

Build Data & AI Applications in .NET

Luis Quintanilla

@ljquintanilla

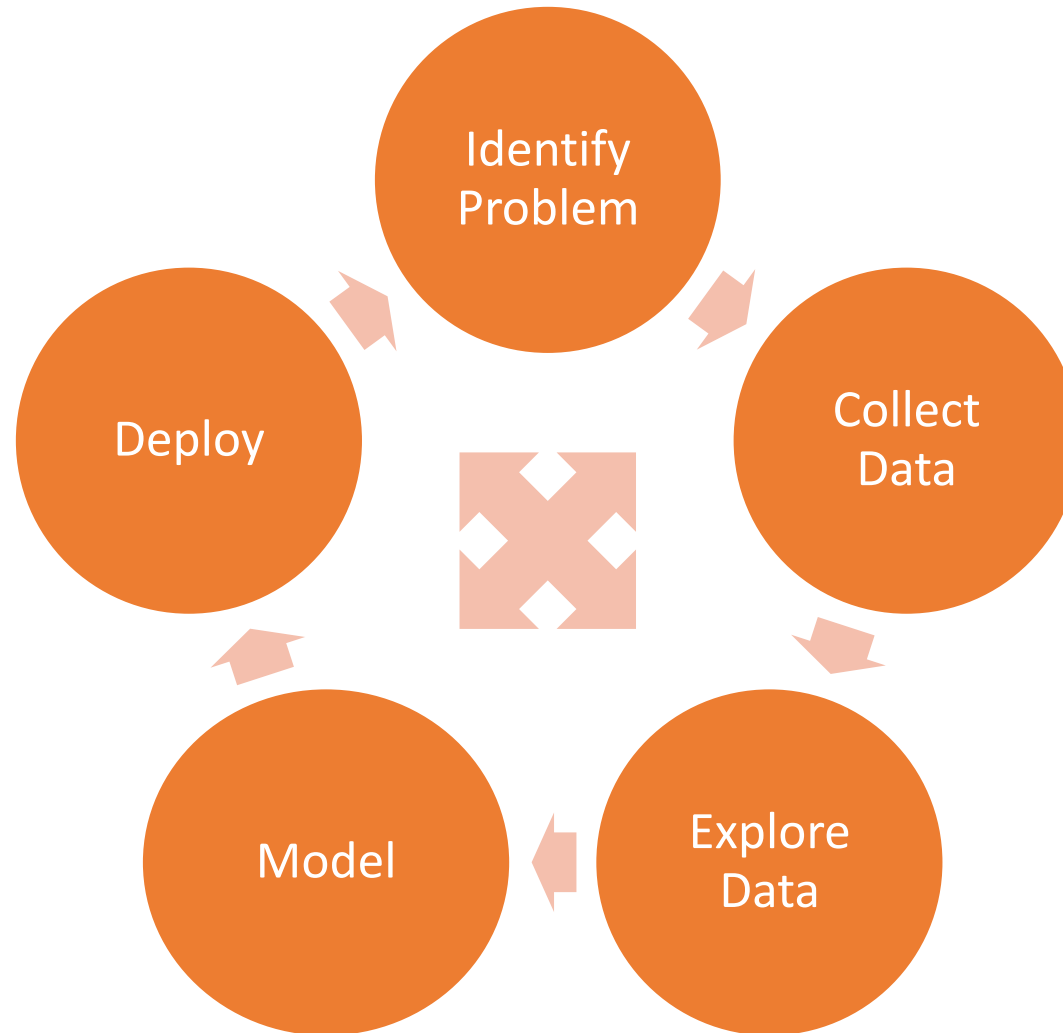
Hello 😊

- Content Developer, ML.NET & Azure Machine Learning
- Twitter: [@ljquintanilla](https://twitter.com/ljquintanilla)
- GitHub: [lqdev](https://github.com/lqdev)
- Blog: <https://luisquintanilla.me>
- Email: lqdev@outlook.com

