



Pair RDDs

Pair RDDs

- **Pair RDDs** are special kinds of RDDs that store data in a key-value style.



Pair RDDs

- **Pair RDDs** are special kinds of RDDs that store data in a key-value style.

(India, DELHI)

(India, MUMBAI)

(USA, WASHINGTON)

(SA, CAPE TOWN)

Pair RDDs

- **Pair RDDs** are special kinds of RDDs that store data in a key-value style.
- They are are similar to dictionaries in Python.



Pair RDDs

- **Pair RDDs** are special kinds of RDDs that store data in a key-value style.
- They are similar to dictionaries in Python.

```
{  
  India: DELHI,  
  India: MUMBAI,  
  USA: WASHINGTON,  
  SA: CAPE TOWN  
}
```

Pair RDDs

- **Pair RDDs** are special kinds of RDDs that store data in a key-value style.
- They are similar to dictionaries in Python.
- Allows to perform operation on each key in parallel.



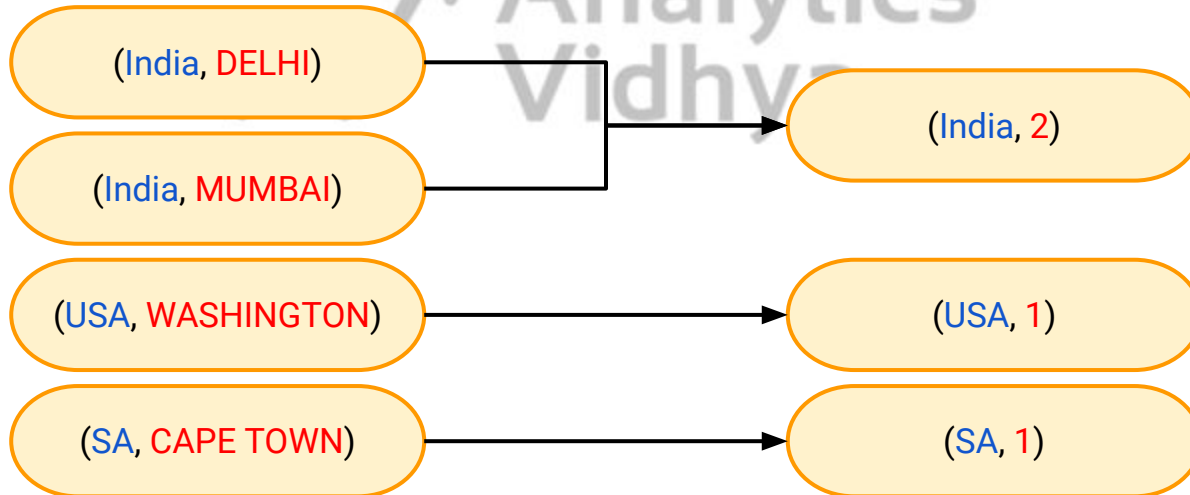
Pair RDDs

- **Pair RDDs** are special kinds of RDDs that store data in a key-value style.
- They are similar to dictionaries in Python.
- Allows to perform operation on each key in parallel.
- Used to aggregate values for elements with same key.



Pair RDDs

- **Pair RDDs** are special kinds of RDDs that store data in a key-value style.
- They are similar to dictionaries in Python.
- Allows to perform operation on each key in parallel.
- Used to aggregate values for elements with same key.



Pair RDDs

- **Pair RDDs** are special kinds of RDDs that store data in a key-value style.
- They are similar to dictionaries in Python.
- Allows to perform operation on each key in parallel.
- Used to aggregate values for elements with same key.
- It has its own special Transformation and Action functions in addition to those for RDDs.



Pair RDDs

- **Pair RDDs** are special kinds of RDDs that store data in a key-value style.
- They are similar to dictionaries in Python.
- Allows to perform operation on each key in parallel.
- Used to aggregate values for elements with same key.
- It has its own special Transformation and Action functions in addition to those for RDDs.

`groupByKey(), reduceByKey(), countByKey(), join()` and many more..

