Microsoft Azure Instructions for Untapped Energy Datathon

Mahsa Rouzbahman, PhD

Cloud Solution Architect – Advanced Analytics & AI
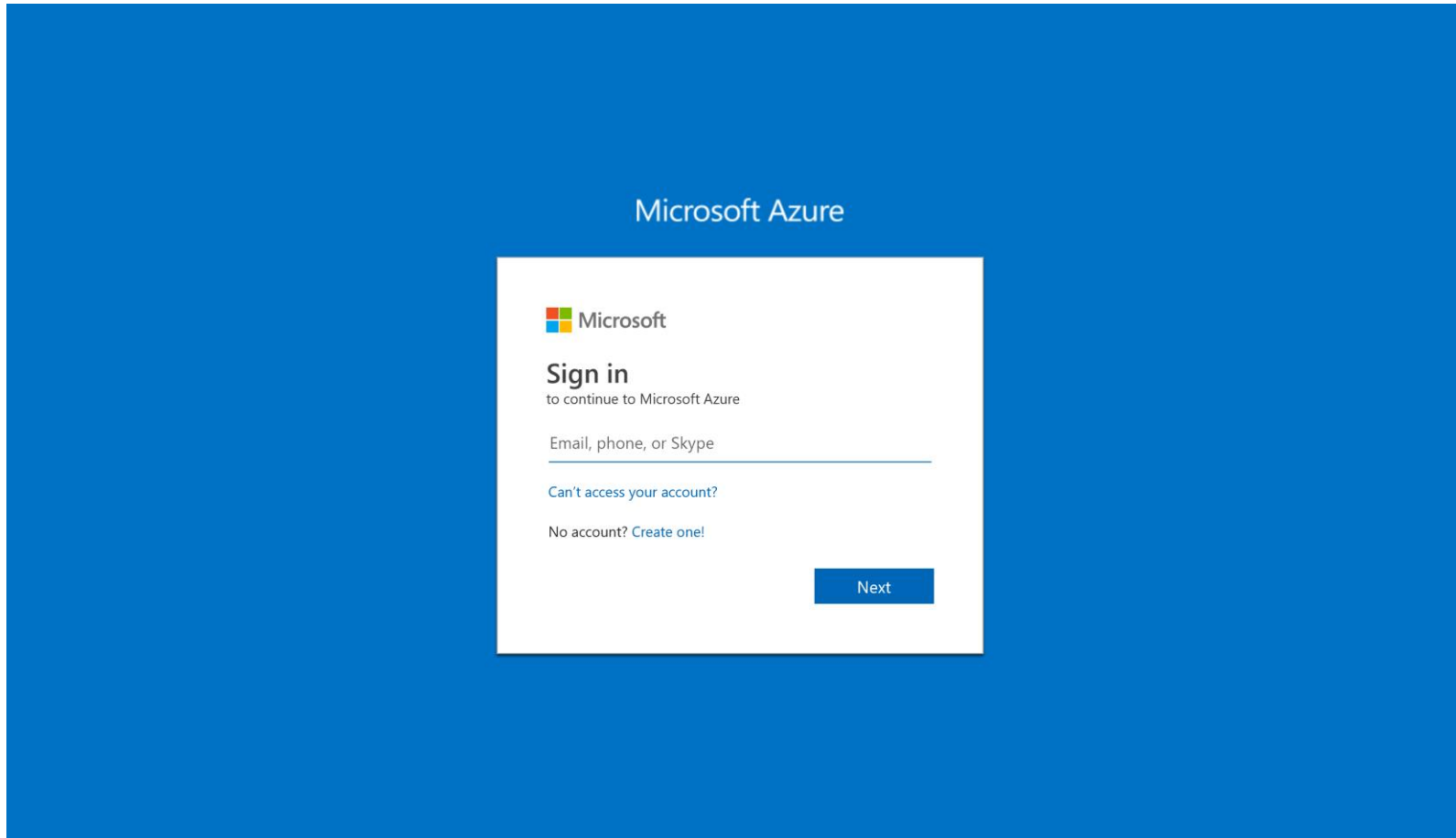
Customer Success Unit

Oct 2018

Login to portal.azure.com

A resource group (RG) was created for each team. Use the assigned RG to create all your resources.