Agenda

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

Data Science & ML Workflows



Data Science & ML Ecosystem

Data



Compute



Deployment



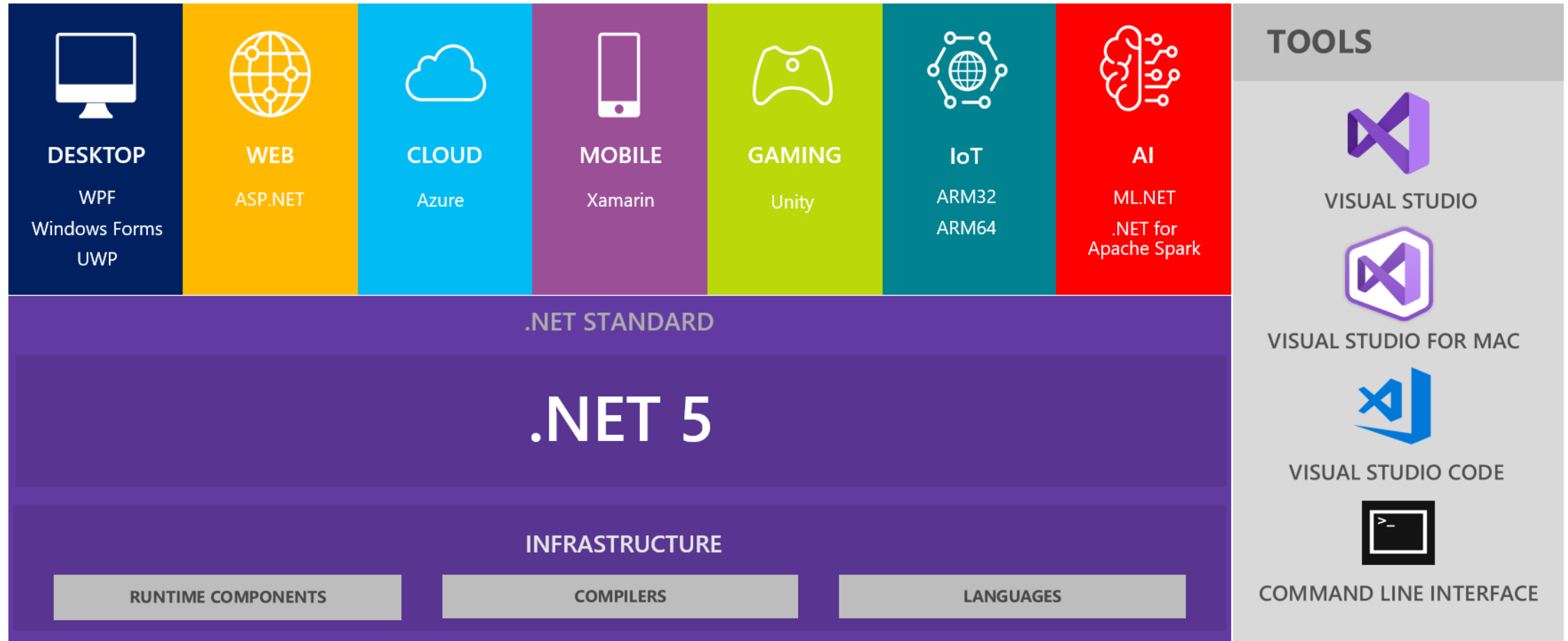
Tooling



Models



.NET – A unified platform



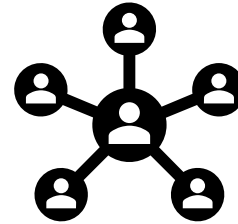
Apache Spark



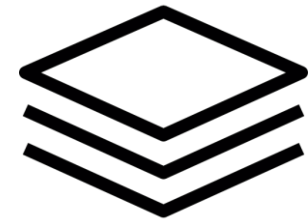
Open
Source



Cross
Platform

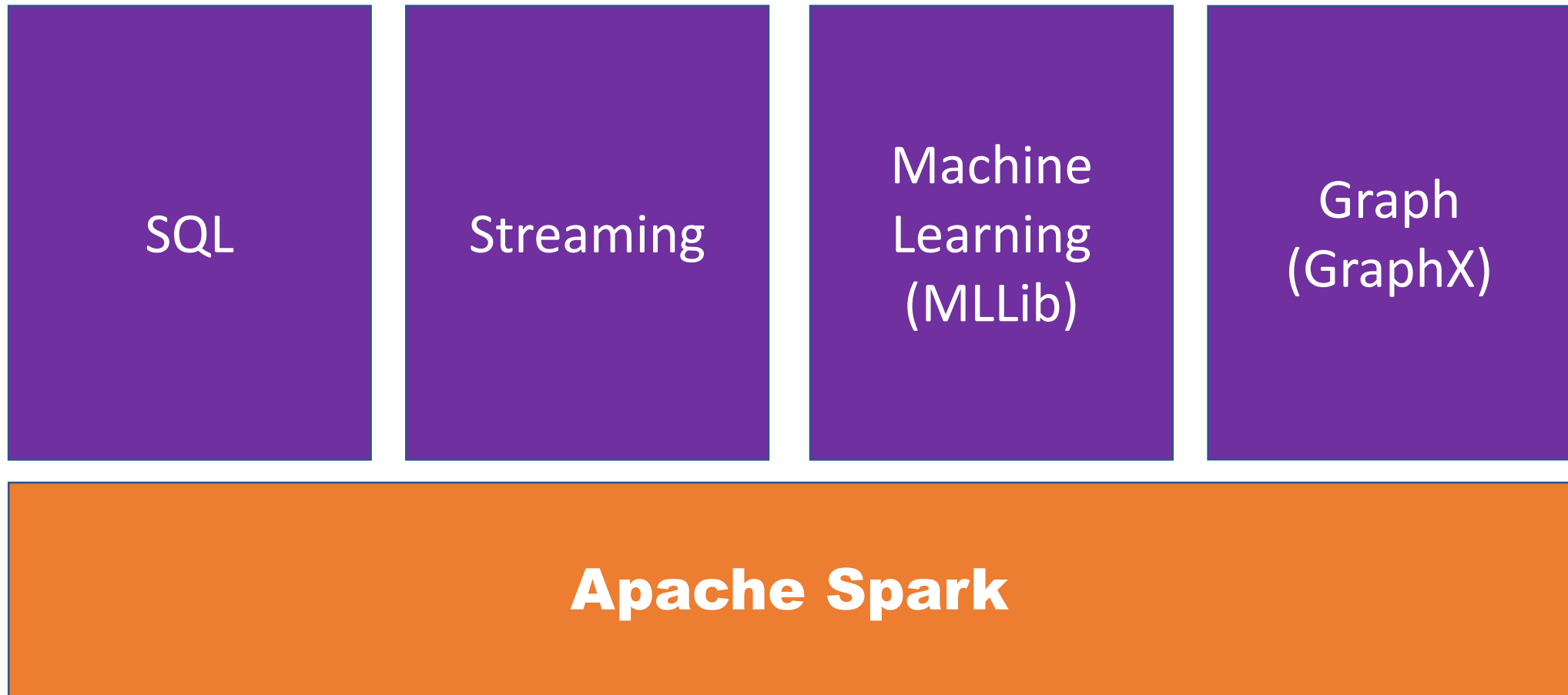


Distributed*

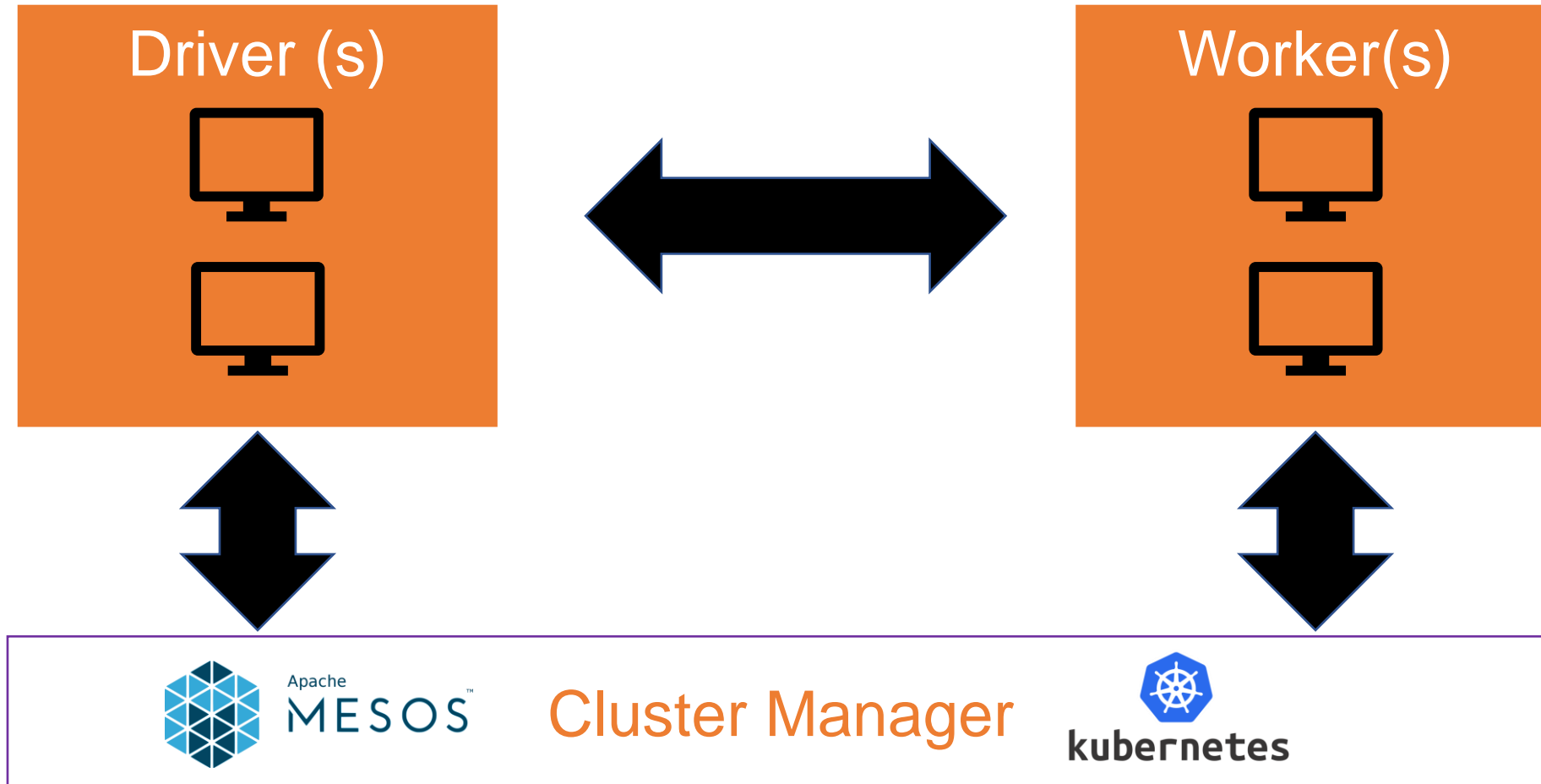


Extensible

Apache Spark Architecture

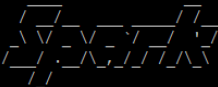


Apache Spark Execution



Interacting with Apache Spark

```
