



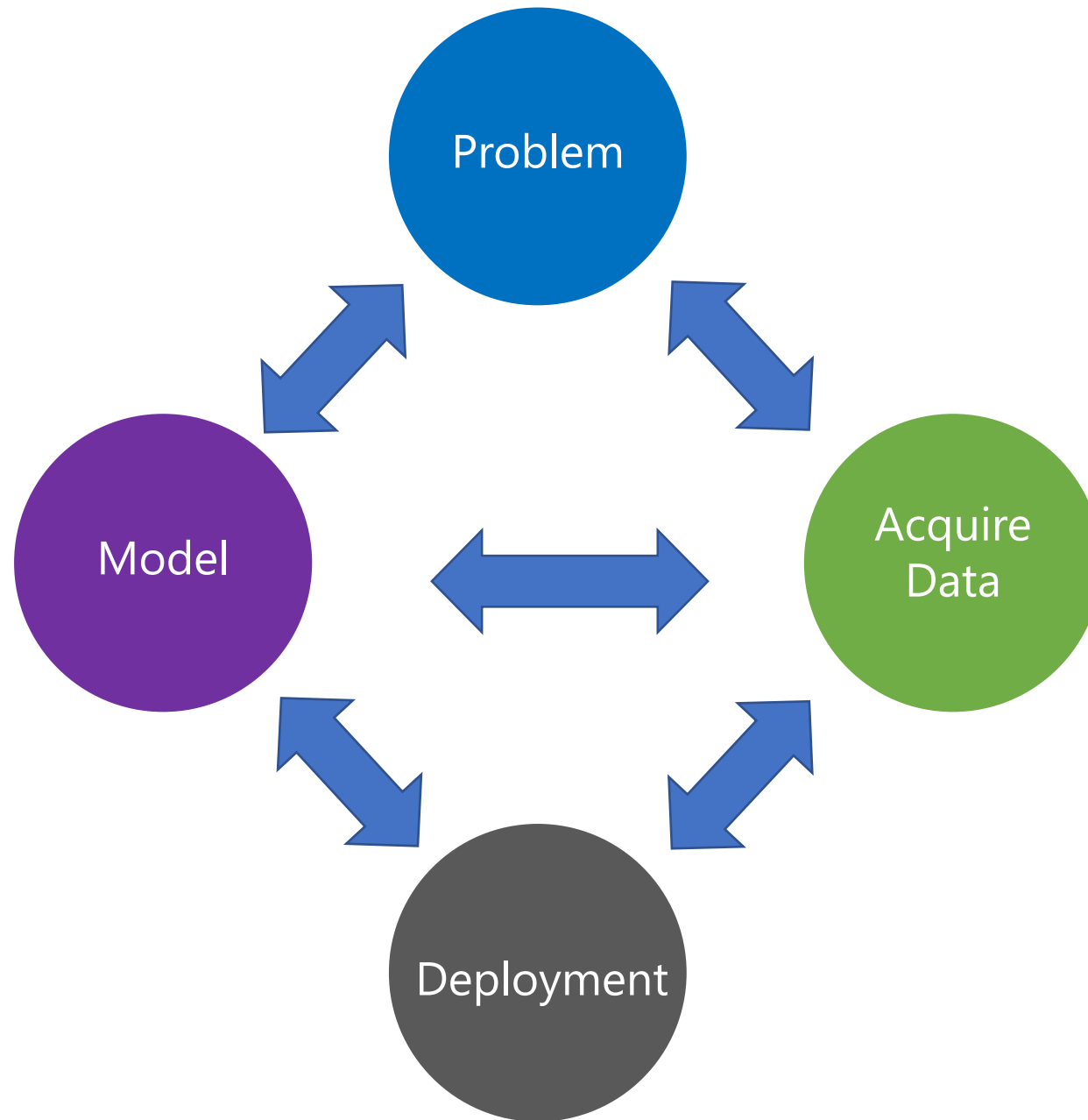
85%



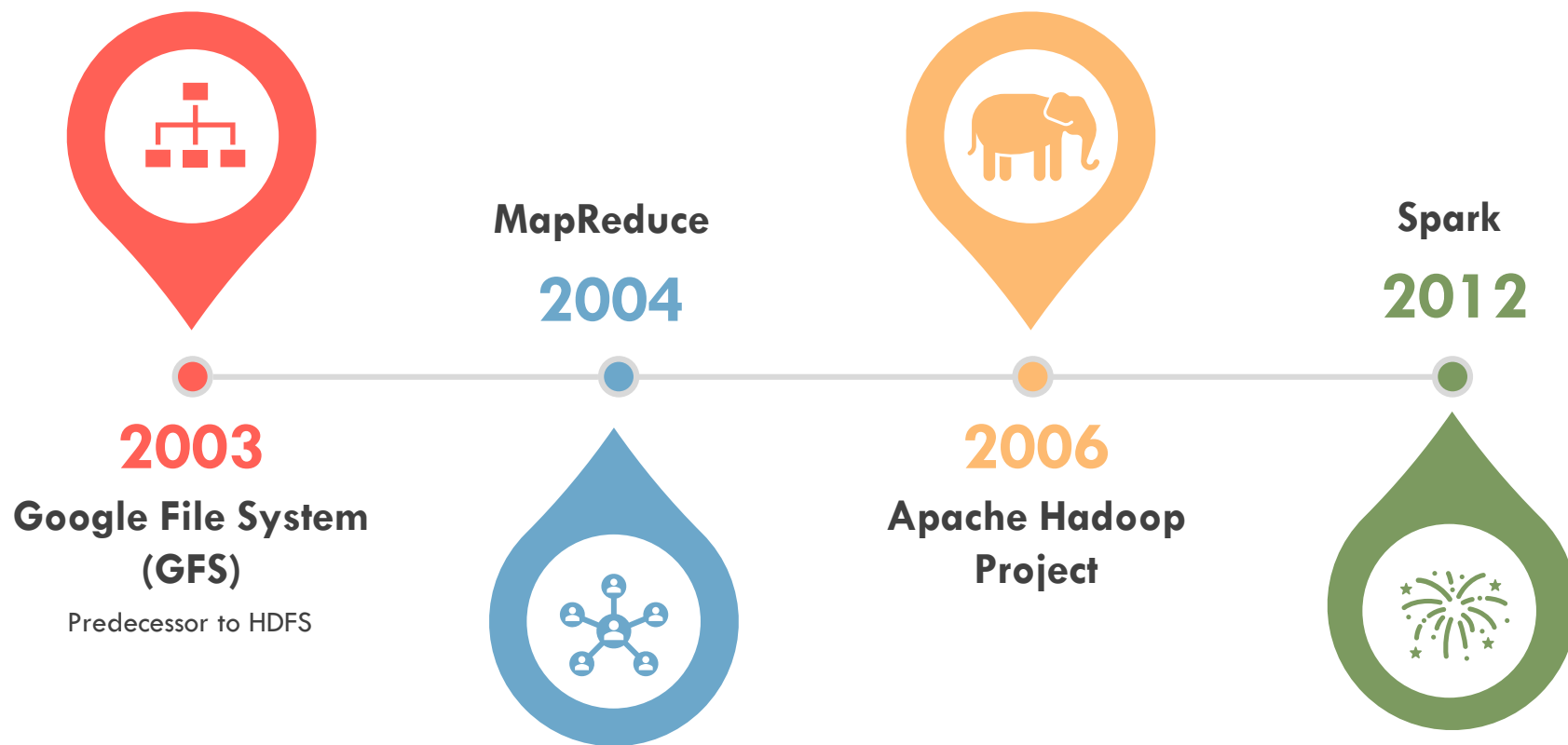
Getting your R Project into Production using Databricks