You can create a resource in different categories (Storage, Analytics, AI, IoT, etc.)
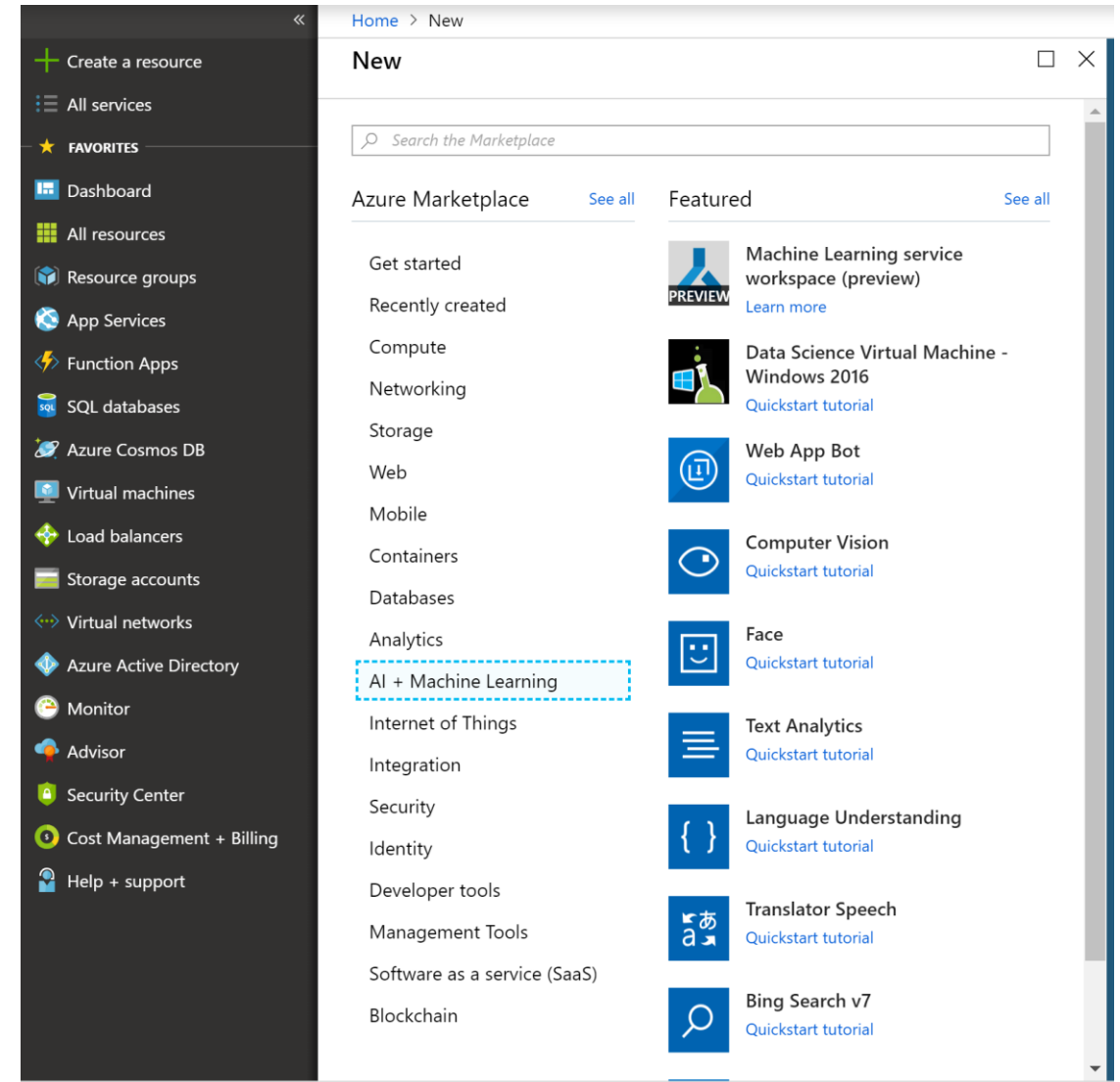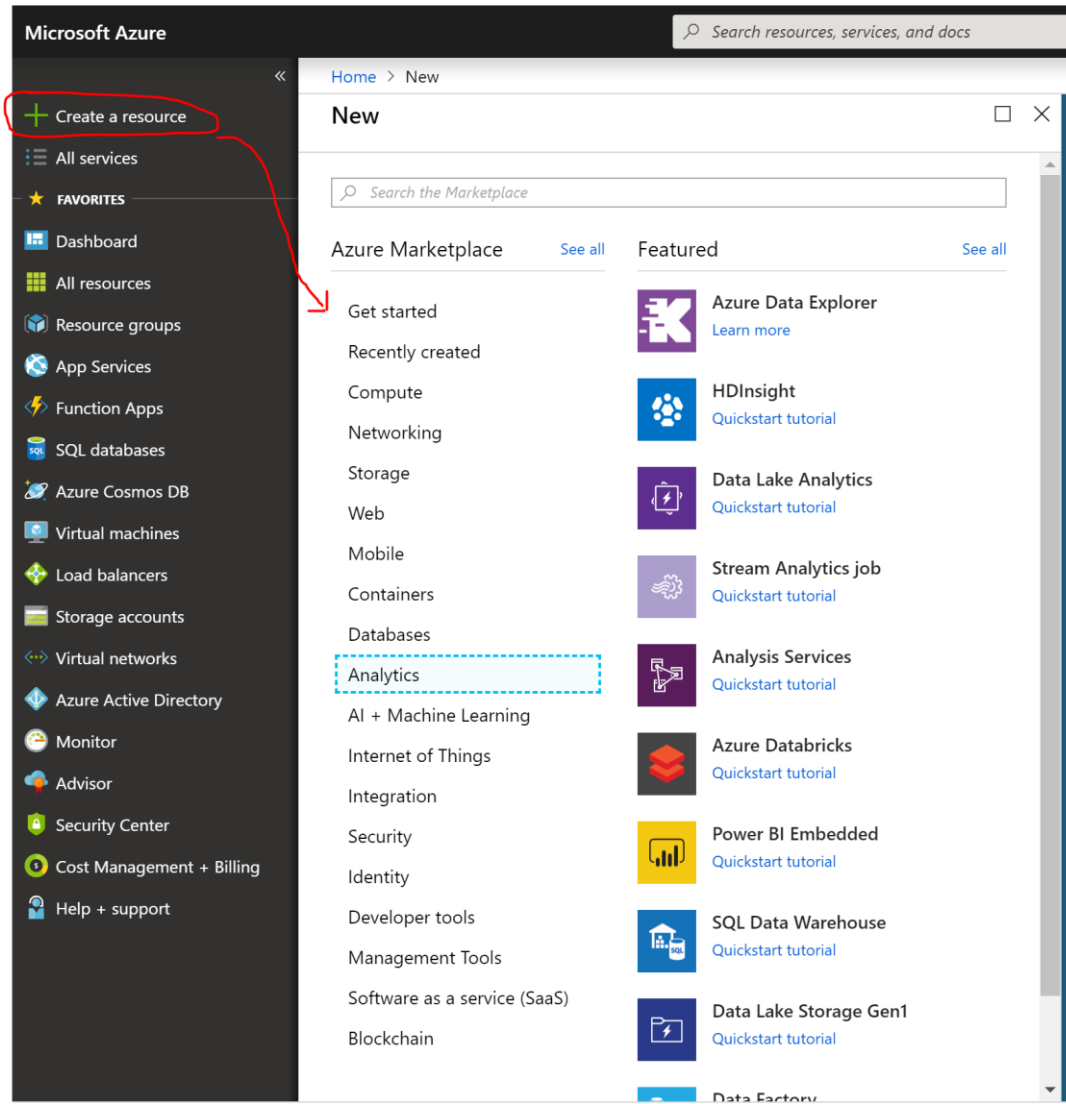
ADVANCED ANALYTICS PATTERN IN AZURE

