Build Data & AI Applications in .NET

Luis Quintanilla

@ljquintanilla

Hello ☺

- Content Developer, ML.NET & Azure Machine Learning
- Twitter: @ljquintanilla
- GitHub: lqdev
- Blog: https://luisquintanilla.me
- Email: lqdev@outlook.com

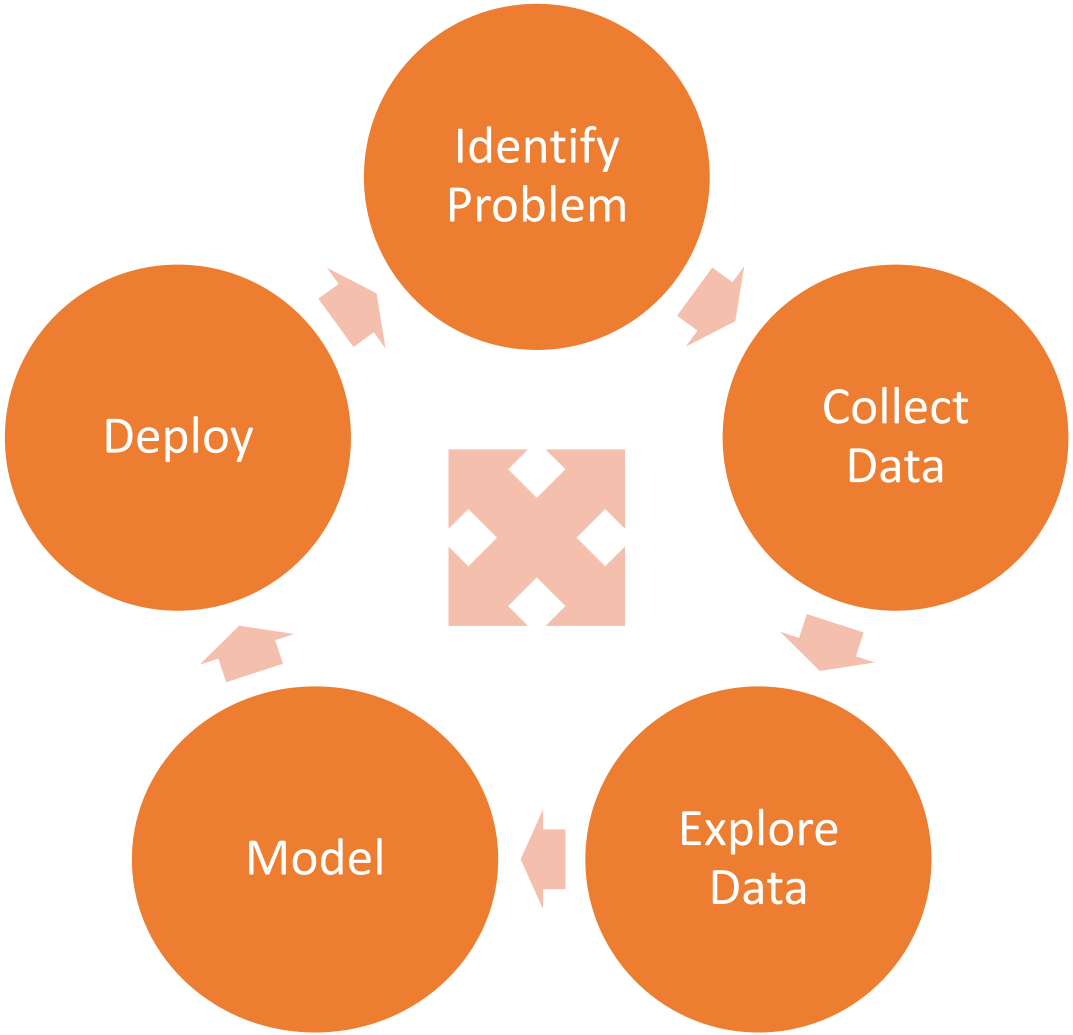
Agenda

- Data Science / ML Workflows & Ecosystem
- .NET Ecosystem
- Apache Spark
- Jupyter Notebooks
- Machine Learning

Code & Slides

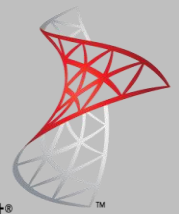
bit.ly/dotnetdataml042020

Data Science & ML Workflows



Data Science & ML Ecosystem

Data



Microsoft® SQL Server®

mongoDB.

