Какие команды отвечают за:  
**1)Сохранение результата в текстовый файл** (Это Action или Transformation?) –

**Ответ:**

saveAsTextFile(outputFile)

Action

**2)Как получить первые n-элементов массива** (Это Action или Transformation?)

**Ответ:**

pyspark.sql.functions.slice(x, 0, n)

берем n элементов из x

Transformation

**3)Объединить два RDD в один** (Это Action или Transformation?)

**Ответ:**

rdd1.union(rdd2)

Transformation

**4)В чем разница между Reduce и CoGroup-операторами** (Это Action или Transformation?)

**Ответ:**

Reduce проходит по всему списку элементов и на каждой итерации модифицирует результирующее значение.

Пример подсчета суммы элементов:

numbers = sc.parallelize([1,7,8,9,5,77,48])

sum = numbers.reduce(lambda a,b:a+b)

\*) здесь sc - это SparkContext

Reduce **–** Action

Операции группировки по ключу –получаем список с элементами, в которых есть ключ(какое-то поле) и список значений этого ключа.

Например, groupBy("city")

cogroup - работают так же как и group операции, только с несколькими rdd.

Например, a.cogroup(b)

CoGroup-операторы **–** Transformation

Для подсчета количества значений каждого из ключей – можно использовать reduceByKey и groupByKey.

Разница между ними – reduceByKey производит агрегацию до перемешивания, а groupByKey в конце сразу всех.

**Нарисовать DAG для Spark’а для подсчета количества уникальных слов в файле**

