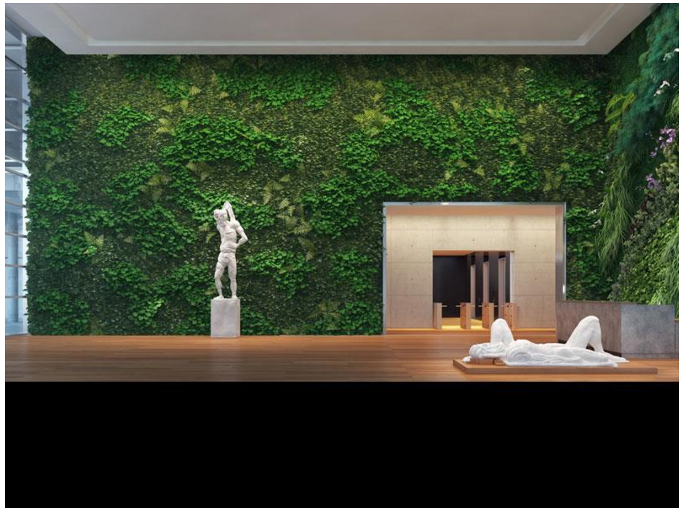
PCA for dimensionality reduction



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

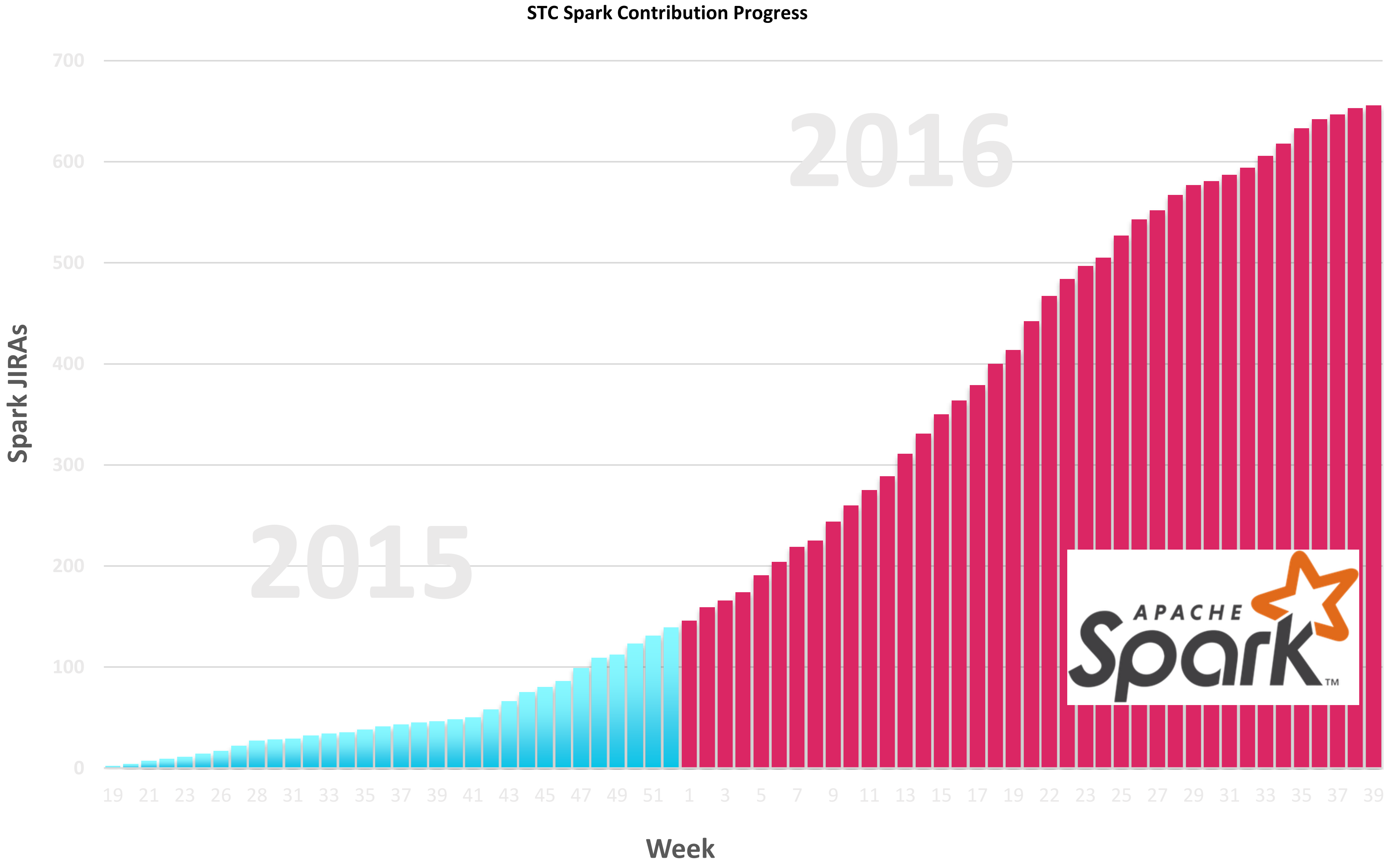
Founding member of AMPLab (and upcoming RISE Lab)