Compute

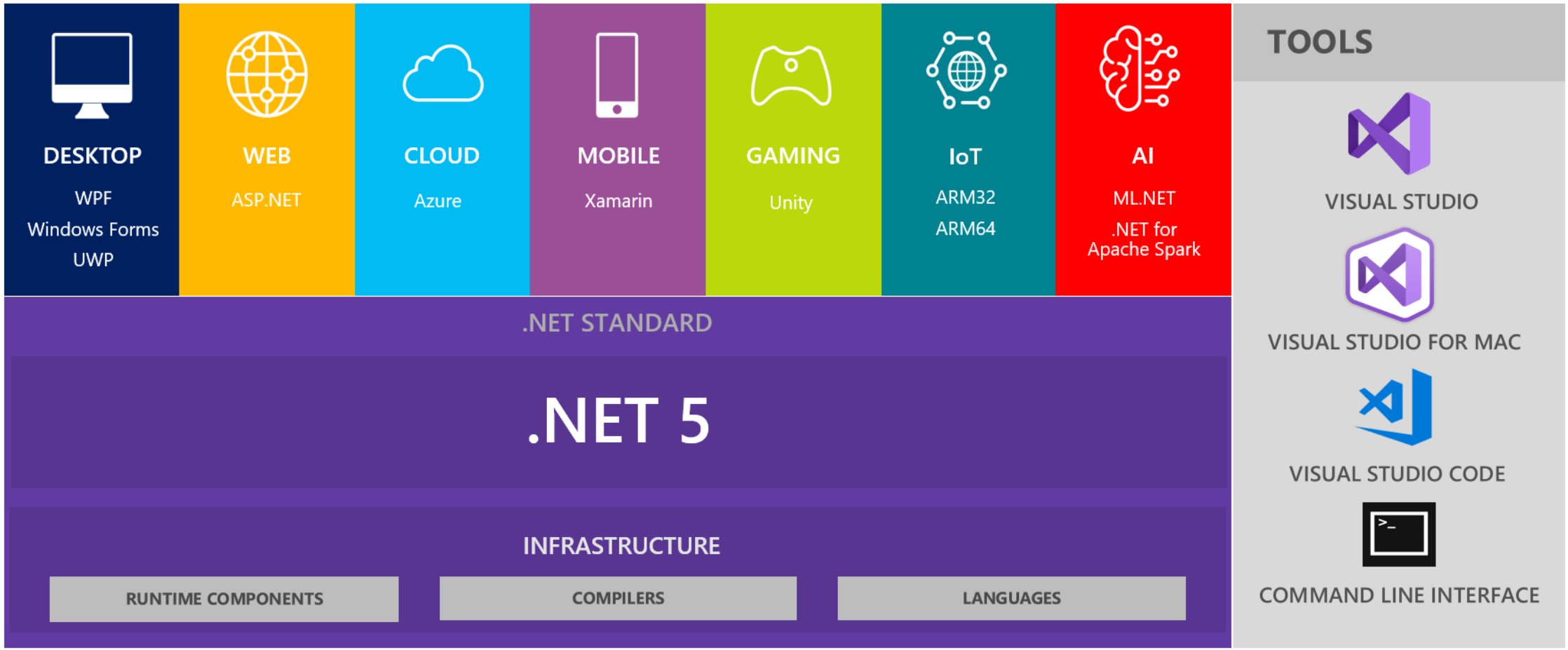
Deployment

Tooling

ANACONDA®

Models

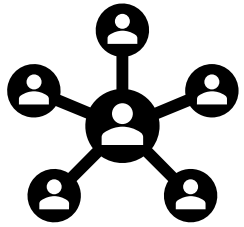
.NET – A unified platform



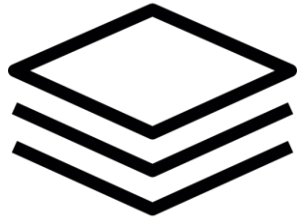
Apache Spark