Pair RDD Concept: Word Count



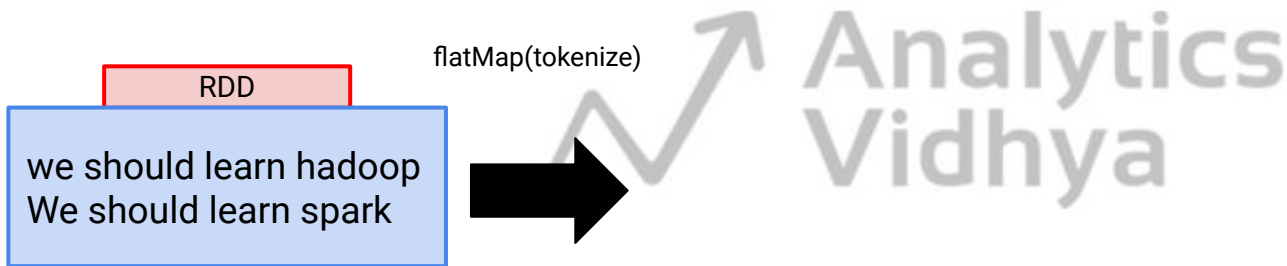
Pair RDD Concept: Word Count

RDD

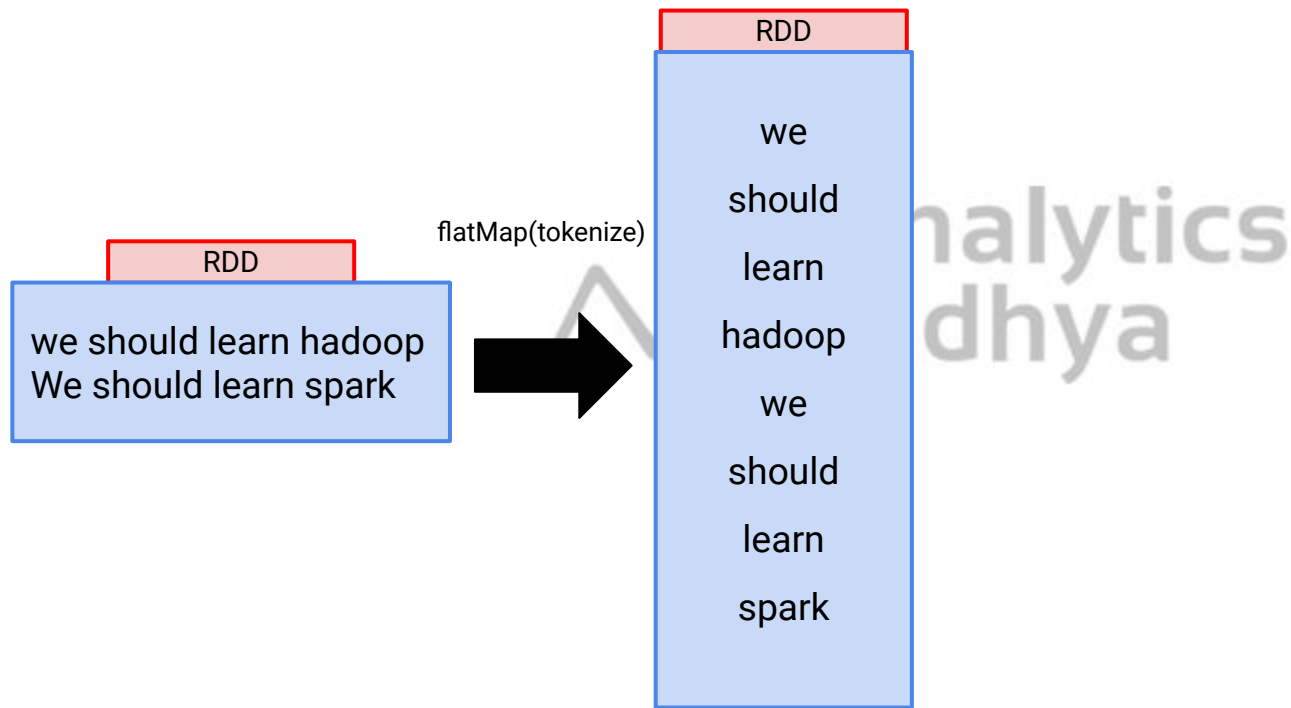
we should learn hadoop
We should learn spark