Member of R Consortium

Founding member of Scala Center

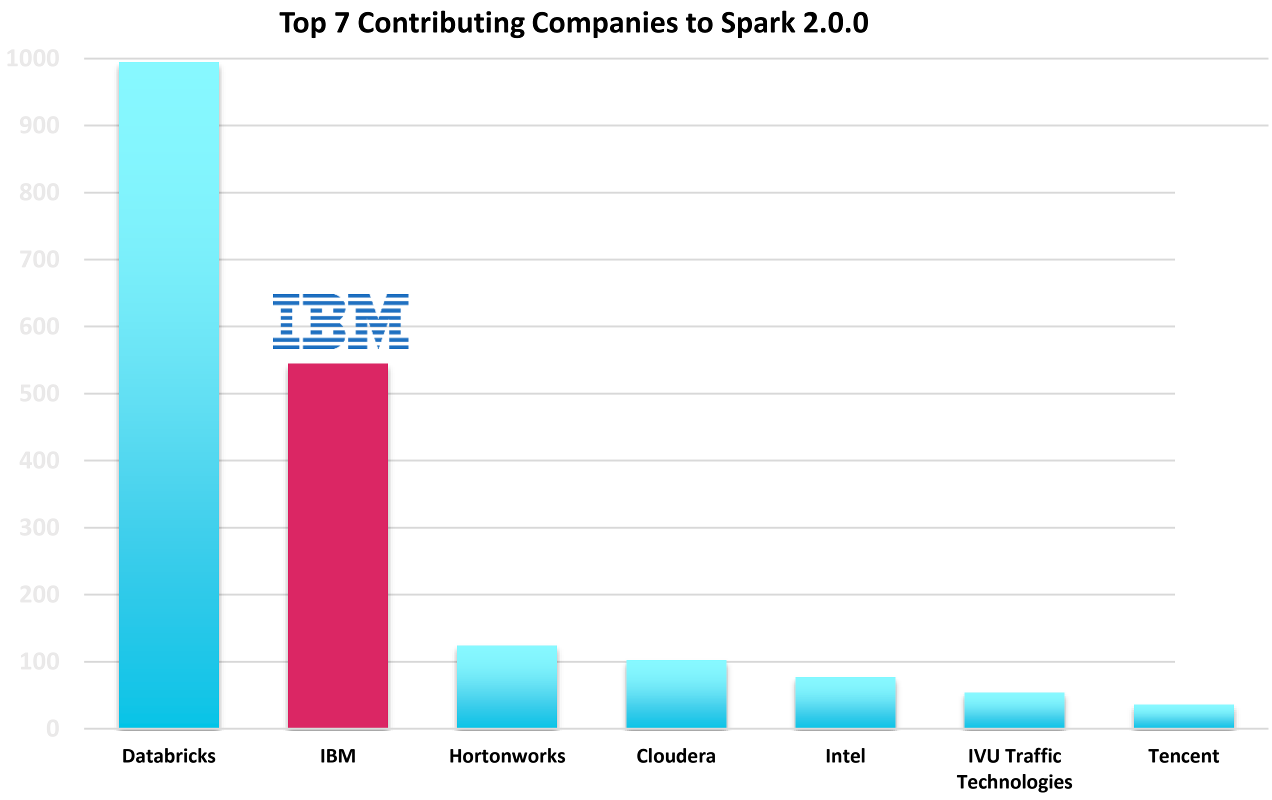
Partnerships in the ecosystem

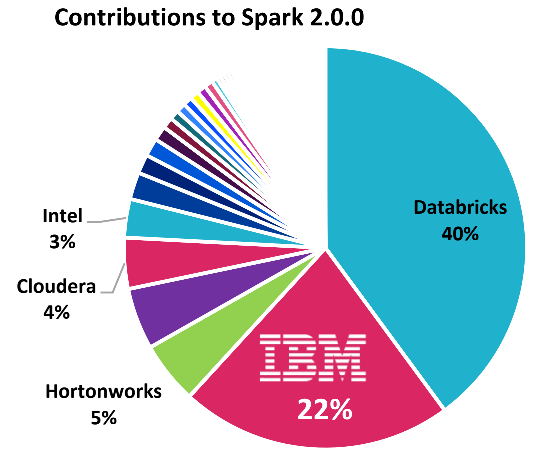
Spark Technology Center contributions have grown over 400% since start in June 2015