Open Source

Cross Platform

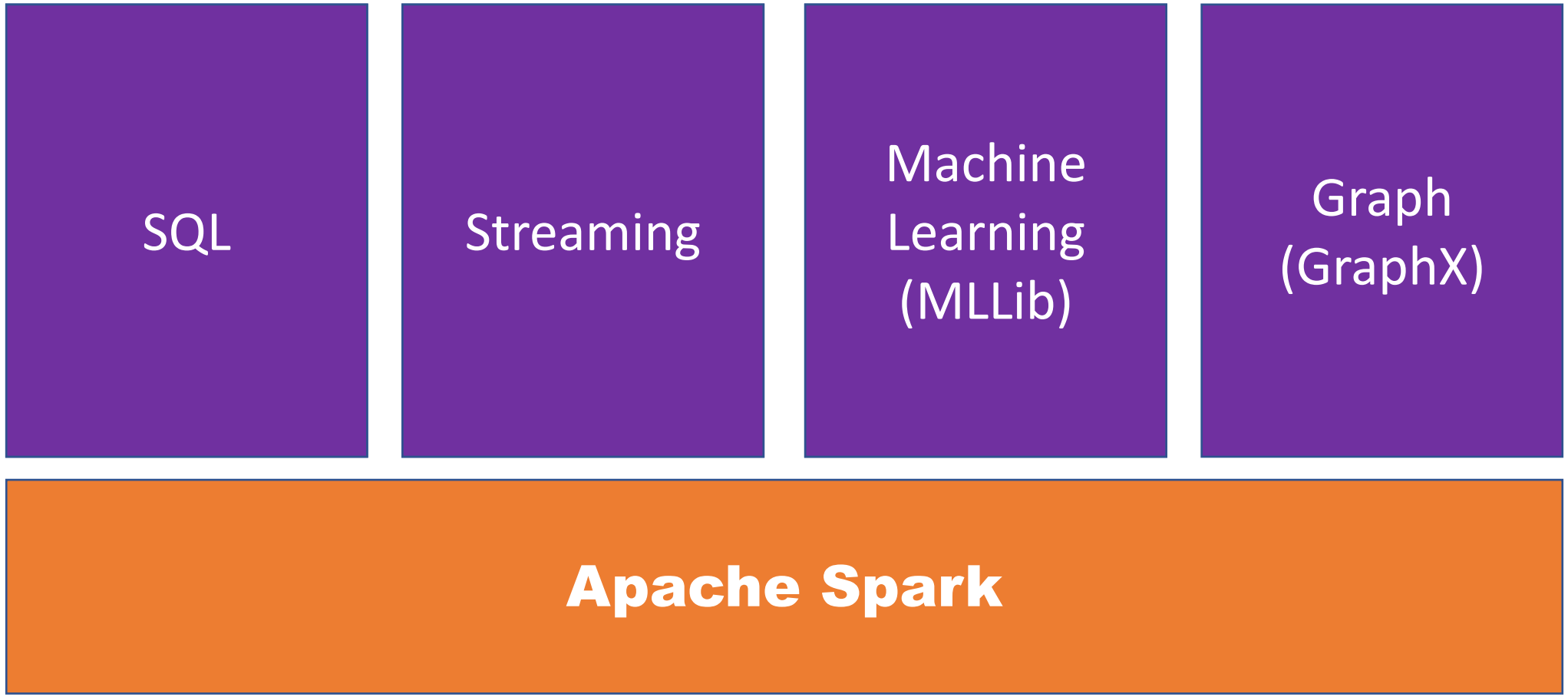


Distributed*

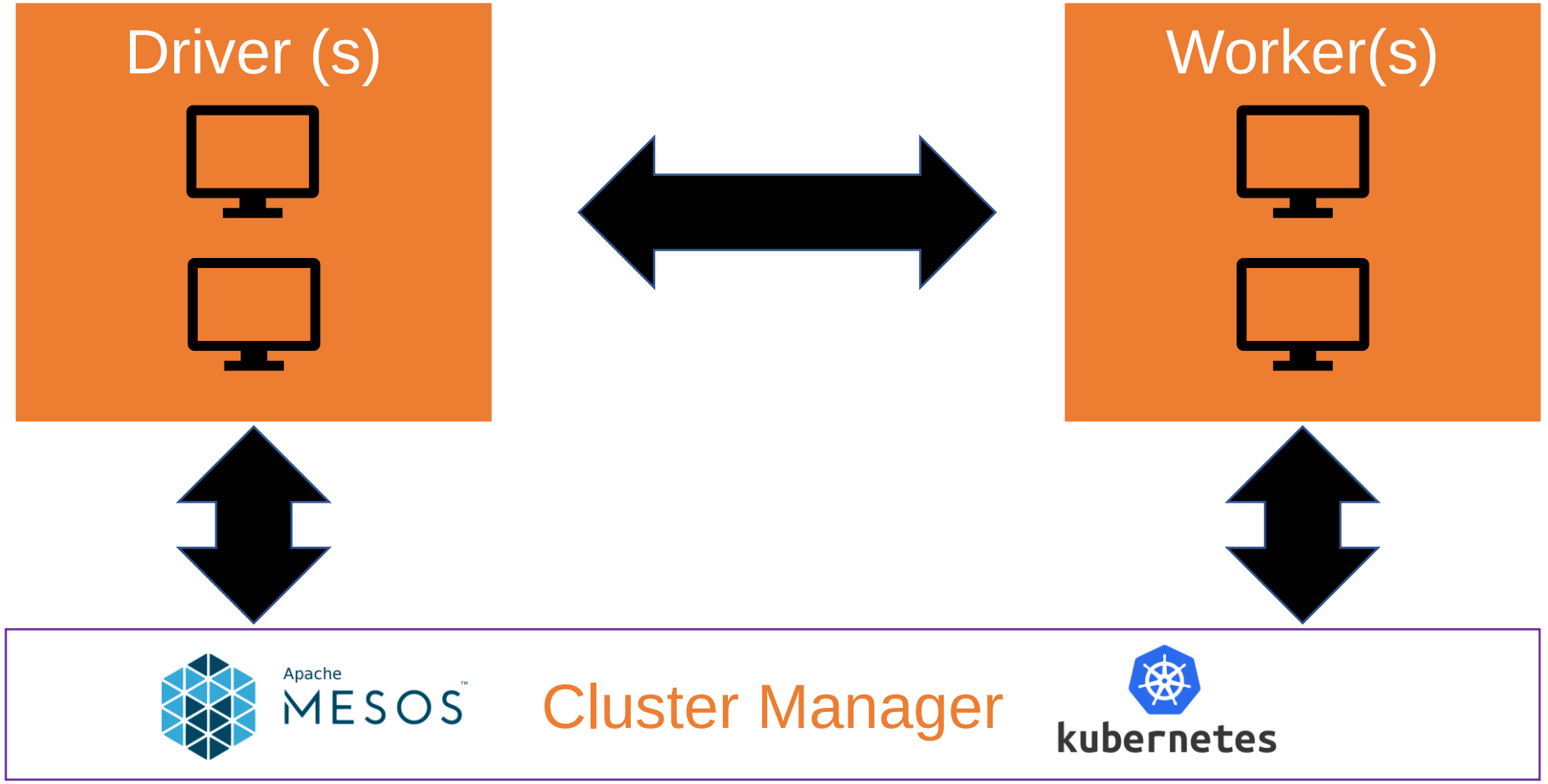


Extensible

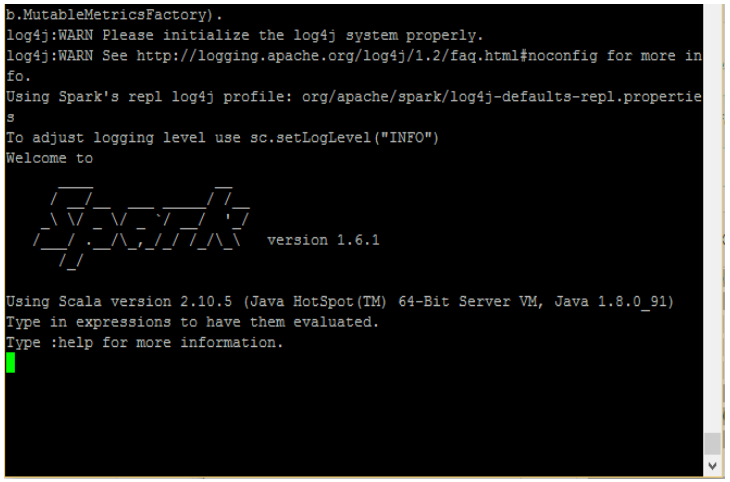
Apache Spark Architecture