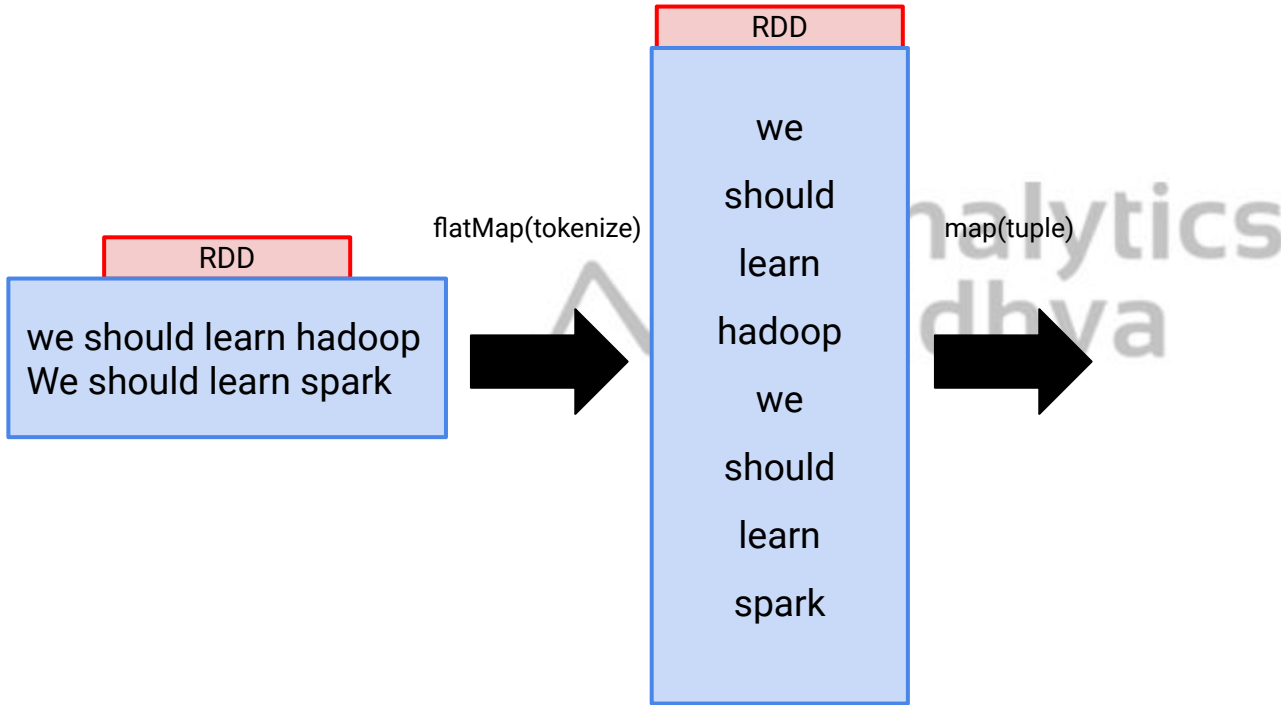
Pair RDD Concept: Word Count



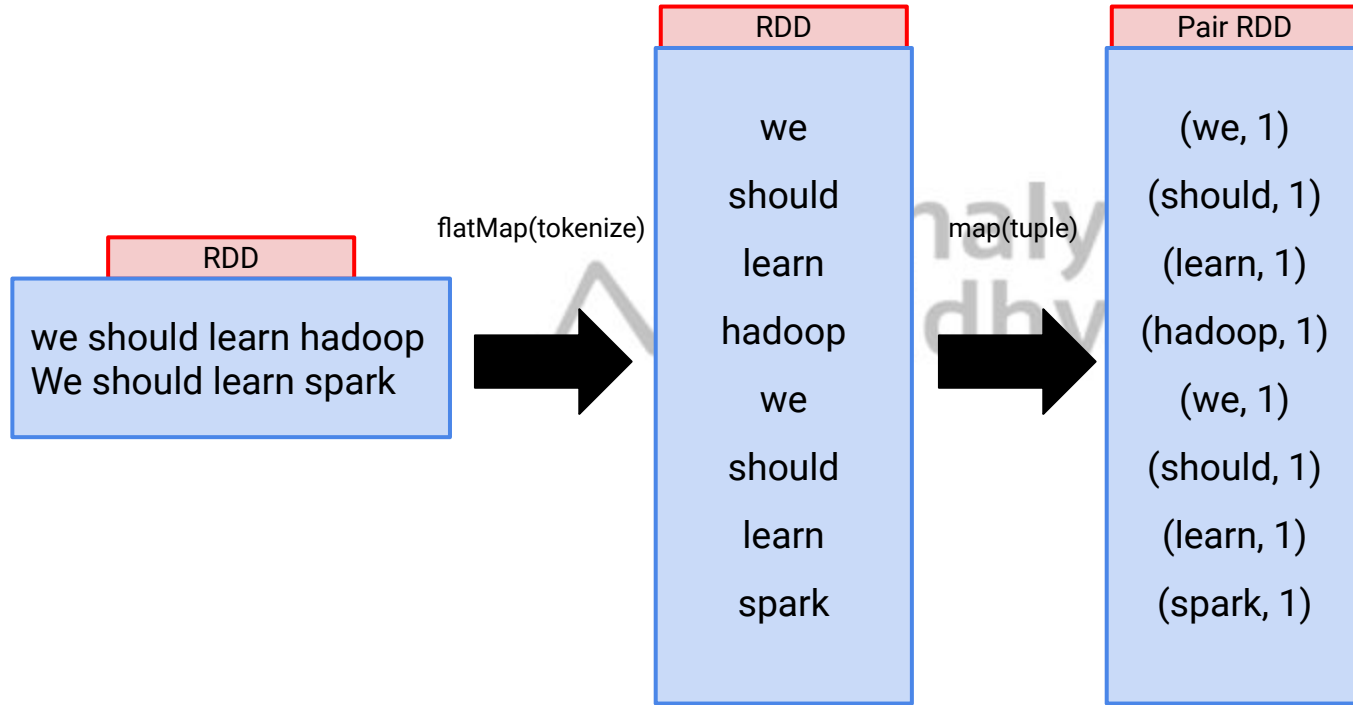