IBM had a significant impact on Spark 2.0

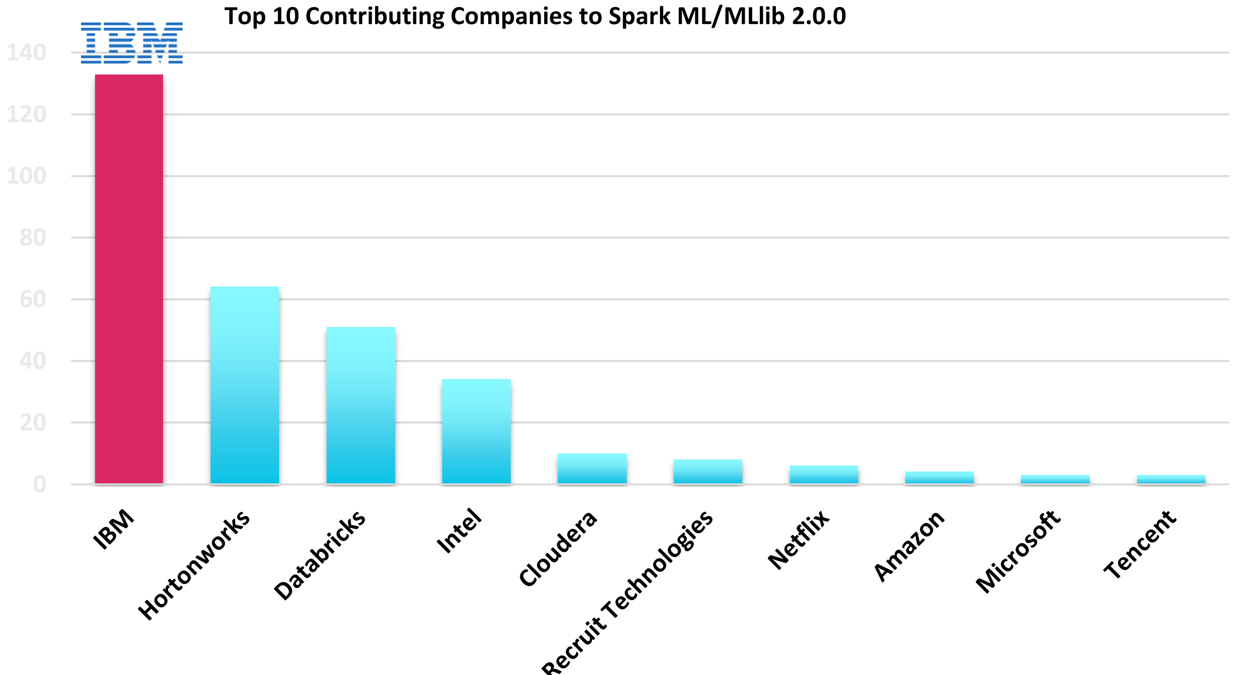
- IBM is **#2 contributor** to Apache Spark
- IBM was the leading contributor in Spark 2.0 to SparkML, PySpark, and SparkR

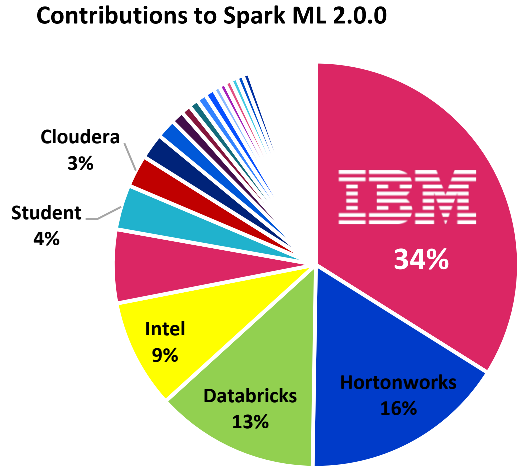




IBM impact on SparkML / MLlib 2.0

- **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:**

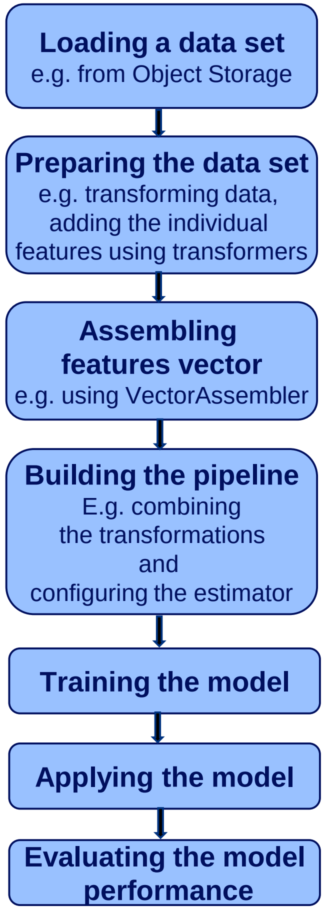
Dataframe -> Dataframe with predictions

- **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