Tools Used in This Course

After this video you will be able to..

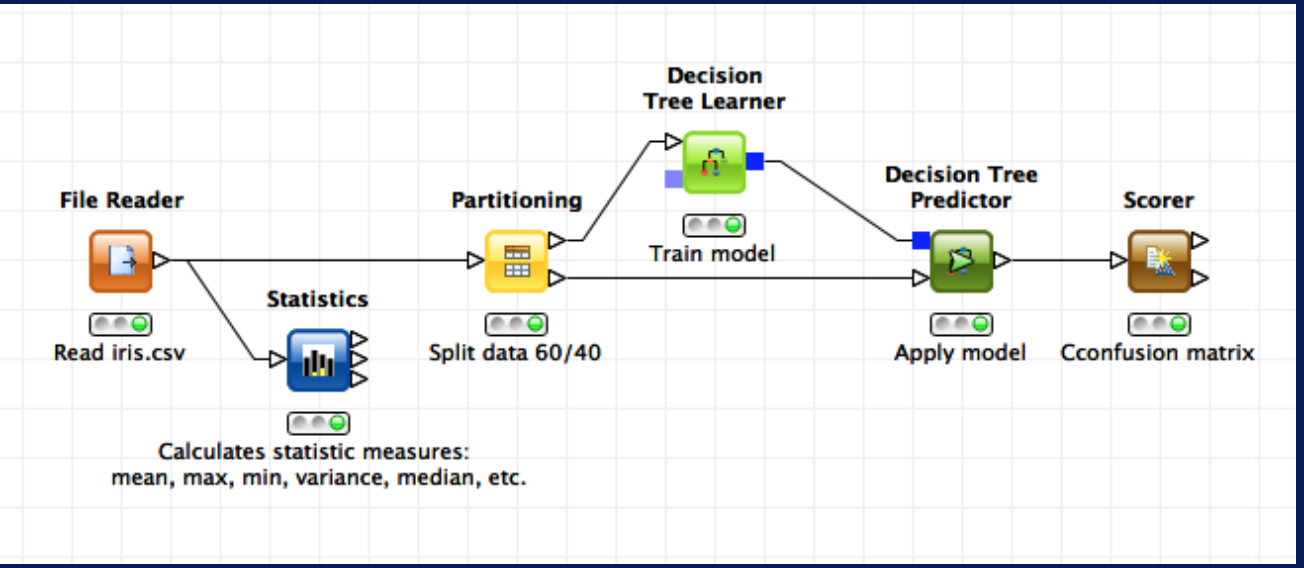
- Describe what KNIME is
- Describe what Spark MLlib is
- Contrast KNIME and ML as machine learning tools

Tools for This Course

KNIME Analytics Platform

- Platform for data analytics, reporting, and visualization
- GUI-based approach with drag-and-drop interface
- Nodes provide functionality
- Nodes are assembled to create workflows