b.MutableMetricsFactory).  
log4j:WARN Please initialize the log4j system properly.  
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more in  
fo.  
Using Spark's repl log4j profile: org/apache/spark/log4j-defaults-repl.properties  
To adjust logging level use sc.setLogLevel("INFO")  
Welcome to
```



```
version 1.6.1
```

```
Using Scala version 2.10.5 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_91)  
Type in expressions to have them evaluated.  
Type :help for more information.
```

Shell

localhost8888/notesbook/Spark!!: jupyter

File Edit View Insert Cell Kernel Help

File Edit View Insert Cell Kernel Help

Python 2

```
In [1]: import os
import sys

spark_home = os.environ.get('SPARK_HOME', None)
if not spark_home:
    raise ValueError("SPARK_HOME environment variable is not set")
sys.path.insert(0, os.path.join(spark_home, 'python'))
sys.path.insert(0, os.path.join(spark_home, 'python/lib/py4j-0.8.1-src.zip'))
sys.path.insert(0, os.path.join(spark_home, 'C:/spark-1.6.0-bin-hadoop2.6/python/lib/py4j-0.9-src.zip'))
execfile(os.path.join(spark_home, 'python/pyspark/shell.py'))
```

Welcome to

version 1.6.0

Using Python version 2.7.10 (default, Oct 21 2015 17:08:47)
SparkContext available as sc, HiveContext available as sqlContext.