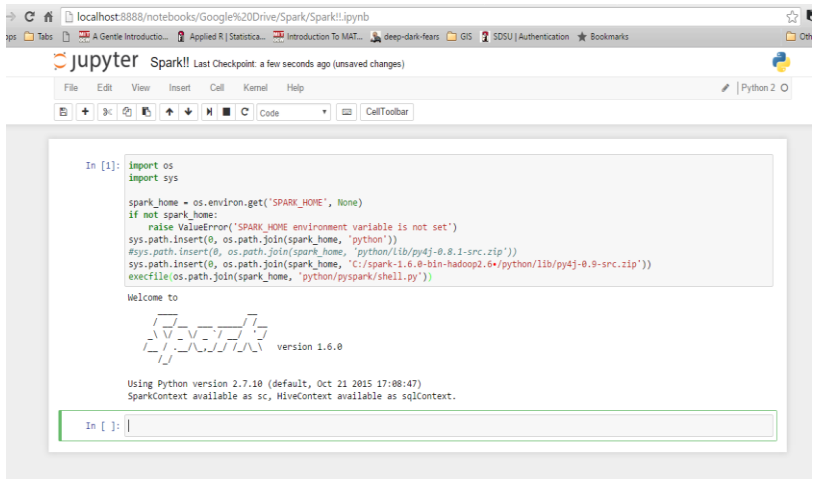


Apache Spark Execution



Interacting with Apache Spark





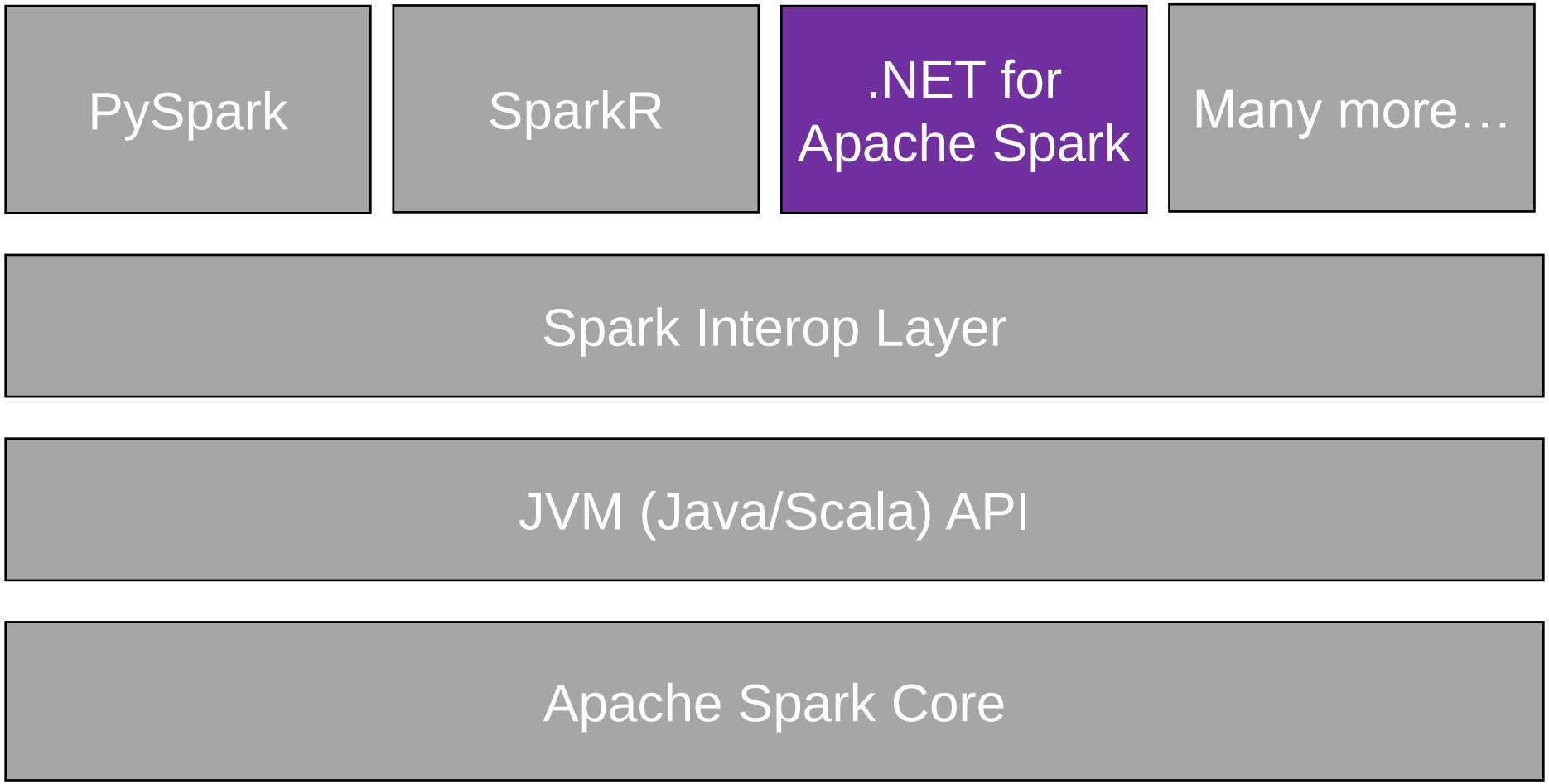

```scala
// Create SparkSession
var sc =
    SparkSession
        .Builder()
        .AppName("Restaurant_Inspections_ETL")
        .GetOrCreate();

// Load data
DataFrame df =
    sc
    .Read()
    .Option("header", "true")
    .Option("inferSchema", "true")
    .Csv("Data/NYC-Restaurant-Inspections.csv");
```