Performing data collection/understanding, modeling and deployment

SENSORS AND IOT (UNSTRUCTURED)

LOGS, FILES AND MEDIA (UNSTRUCTURED)

BUSINESS / CUSTOM APPS (STRUCTURED)

MODEL TRAINING

- AZURE ML
- AZURE ML STUDIO
- ML SERVER
- AZURE DATABRICKS (Spark ML)
- SQL Server (In-database ML)
- DATA SCIENCE VM
- BATCH AI

LONG TERM STORAGE

- DATA LAKE STORE
- AZURE STORAGE
- COSMOS DB
- SQL DB

DATA PROCESSING

- DATA LAKE ANALYTICS
- AZURE DATABRICKS
- HDINSIGHT

ORCHESTRATION

- DATA FACTORY

TRAINED MODEL HOSTING

- AZURE CONTAINER SERVICE
- SQL Server (In-database ML)

SERVING STORAGE

- COSMOS DB
- SQL DB
- SQL DW
- AZURE ANALYSIS SERVICES

APPLICATIONS

Power BI

DASHBOARDS

Machine Learning on Azure

Sophisticated pretrained models
To simplify solution development

Vision   Speech   Language   Search

Popular frameworks
To build advanced deep learning solutions

Pytorch   TensorFlow   Keras   Onnx

Productive services
To empower data science and development teams

Azure Databricks   Azure Machine Learning   Machine Learning VMs

Powerful infrastructure
To accelerate deep learning

GPU and FPGA won't be available for the datathon event

CPU   GPU   FPGA
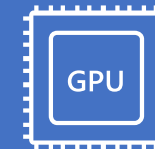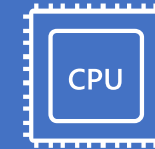
Flexible deployment
To deploy and manage models on intelligent cloud and edge

On-premises   Cloud   Edge

Possible Options to Consider for Data Analysis

1. Azure Machine Learning Studio (visual drag and drop)
2. Azure Machine Learning Service (code-first)
3. Data Science Virtual Machine
4. Azure Databricks
5. Cognitive Services
6. Time Series Insight

Azure Machine Learning Studio

Collaborative, drag-and-drop tool to build, test, and deploy predictive analytics solution

