**Домашнее задание №5**

1.1 Какие команды отвечают за:

|  |  |  |
| --- | --- | --- |
| **Описание** | **Команда** | **Тип команды** |
| Сохранение результата в текстовый файл | saveAsTextFile(path) | Action |
| получение первых n-элементов массива | take(n) | Action |
| Объединение двух RDD в один | union(otherDataset) | Transformation |

1.2 В чем разница между Reduce и CoGroup-операторами:

|  |  |  |
| --- | --- | --- |
| **Команда** | **Описание** | **Тип команды** |
| reduce(func) | Агрегирует значения датасета с помощью вложенной функции, которая принимает на вход два значение, а возвращает одно.  Пример: ds1 = [1,2,3,4,5,6]  C помощью reduce(sum a,b: a+b) получим ds2 = {21} | Action |
| cogroup(otherDataset, [numPartitions]) | Объединяет 2 датасета в результирующий датасет, который содержит кортежи со списком значений для каждого из ключей обоих датасетов.  Пример: ds1 = [{A,1}, {B, 1}, {C, 1}, {D,1}]  ds2 = [{A,2}, {C,5}, {E,17}]  Результатом функции будет являться:  ds3 = [{A, {1,2}}, {B, 1}, {C, {1, 5}}, {D,1}, {E,17}] | Transformation |

2. Нарисуйте DAG для Spark для подсчёта количества уникальных слов в файле.

Данное задание выполнено на отдельном листе.