

User Defined Functions



Why

DataFrames, Datasets and Spark SQL can't express everything

Register custom code to run inside structured APIs



UDFs

User defined functions: custom logic whenever DataFrames API is not enough

```
// 1 - create a Scala function
val countWords = (text: String) =>
    text.split("\\s+").length

// 2 - register the UDF
val countWordsUDF = udf(countWords)

// 3 - use the UDF as a normal Spark function
val moviesWithWordCountDF = moviesDF.select(
    col("Title"),
    countWordsUDF(col("Title")).as("Title_Words")
)
```

Can run on multiple columns – same steps, just multi-arg Scala functions

Spark rocks