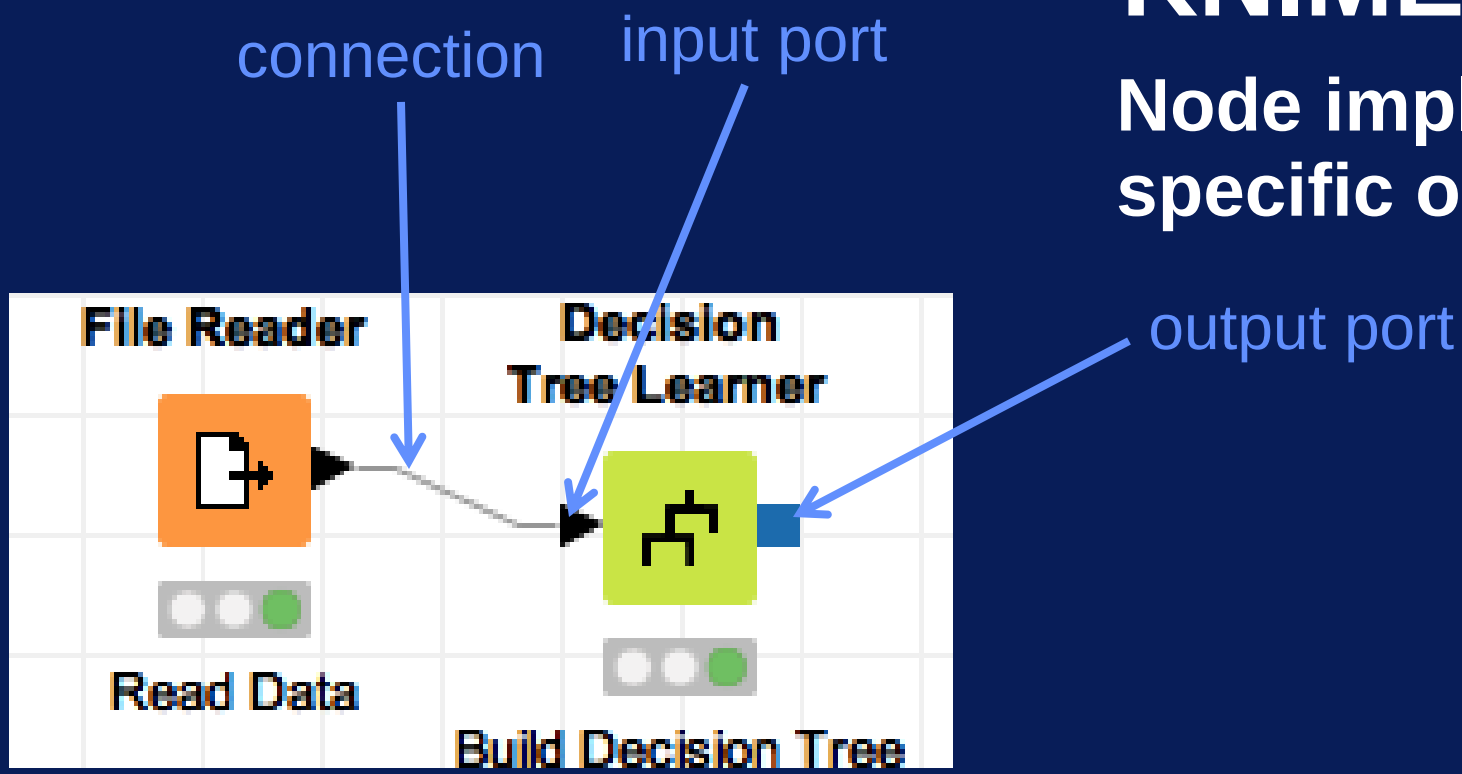
KNIME Workflow

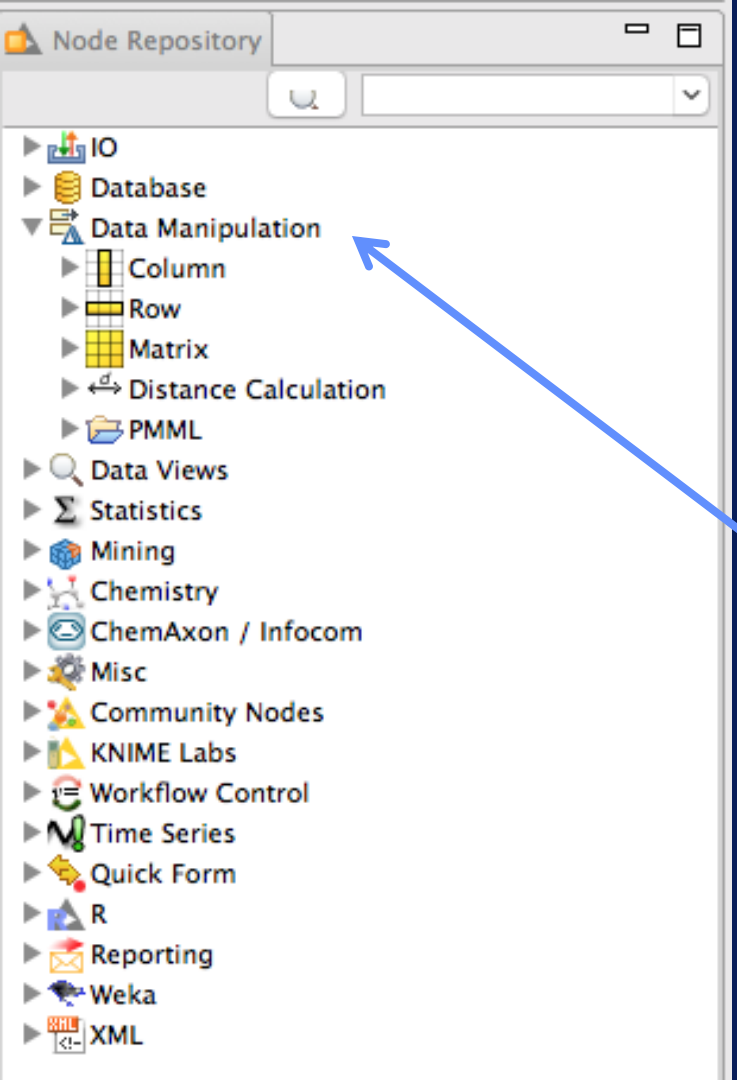


- Visual representation of steps in analysis process
- Workflow is composed of nodes

KNIME Node

Node implements specific operation





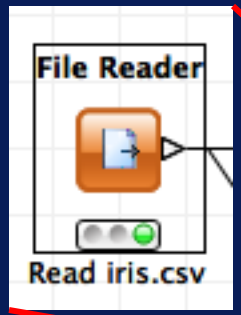