- Browser-based, no coding, graphical tool to build predictive analytics apps
- Serverless Azure service
- 99.95% SLA
- Supports multiple algorithms, data source and data formats.
- Trained models can be deployed as Request-Response Service (RRS), Batch Execution Service (BES).
- Gallery includes community contributed solutions.
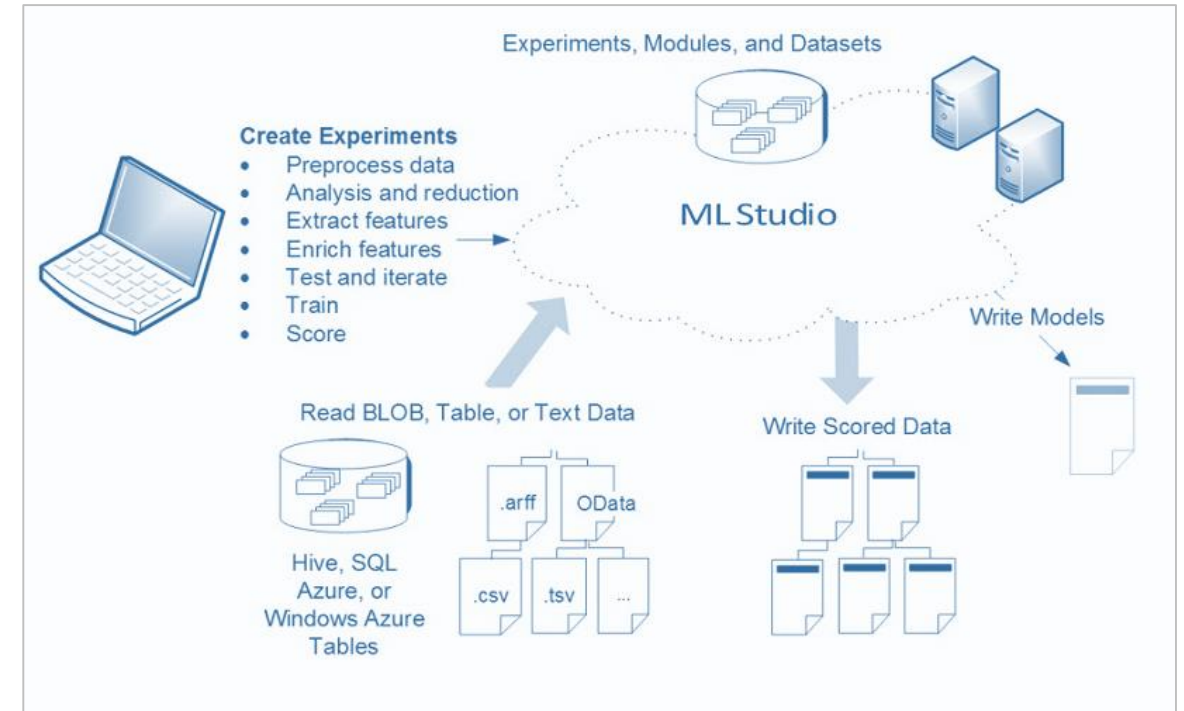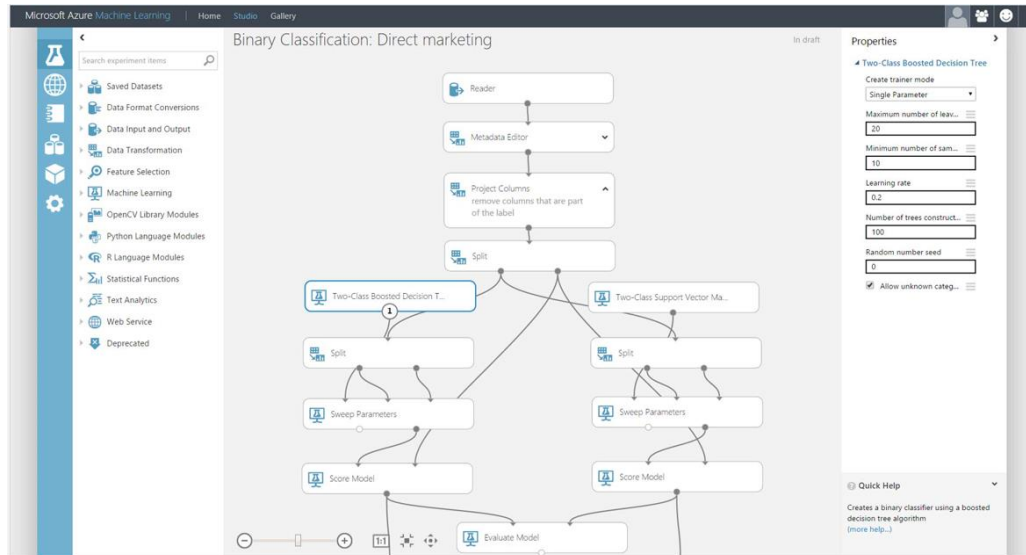
 Microsoft Azure Machine Learning Studio Capabilities Overview

