

# № 1

## **aggregateByKey** operation:

- This operation gathers available JavaPairRDDs together according to the key. Gathered values are aggregated according to given combining rules (functions) and initial element. As a result, kind of map is generated, where keys are keys, which were used for combining and values – aggregated values.
- `aggregateByKey(new UserSet(), (agg, value) -> agg.add(value), (agg1, agg2) -> agg1.add(agg2));`
  - `new UserSet()` – initial value of aggregator, initialized for each key
  - `(agg, value) -> agg.add(value)` – combining function to combine aggregator with the new value
  - `(agg1, agg2) -> agg1.add(agg2)` – combining function to combine two aggregator.

## **Map and Reduce:**

- **aggregateByKey** corresponds to the **Reduce** phase. This function tells how to reduce previously mapped values.