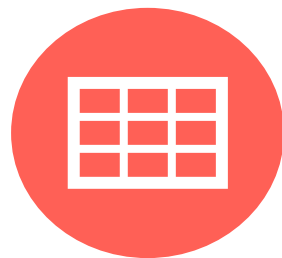
Michael Johnson
@MikeJohnsonZA

Data Science Lifecycle









DataFrames API / SQL



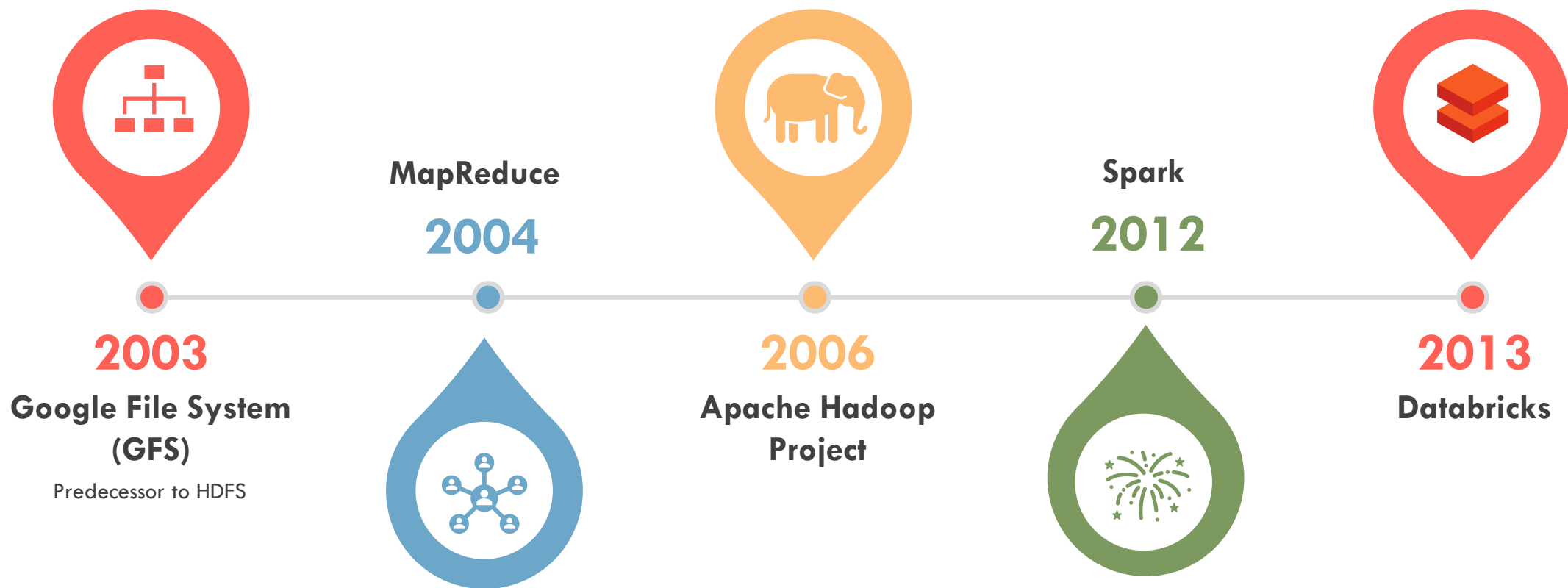
Streaming



Graph



Machine Learning