You can sign in here: https://studio.azureml.net/

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.
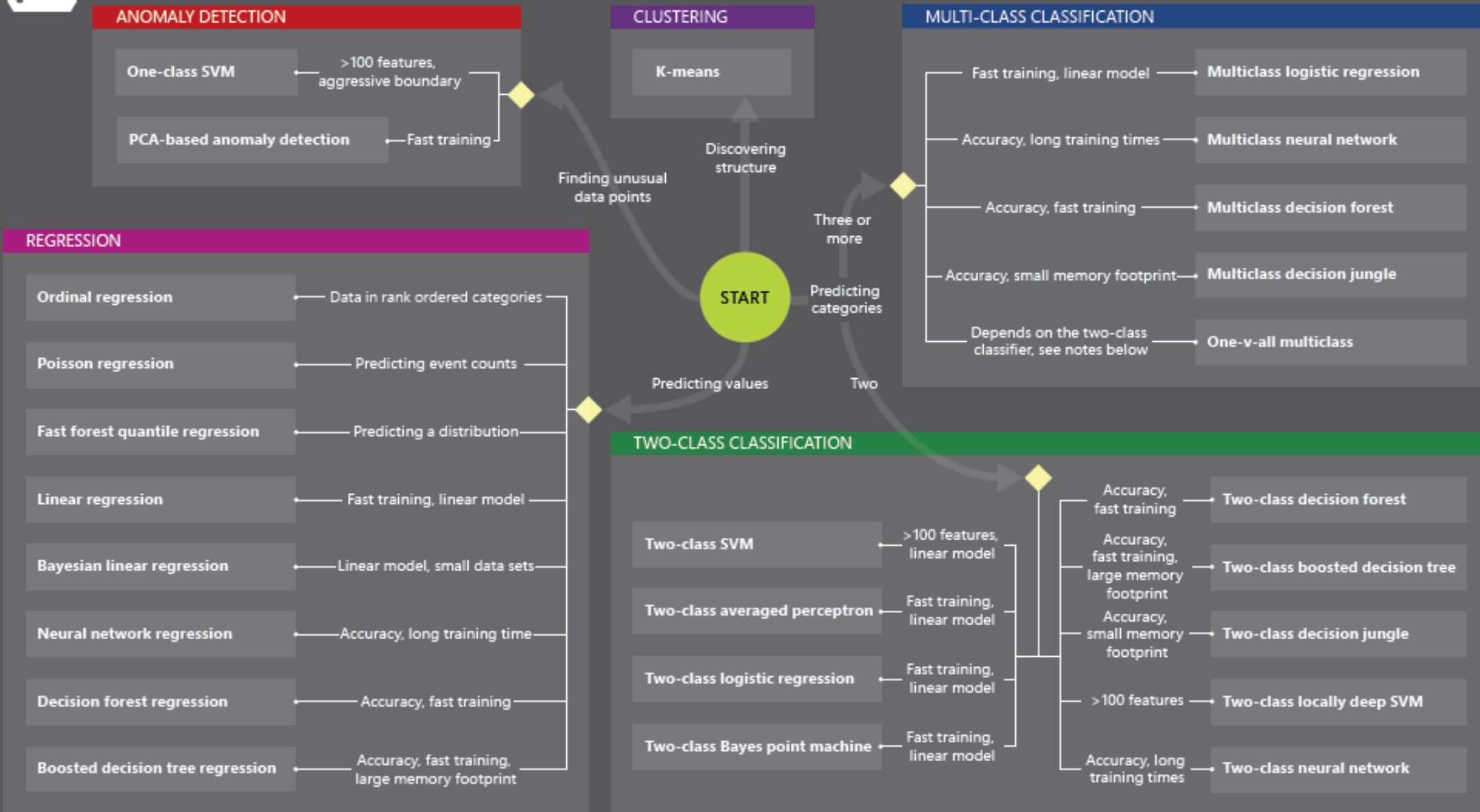
http://aka.ms/MLCheatSheet

ANOMALY DETECTION

- One-class SVM — >100 features, aggressive boundary
- PCA-based anomaly detection — Fast training

CLUSTERING

- K-means

MULTI-CLASS CLASSIFICATION

- Fast training, linear model — **Multiclass logistic regression**
- Accuracy, long training times — **Multiclass neural network**
- Accuracy, fast training — **Multiclass decision forest**
- Accuracy, small memory footprint — **Multiclass decision jungle**
- Depends on the two-class classifier, see notes below — **One-v-all multiclass**

REGRESSION

- **Ordinal regression** — Data in rank ordered categories
- **Poisson regression** — Predicting event counts
- **Fast forest quantile regression** — Predicting a distribution
- **Linear regression** — Fast training, linear model
- **Bayesian linear regression** — Linear model, small data sets
- **Neural network regression** — Accuracy, long training time
- **Decision forest regression** — Accuracy, fast training
- **Boosted decision tree regression** — Accuracy, fast training, large memory footprint

START

- Finding unusual data points
- Discovering structure
- Three or more — Predicting categories
- Two
- Predicting values

TWO-CLASS CLASSIFICATION

- **Two-class SVM** — >100 features, linear model
- **Two-class averaged perceptron** — Fast training, linear model
- **Two-class logistic regression** — Fast training, linear model
- **Two-class Bayes point machine** — Fast training, linear model

- Accuracy, fast training — **Two-class decision forest**
- Accuracy, fast training, large memory footprint — **Two-class boosted decision tree**
- Accuracy, small memory footprint — **Two-class decision jungle**
- >100 features — **Two-class locally deep SVM**
- Accuracy, long training times — **Two-class neural network**