In []: |

Notebooks

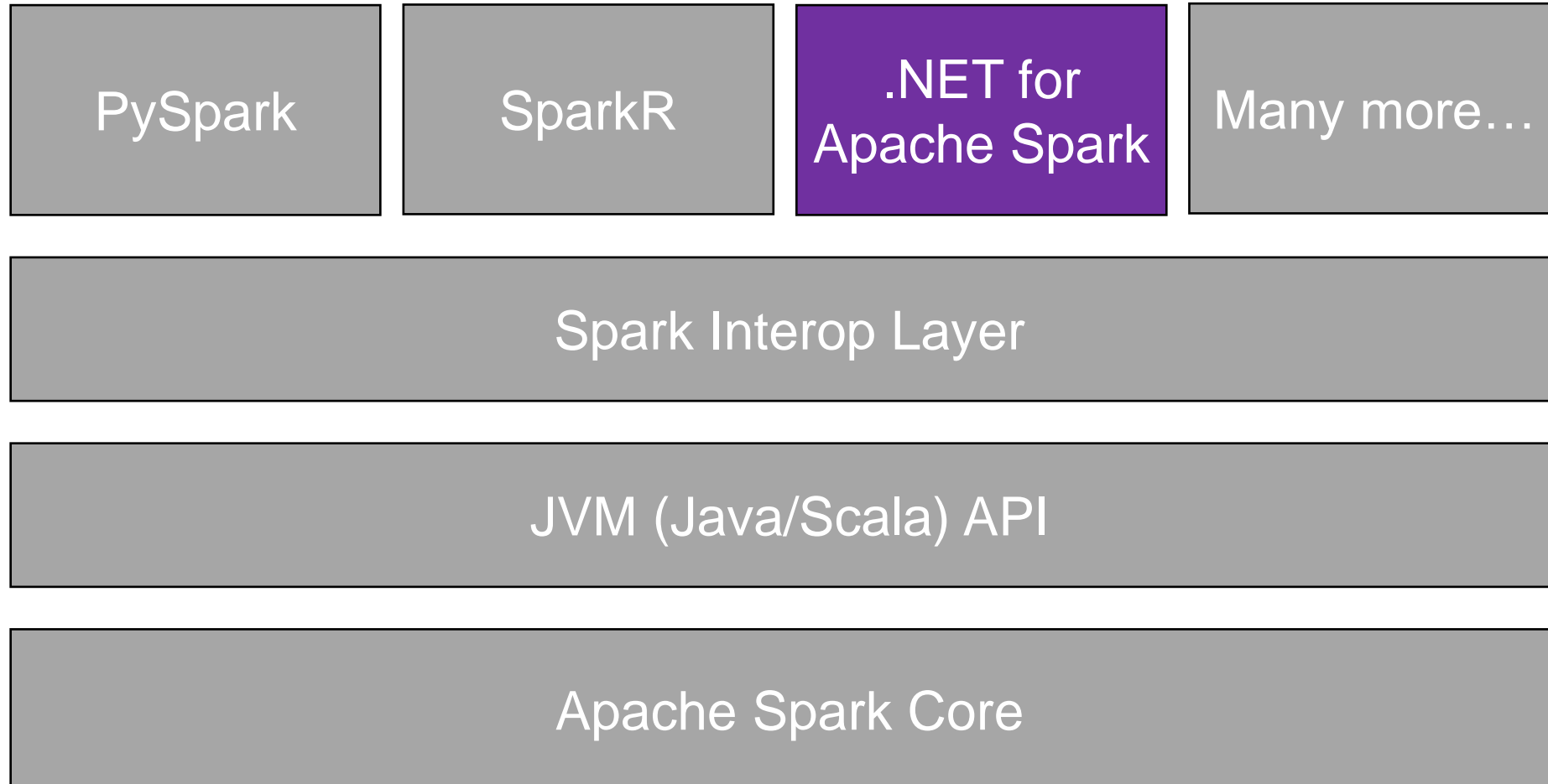
@ljquintanilla

```
// 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");
```