Node Repository

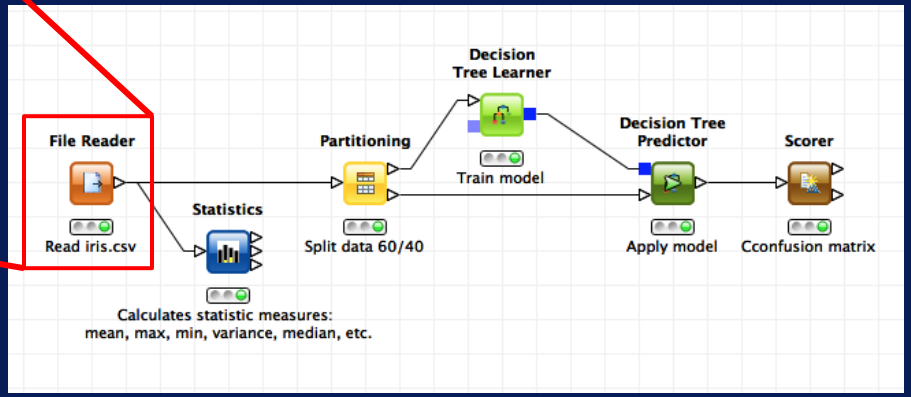
Contains nodes organized by category

Expand category to see subcategories and nodes

KNIME

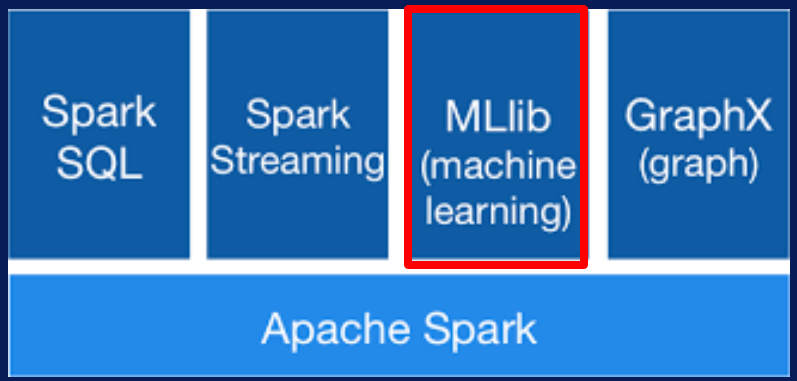
- GUI-based
- Drag-and-drop
- Interactive
- For small datasets