Azure Machine Learning Service

Azure Machine Learning service provides a cloud-based environment you can use to develop, train, test, deploy, manage, and track machine learning models.
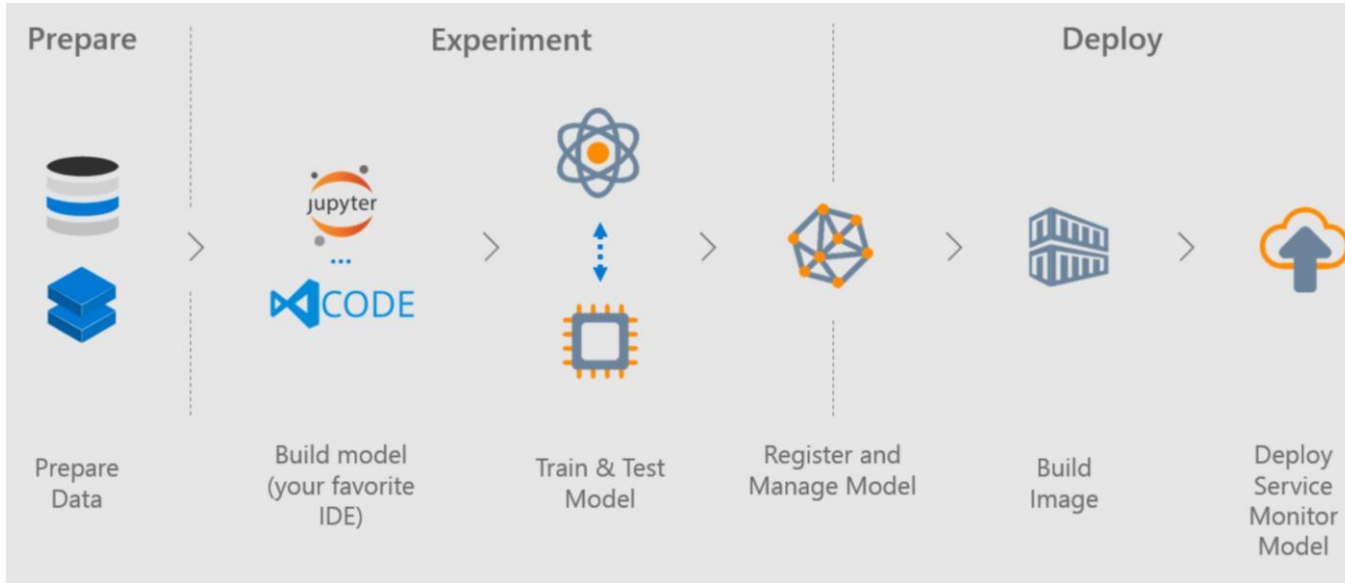
Azure Machine Learning service fully supports open-source technologies, so you can use tens of thousands of open-source Python packages with machine learning components such as TensorFlow and scikit-learn. Rich tools, such as Jupyter notebooks or the Visual Studio Code Tools for AI, make it easy to interactively explore data, transform it, and then develop, and test models. Azure Machine Learning service also includes features that automate model generation and tuning to help you create models with ease, efficiency, and accuracy.
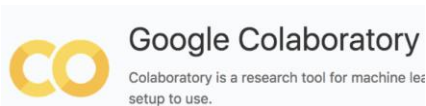
Documentation:
https://docs.microsoft.com/en-us/azure/machine-learning/service/overview-what-is-azure-ml

Use your favorite IDE

CODE   Visual Studio   jupyter   PyCharm

Microsoft Azure Notebooks   Google Colaboratory
Colaboratory is a research tool for machine lea
setup to use.

Augment the tools that you know and love with the Azure ML SDK

Build and train models directly from your environment

Start from prebuilt packages or your own code

+

Azure Machine Learning

Build and train models locally, with Remote VMs or massive Batch AI clusters

Track experiments for reproducibility and auditing needs. Identify and promote your best models

Deploy and manage your models using containers to run them anywhere

Monitor your models performance from a single location

Data Science Virtual Machine?

- Custom VM image on Azure Marketplace

- Contains a set of data science, Azure tools/SDKs

- All pre-configured and ready to use

- Pay for cloud hardware usage only. No separate software charges!

- Pointers to gallery, samples, documentation

- Windows and Linux Versions

- Up and running quickly

- Please start with a small VM. You can always change the size if necessary, later.