Shell

Notebooks

Traditional Apps

Extending Apache Spark

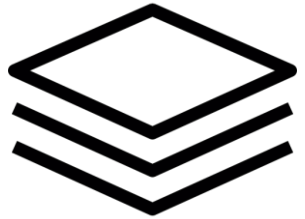


.NET for Apache Spark

Open Source

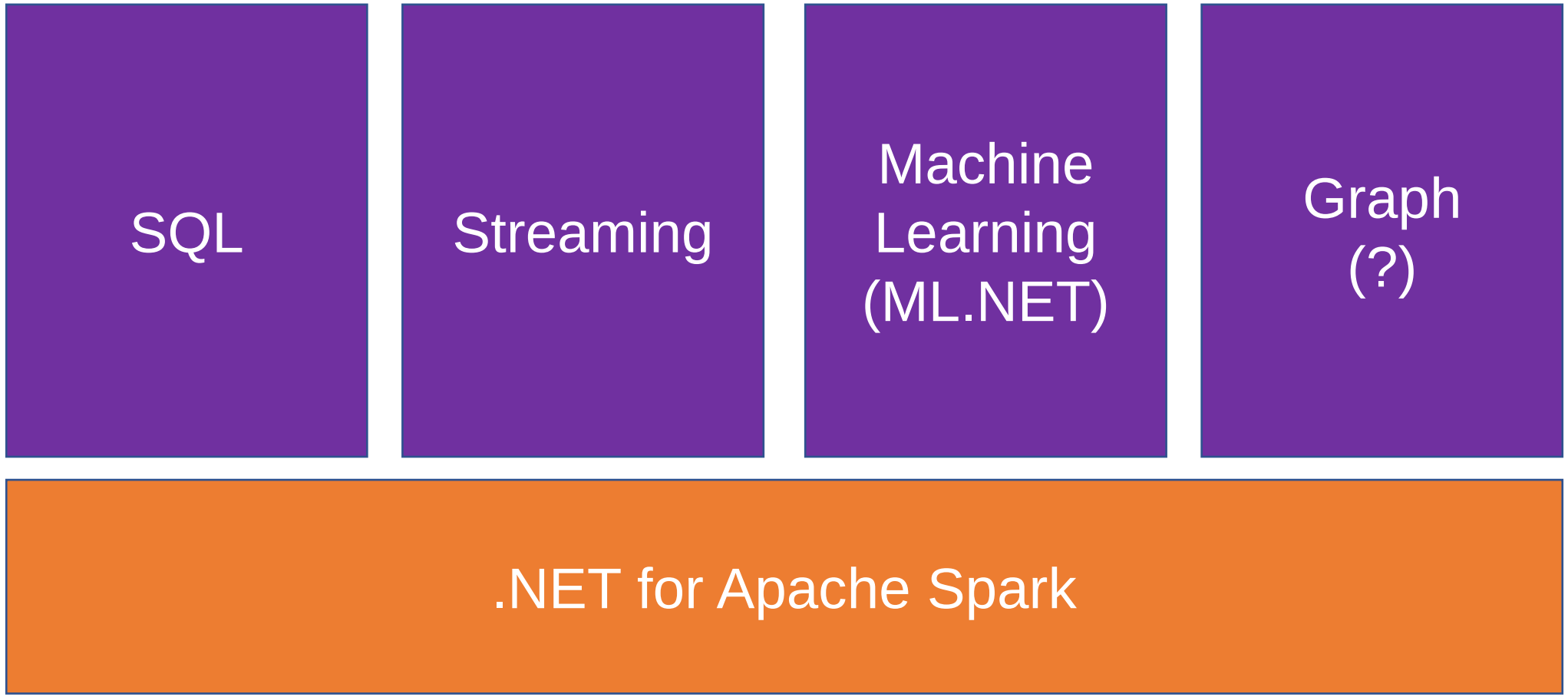
Cross Platform

.NET Standard



Extensible

.NET for Apache Spark Architecture

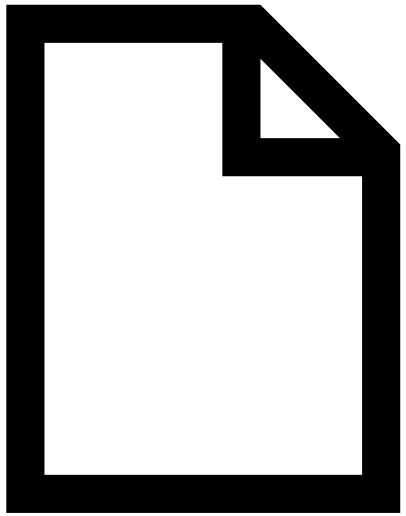


Demo
.NET for Apache Spark

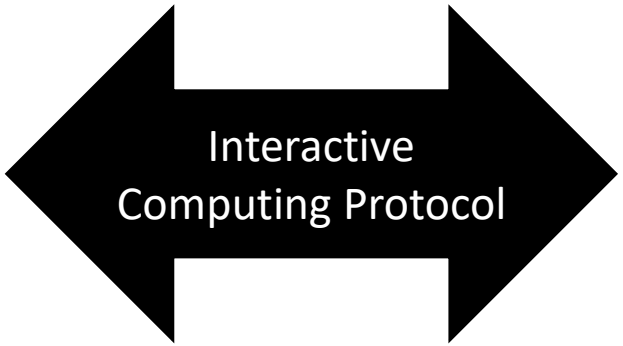
Jupyter Notebooks

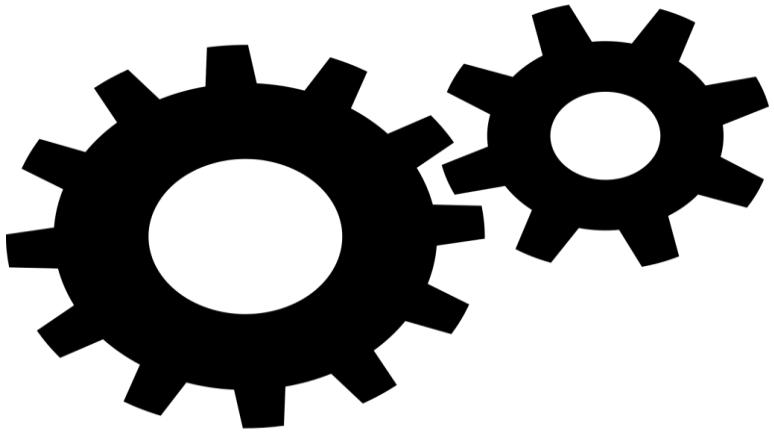
- Open Source Interactive Computing Environment
- 40+ supported programming languages
- Extension Support

