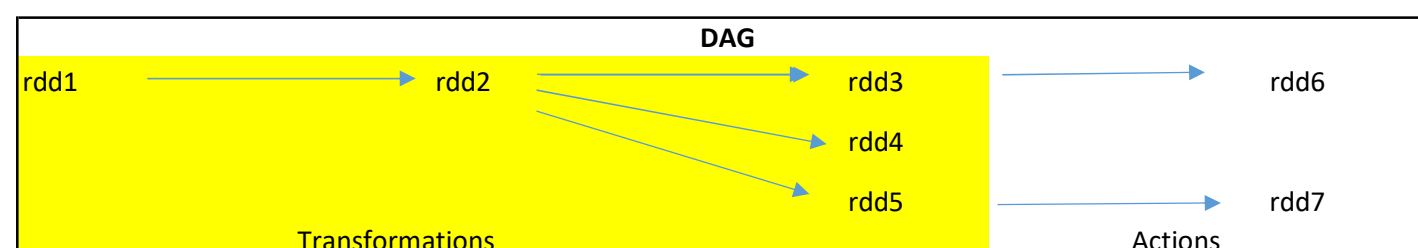


Note: In case of python we can assign change the values of the variable (Mutable).
However, in case of functional programming we cant change the value of a variable.
Therefore, the concept of RDD is very important.

<code>rdd1 = spark.sparkContext.parallelize([1,2,3,3,2,1,35,5,7,8,9])</code>	RDD Creation
<code>rdd2 = rdd1.map(lambda x: x*x)</code>	RDD Transformation 1
<code>rdd3 = rdd2.filter(lambda x: x%2==0)</code>	RDD Transformation 2
<code>rdd4 = rdd2.map(lambda x: x/4)</code>	RDD Transformation 3
<code>rdd5 = rdd2.filter(lambda x: x>0)</code>	RDD Transformation 4
<code>rdd6 = rdd3.collect()</code>	RDD Action
<code>rdd7 = rdd5.count()</code>	RDD Action

Since rdd2 has been used multiple times to create other rdds. Thus , this rdd2 can be cached .

Till the creation of rdd5 there was no execution with the data only DAG graph was built.



Internals of Job Execution In Spark

