* **Narrow and wide transformation:** Diagram

  Description automatically generated

Timeline

Description automatically generated

Diagram

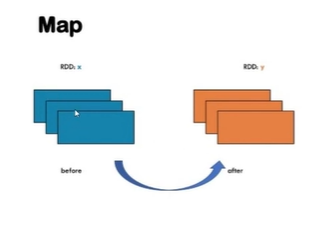
Description automatically generated with low confidence

Chart

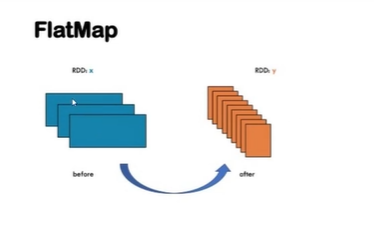
Description automatically generated

* **Narrow Transformation:**

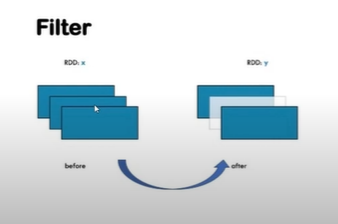
1. **Map:**

****

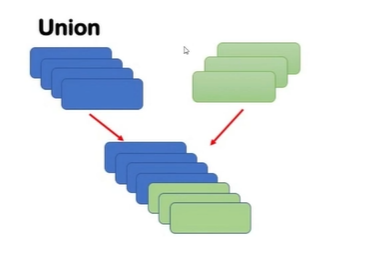
1. **flatMap:**

****

1. **Filter:**

****

1. **Union:**

****

* **Wide Transformation:**

1. **GroupBy:**

A picture containing chart

Description automatically generated

1. **Intersection:**

Diagram

Description automatically generated

1. **Substract:** Removes the common value from the RDDs.

Text, letter

Description automatically generated

1. **Distinct:** It will give the unique values from a RDD.
2. **mapValues:** it is same as map.
3. **ReduceByKey:**

Diagram

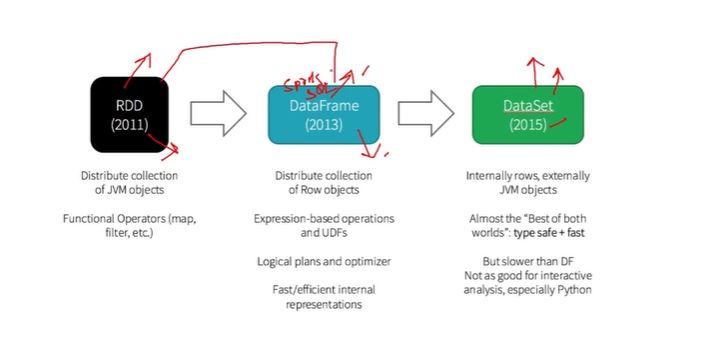
Description automatically generated

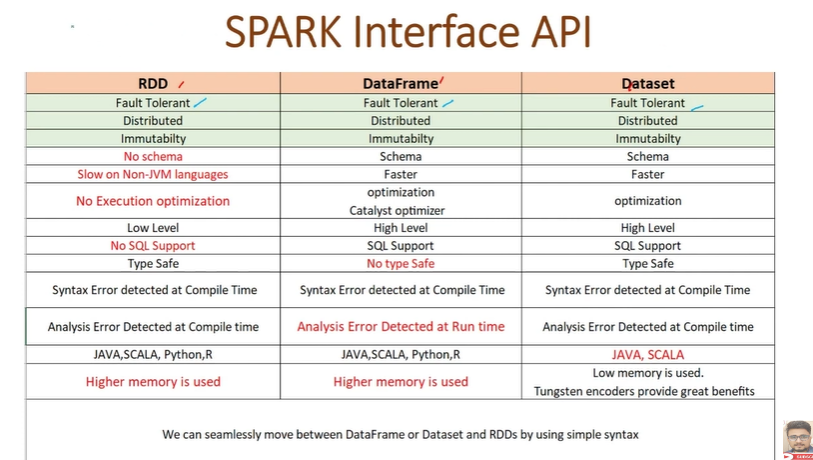
1. **GroupByKey:**

Diagram

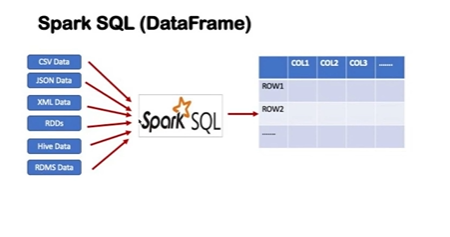
Description automatically generated

* **RDDs vs Dataset vs Dataframe:**





* **Spark SQL Dataframe:**

****

import pyspark

spark = pyspark.sql.SparkSession.builder.master("local").appName('Tutorial').getOrCreate()