Provision a Data Science Virtual Machine on Azure

https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/provision-vm

Azure Databricks

- Introducing Azure Databricks: a quick video
  - https://databricks.com/introducing-azure-databricks
- Azure Databricks Documentation
  - https://docs.microsoft.com/en-us/azure/azure-databricks/
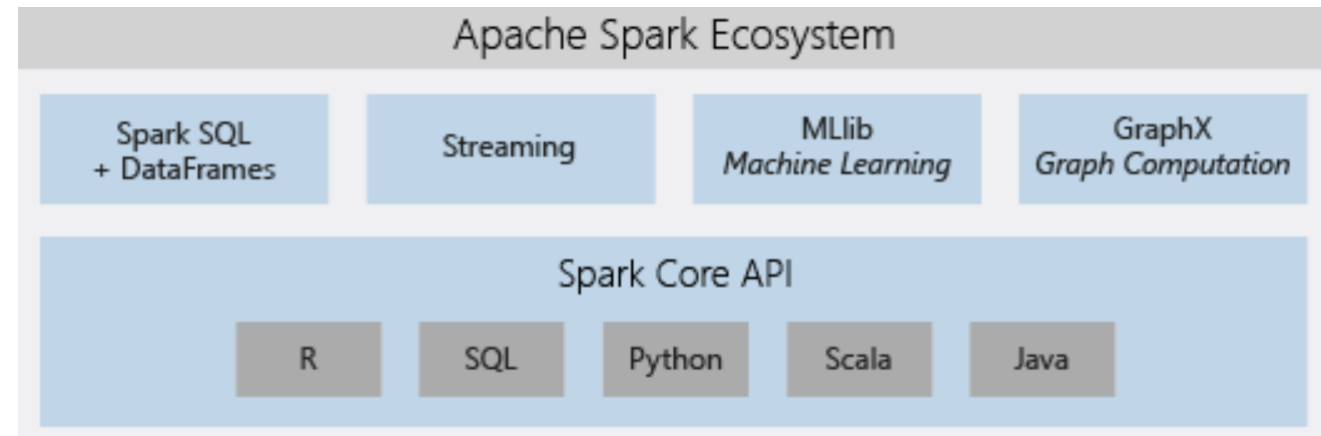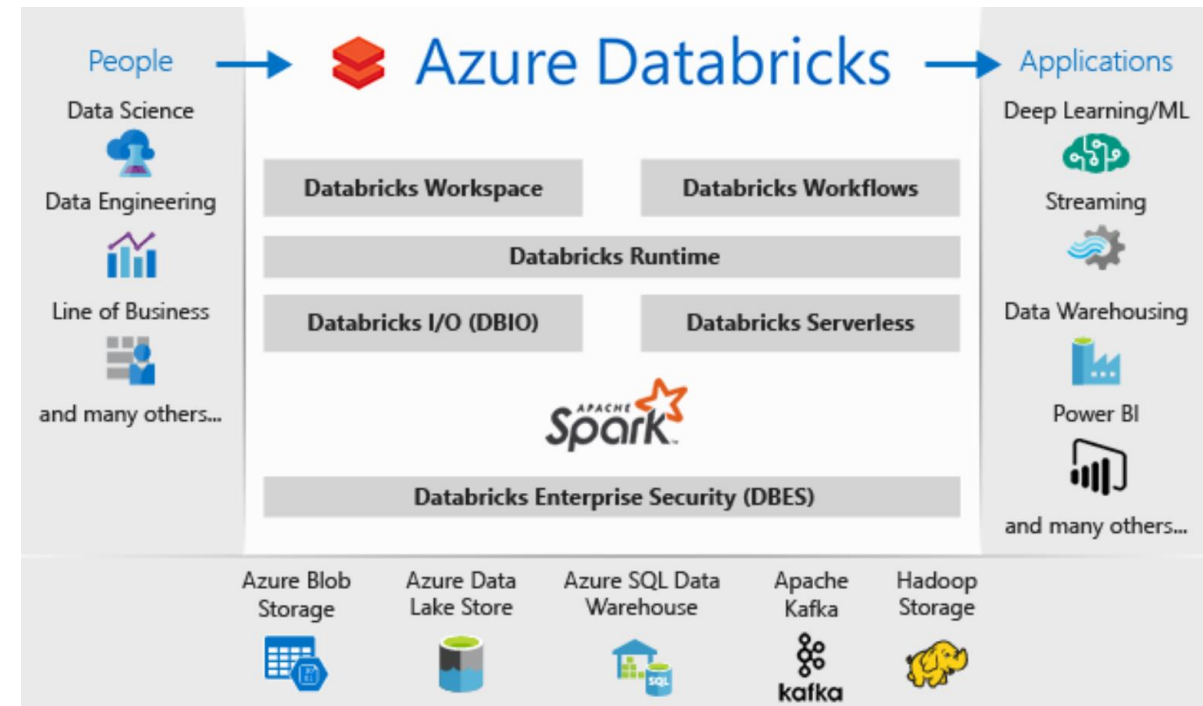
Azure Databricks Tutorial

- Azure Databricks Workshop:
    https://github.com/azeltov/adb_workshop

Workspace

| ADB_WORKSHOP ⌄	workshops
📁 labs ▽	📄 00 Introduce the Business Problem
📁 labs-answers ▽	📄 01 Databricks Overview
📁 workshops ▽	📄 02 Mounting Storage
	📄 03 Reading Data
	📄 04 Streaming Lab with FileSystem
	📄 05 Power Bi
	📄 06 Spark MLlib - ALS
	📁 helpers
	📁 libs
	📄 OPTIONAL - 04 Streaming with Event Hubs