Traditional Apps

Extending Apache Spark



.NET for Apache Spark



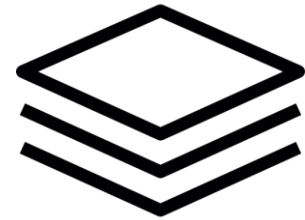
Open
Source



Cross
Platform

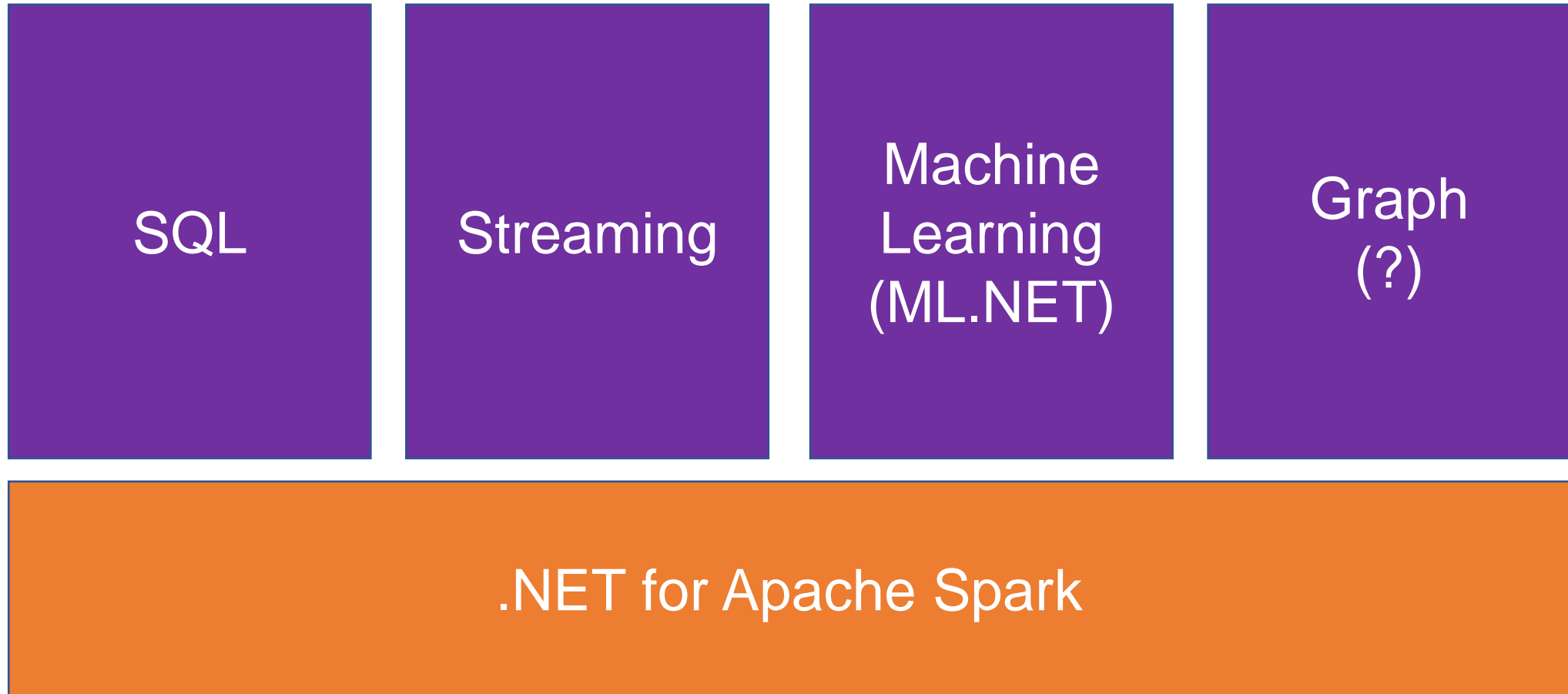


.NET
Standard



Extensible

.NET for Apache Spark Architecture



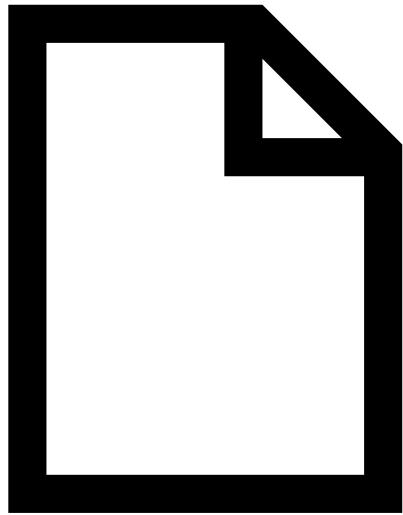
Demo

.NET for Apache Spark

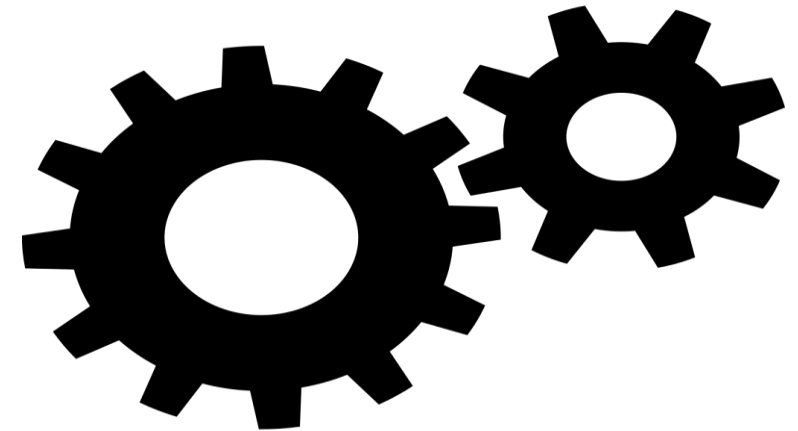
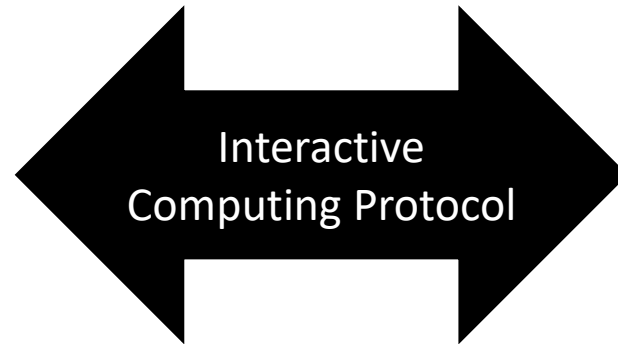
Jupyter Notebooks