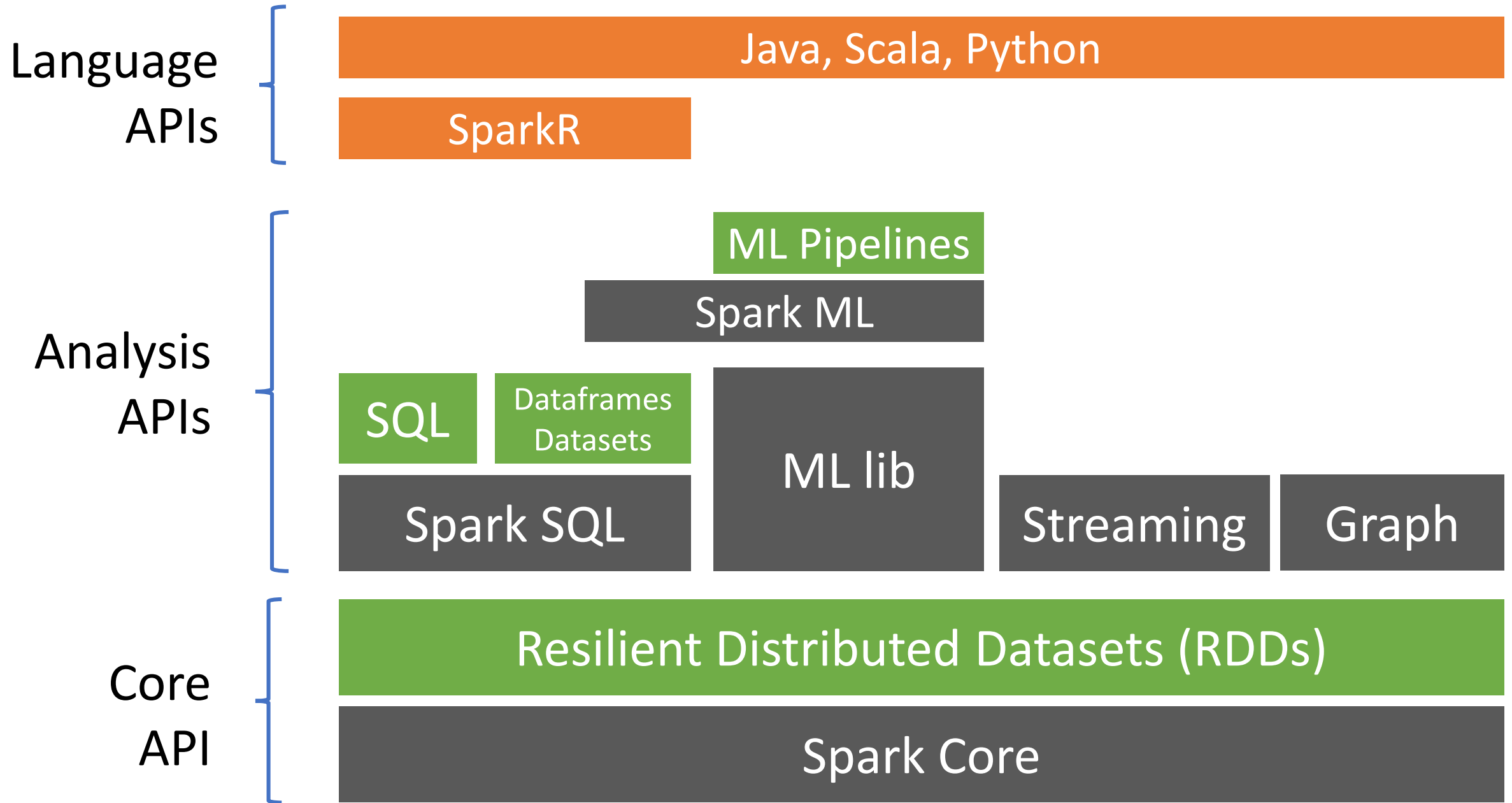
Architecture

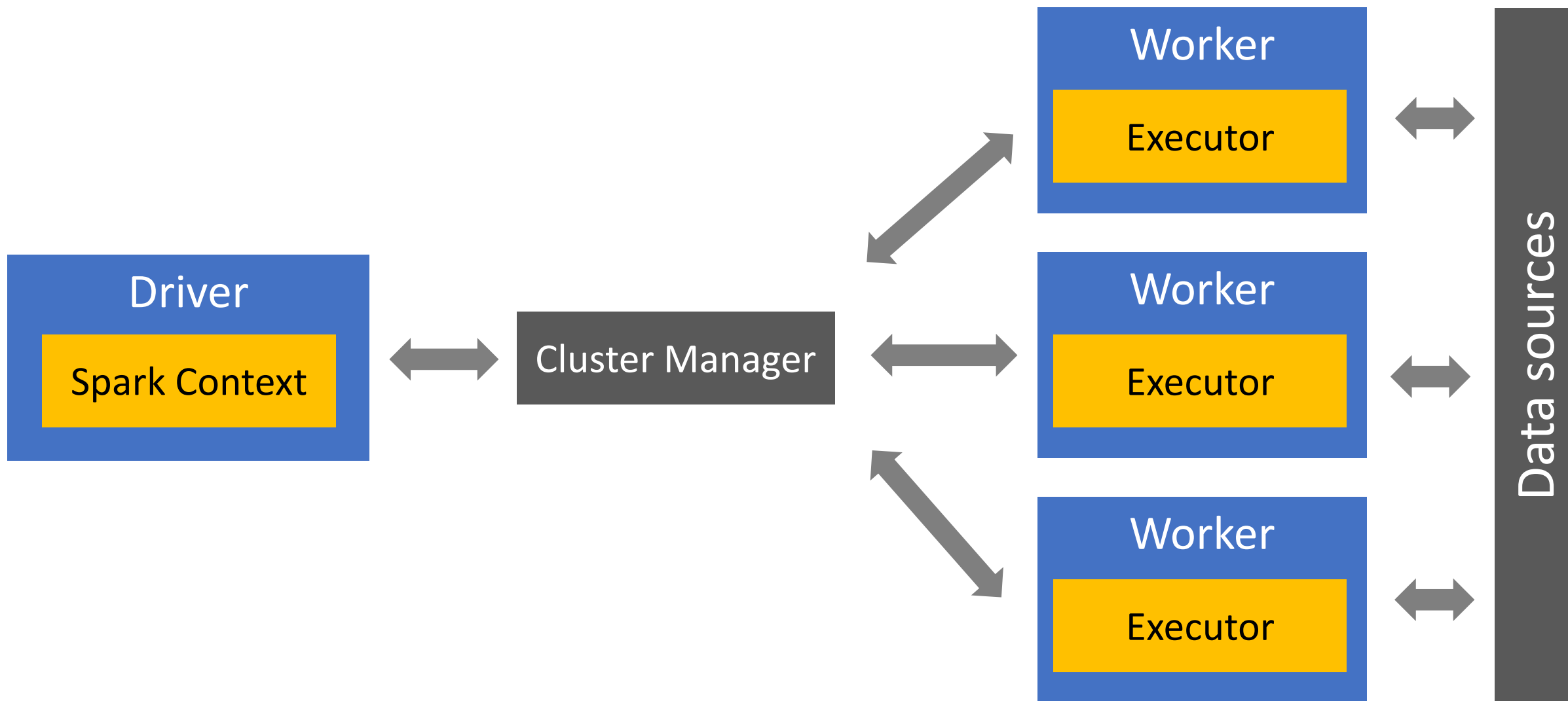
**Resilient
Distributed
Datasets**

**Directed
Acyclic
Graph**

**Resilient
Distributed
Datasets**

**Directed
Acyclic