Cognitive Services

- A group of services and APIs that allow developers to build artificial intelligence applications
  - Easily add intelligent features to applications
  - Vision | Speech | Language | Knowledge | Search
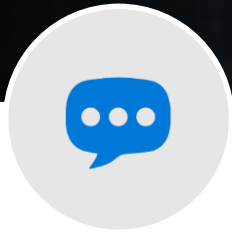- Built on the Microsoft Cognitive Toolkit

https://azure.microsoft.com/en-ca/services/cognitive-services/directory/lang/
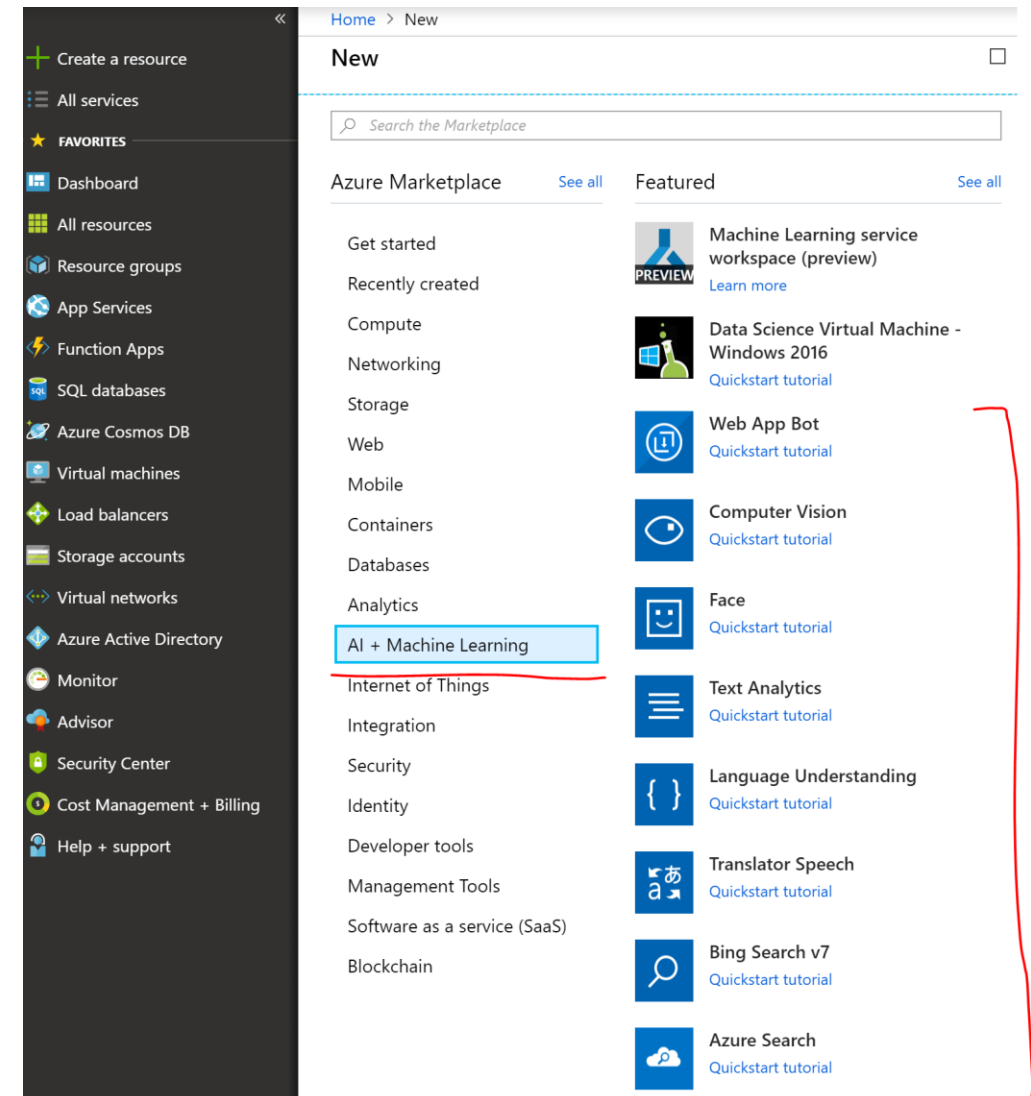
Microsoft Cognitive Services

Give your apps a human side

Vision
Computer Vision
Content Moderator
Emotion
Face
Video
Video Indexer
Custom Vision
Service

Speech
Bing Speech
Custom Speech
Service
Speaker Recognition

Language
Bing Spell Check
Language Understanding
Linguistic Analysis
Translator Text & Speech
Web Language Model
Text Analytics

Knowledge
Academic
Knowledge
Entity Linking
Knowledge
Exploration
Recommendations
QnA Maker
Custom Decision
Service

Search
Bing Autosuggest
Bing Image Search
Bing News Search
Bing Video Search
Bing Web Search
Bing Custom Search

Labs
Project Prague (gesture)
Cuzco (events)
Johannesburg (routing)
Nanjing (Isochrones)
Abu Dhabi (distance matrix)
Wollongong (location)
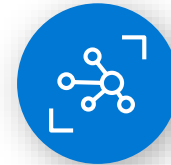
Azure **Time Series Insights**

Documentation:
https://docs.microsoft.com/en-us/azure/time-series-insights/time-series-quickstart

- IoT scale time-series data store
- Schema-less store, just send data
- Easy IoT Hub connection
- Store, query, and visualize billions of events
- Get near real-time insights in seconds
- Build apps using Time Series Insights APIs