Spark MLlib

- **Scalable machine learning library**
- **Runs on Spark**
  - Distributed computing platform



Spark MLlib

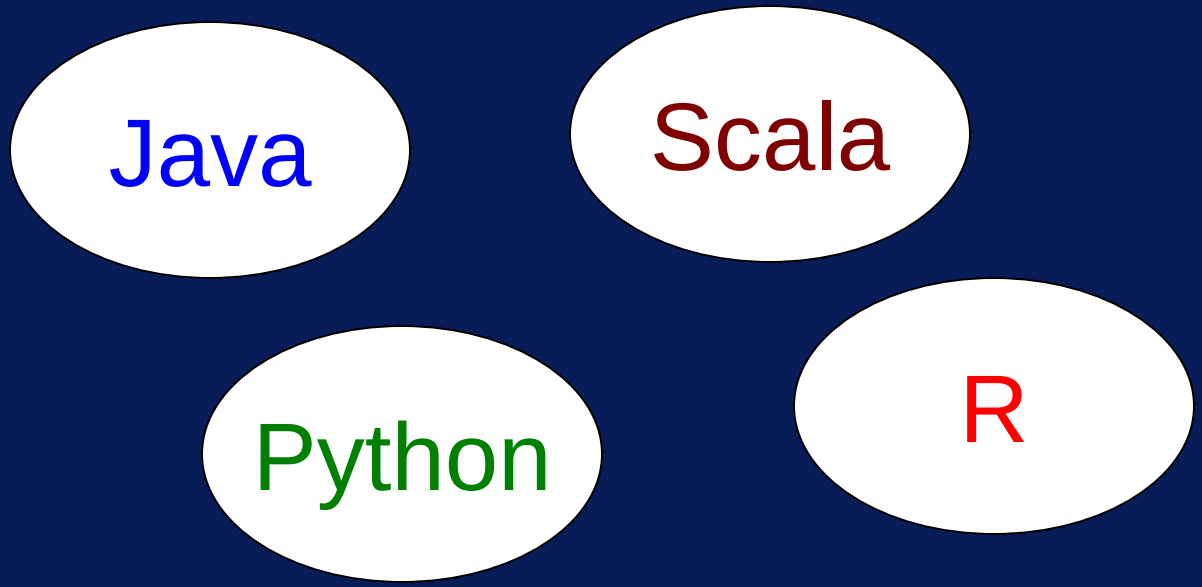
- **Write code to implement machine learning operations**

```
# Read and parse data
data = sc.textFile("data.txt")

# Decision tree for classification
model = DecisionTree.trainClassifier
        (parsedData, numClasses=2)
print(model.toDebugString())
model.save(sc, "decisionTreeModel")
```

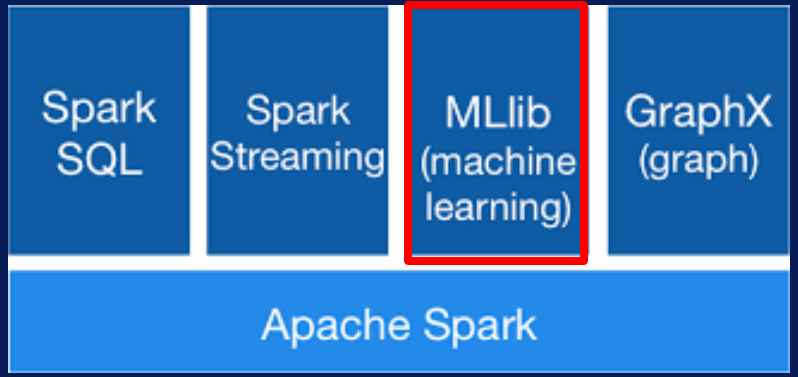
Spark MLlib

- **Provides APIs for**



Spark MLlib

- Distributed platform
- Scalable algorithms & techniques
- For large datasets
- Requires coding



KNIME & Spark MLlib