Graph**





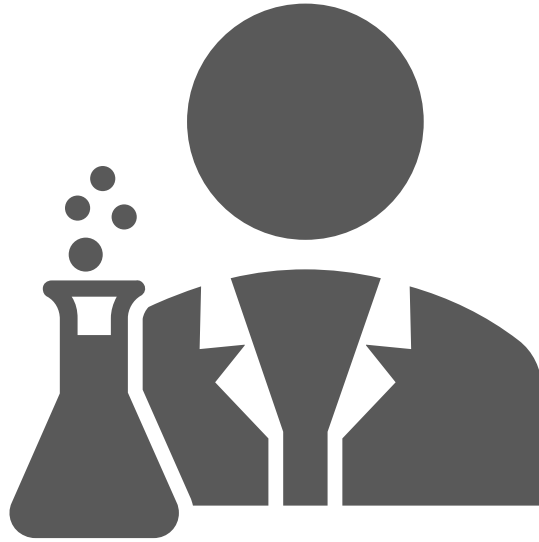
Databricks



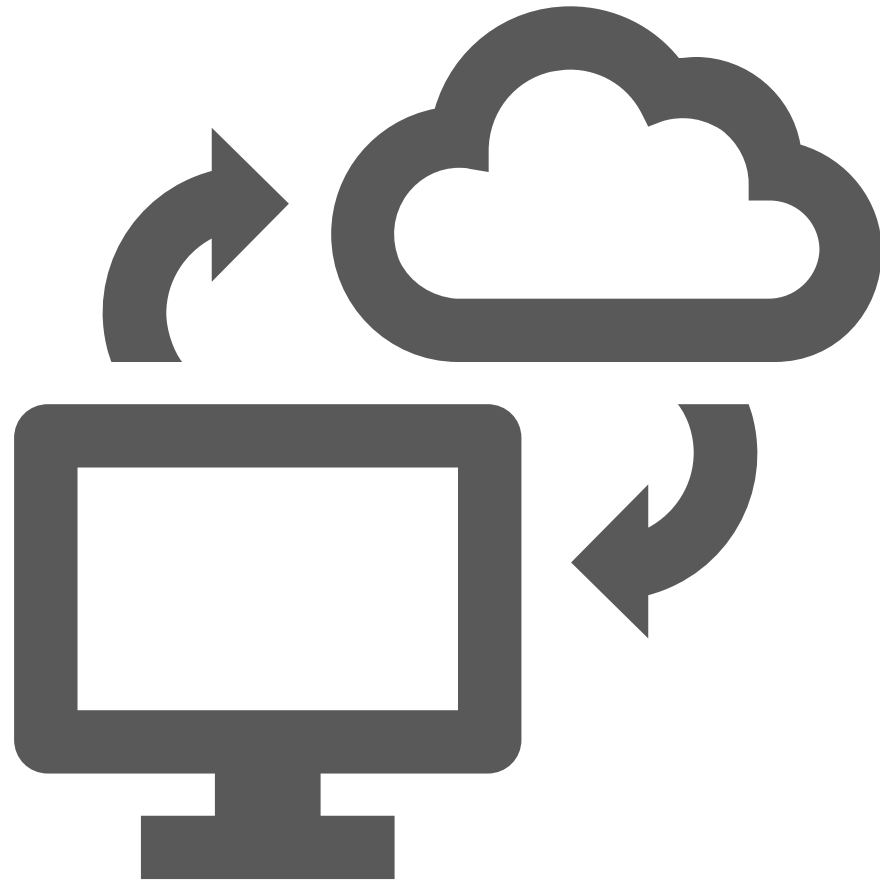
Implementation



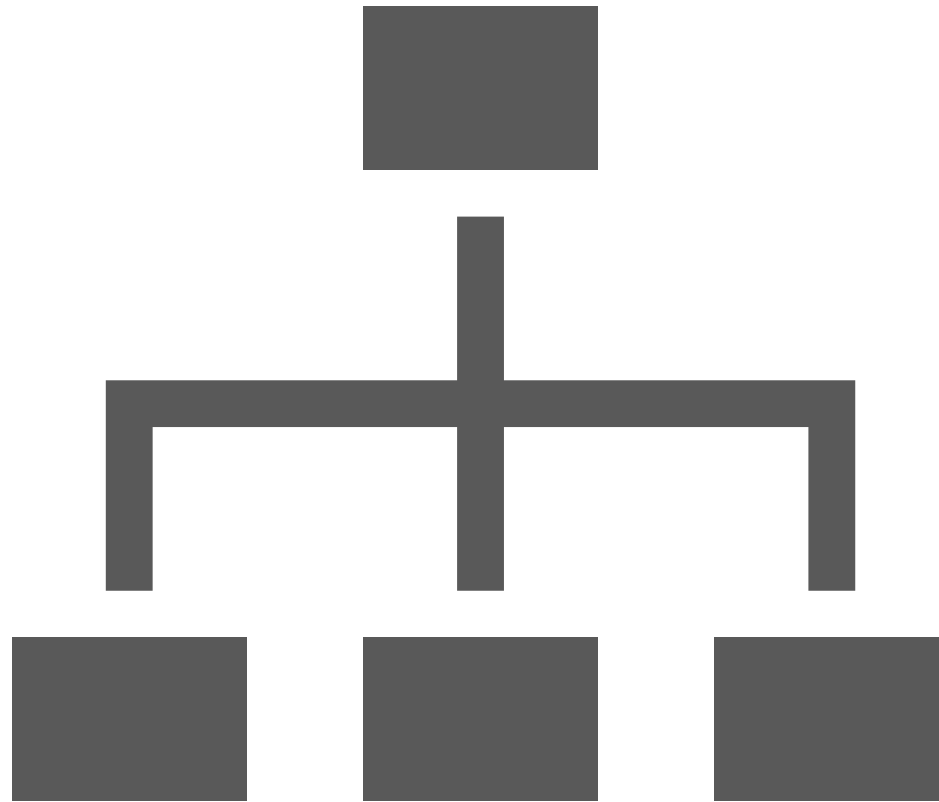
One Environment for the Whole Team



One Click Setup



Scale



Security



Interactive Notebooks

Create Notebook

Name

Language

Python

Scala

SQL

✓ R

Cluster

customer-analysis (270 GB, Run ⬆️⬆️)

Cancel

Create

R Studio Integration