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

Jupyter Notebooks Internals

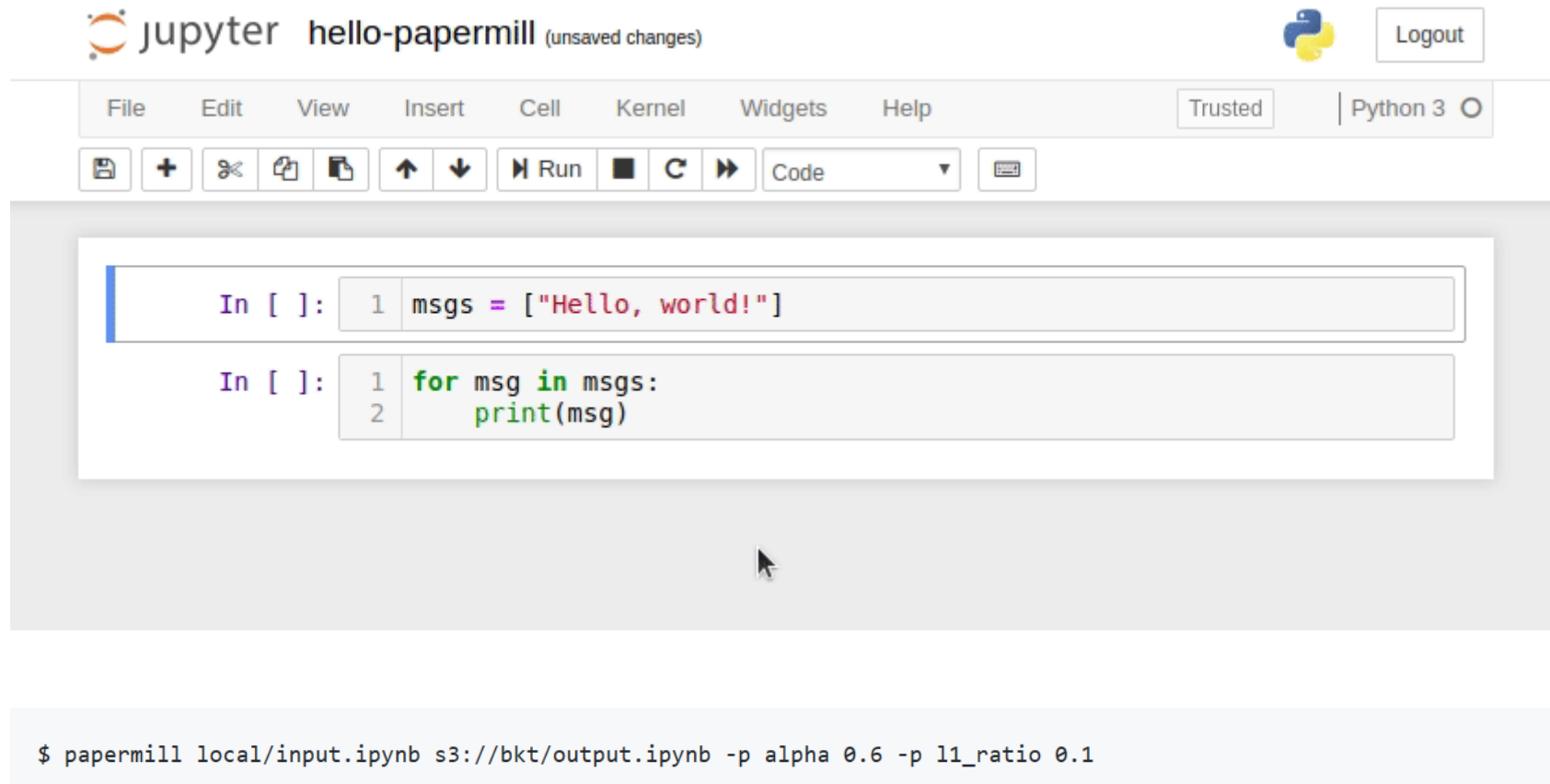


Notebook
Document
Format



Kernel

Extensions



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

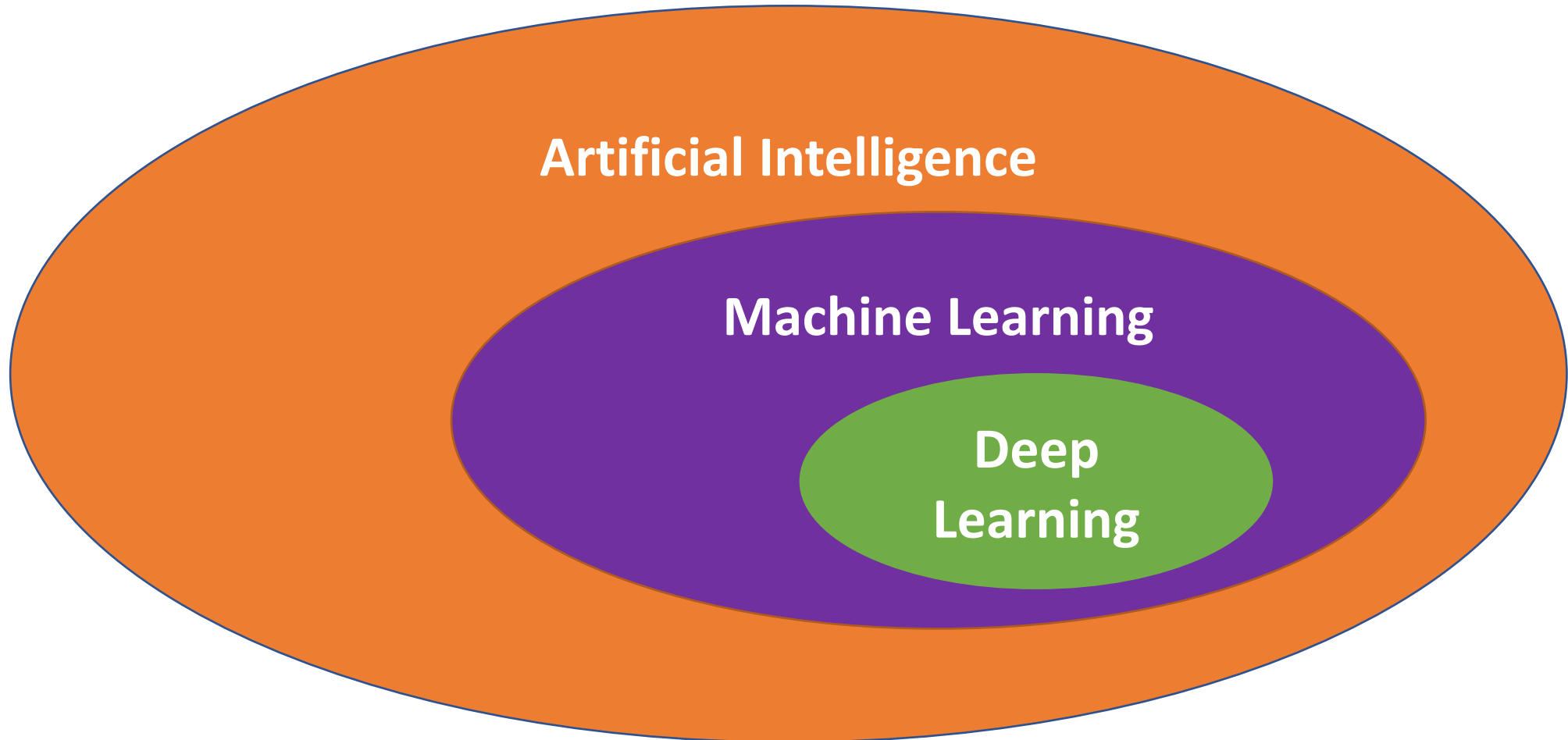
.NET Interactive

- 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?

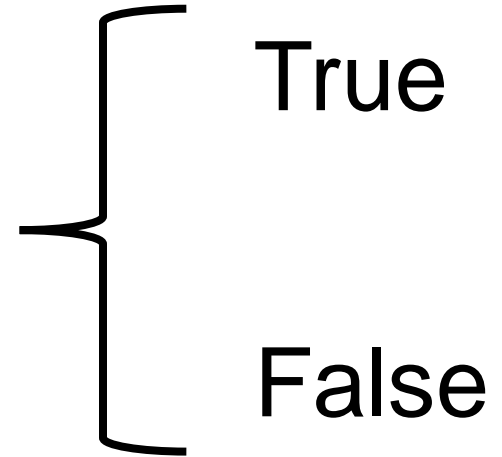


Input



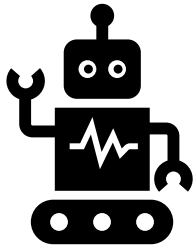
$f(x)$

Model



Output

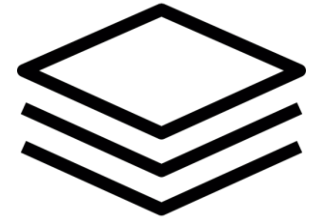
ML.NET



Framework for
Machine Learning



.NET Standard



Proven &
Extensible



Open
Source



Cross
Platform

ML.NET - Framework

Transforms

- Missing Values
- Feature Selection
- Normalization

Trainers

- SVM
- K-Means
- Boosted Trees
- Image Classification

Misc

- Data Loaders
- Evaluators

Extensions

- 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/dotnet-dataml042020>
- <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>