Jupyter Notebooks Internals



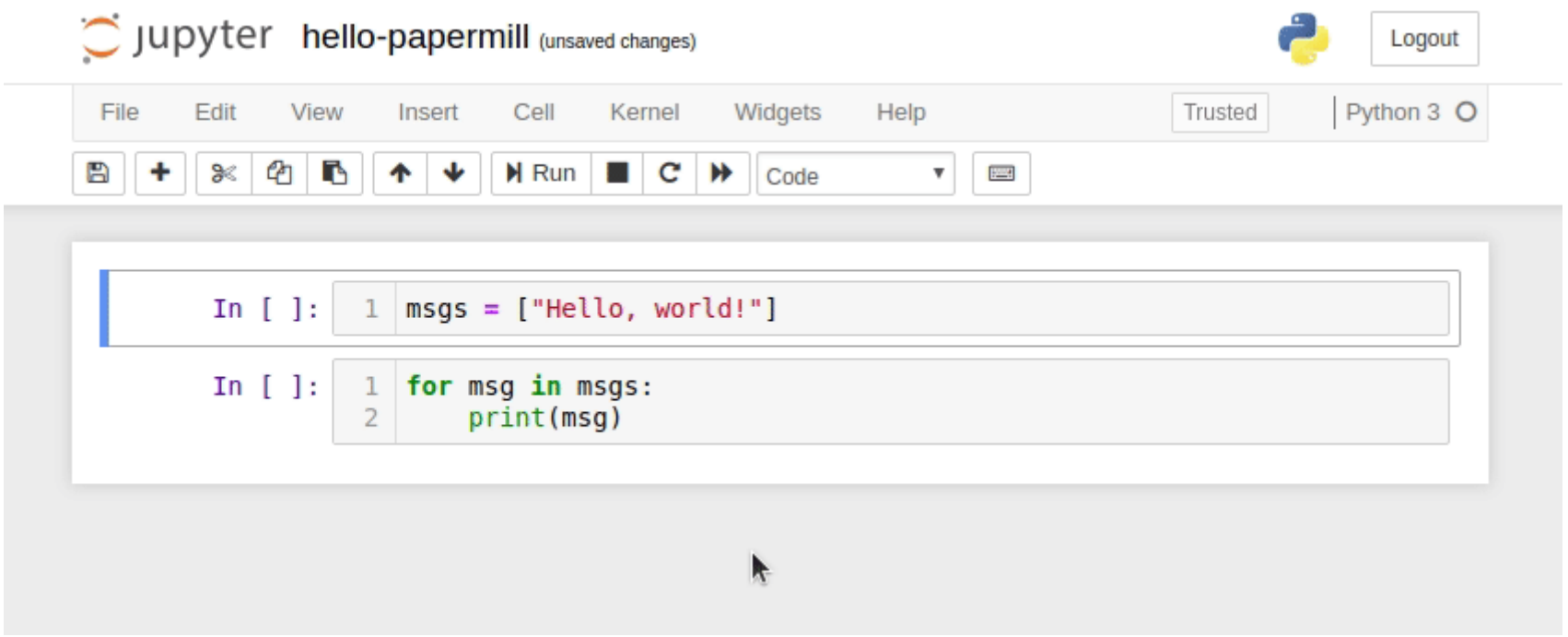
Notebook
Document
Format





Kernel

Extensions




```
$ papermill local/input.ipynb s3://bkt/output.ipynb -p alpha 0.6 -p l1_ratio 0.1
```

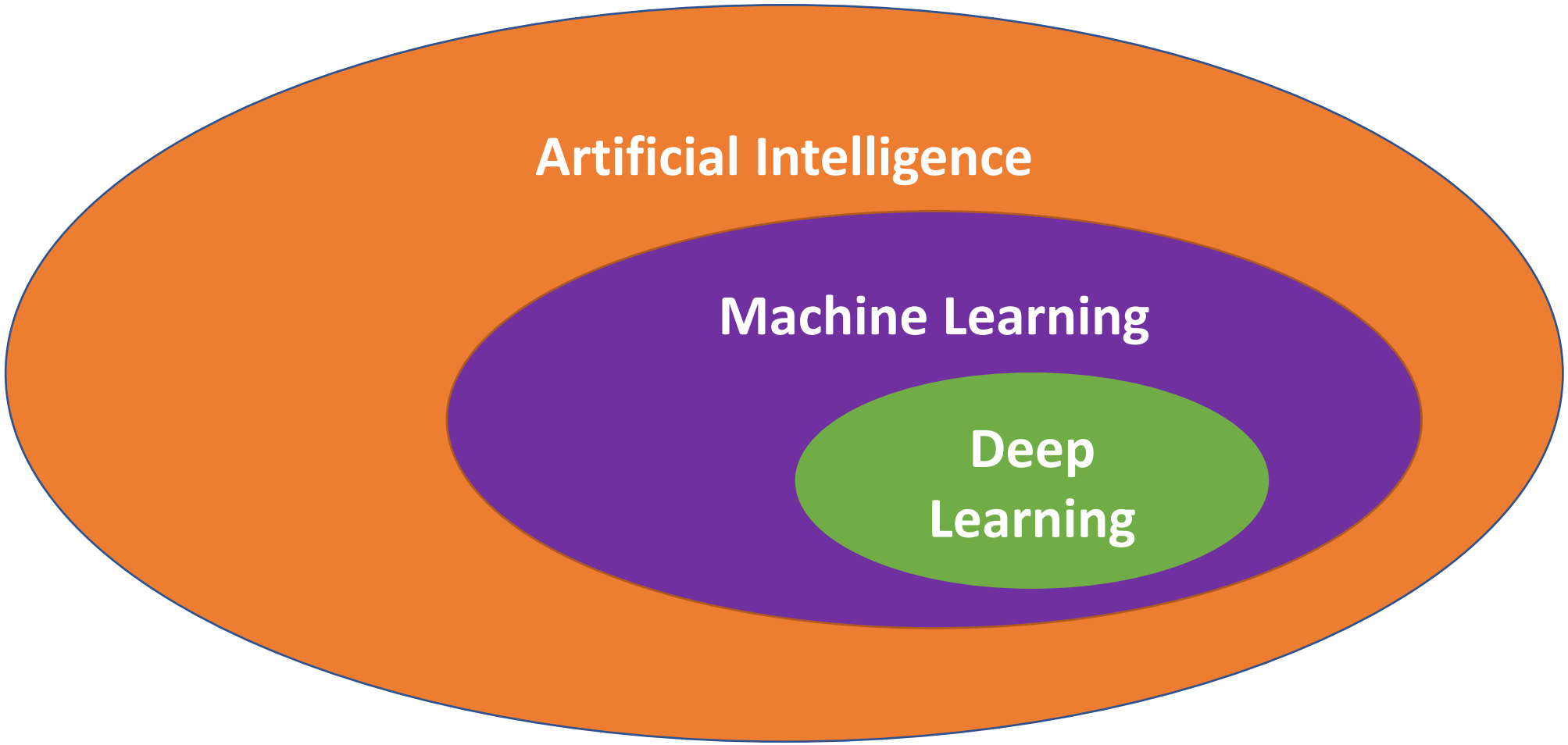
Source: https://github.com/nteract/papermill

.NET Interactive

- Open Source
- .NET Standard
- Interactive programming environment
- C#, F#, PowerShell Kernels for Jupyter Notebooks

Demo
.NET Interactive

AI vs ML vs DL



Machine Learning Tasks

Supervised Learning

Regression

What is the price of a home in NYC?

Classification

Is this a dog or cat?

Unsupervised Learning

Clustering

Customer segments in a database

Classification Example

Training Data

Species	Is Independent	Class
Canine	False	Dog
Feline	True	Cat
Feline	True	Cat
Canine	False	Dog
Canine	True	Dog

Features (input)

Label (output)

New Data

Species	Is Independent
Canine	False

Prediction

Class
Dog

What is a model?

$f(x)$

True

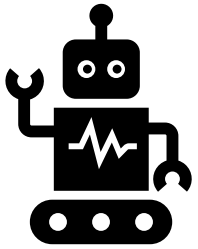
False

Input

Model

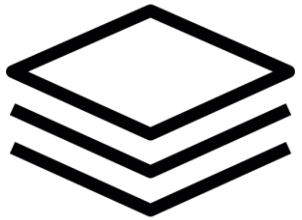
Output

ML.NET



Framework for Machine Learning

.NET Standard



Proven & Extensible

Open Source

Cross Platform

ML.NET - Framework

Transforms	Trainers	Misc	Extensions
• Missing Values • Feature Selection • Normalization	• SVM • K-Means • Boosted Trees • Image Classification	• Data Loaders • Evaluators	• TensorFlow • ONNX

Build Models with ML.NET

CLI

Model Builder

Automated ML

ML.NET API

Jupyter Notebooks

Demo
Train Time Series Model

Next Steps

- Use cloud resources
- Deploy
- MLOps

Takeaways

- .NET ecosystem for data analytics and machine learning is nascent but there is a lot of potential
- You can build a wide range of applications with ML.NET
- Using data from these applications, .NET Interactive and .NET for Apache Spark can be leveraged for descriptive data analytics
- Predictive analytics models can be built using ML.NET and deployed to one of the many .NET deployment targets

Questions

Resources

- https://github.com/luisquintanilla/Presentations/tree/master/dotnetdataml042020
- https://docs.microsoft.com/en-us/dotnet/machine-learning/
- https://docs.microsoft.com/en-us/dotnet/spark/
- https://github.com/dotnet/interactive
- https://github.com/dotnet/machinelearning-samples
- https://github.com/dotnet/machinelearning
- https://github.com/dotnet/spark