Pair RDD Concept: Word Count



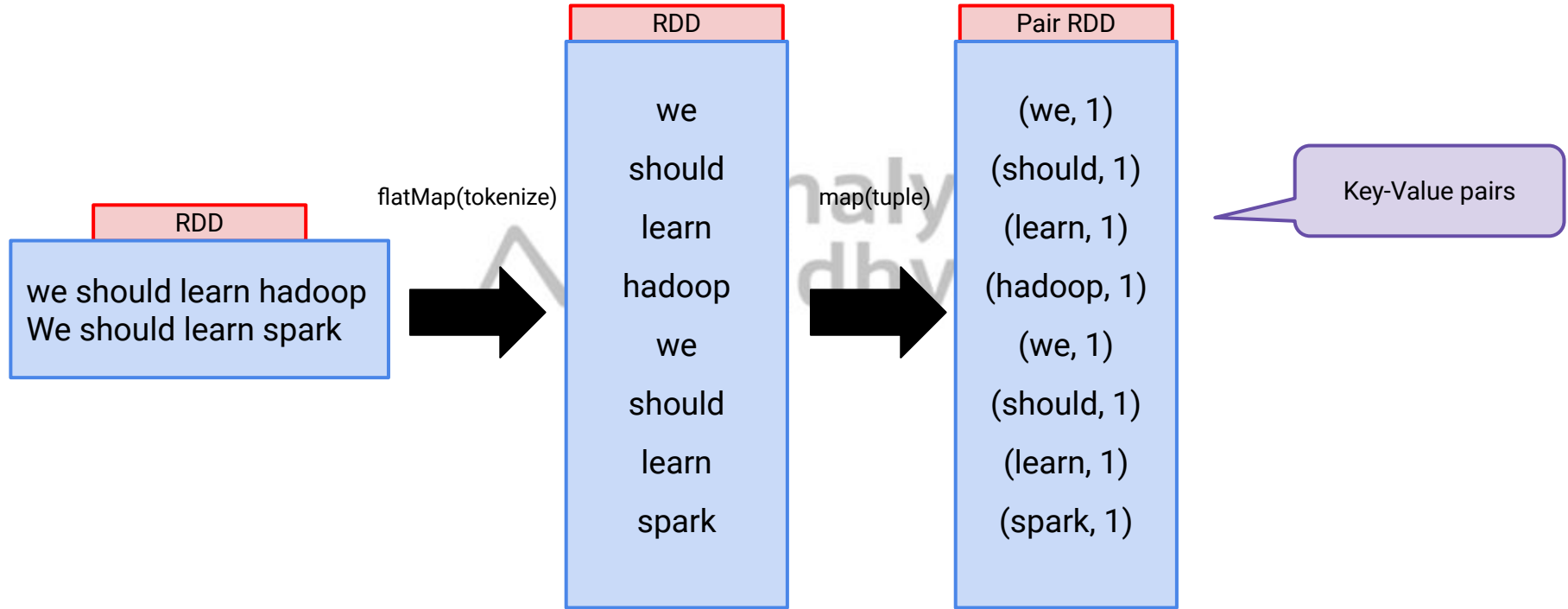
Pair RDD Concept: Word Count



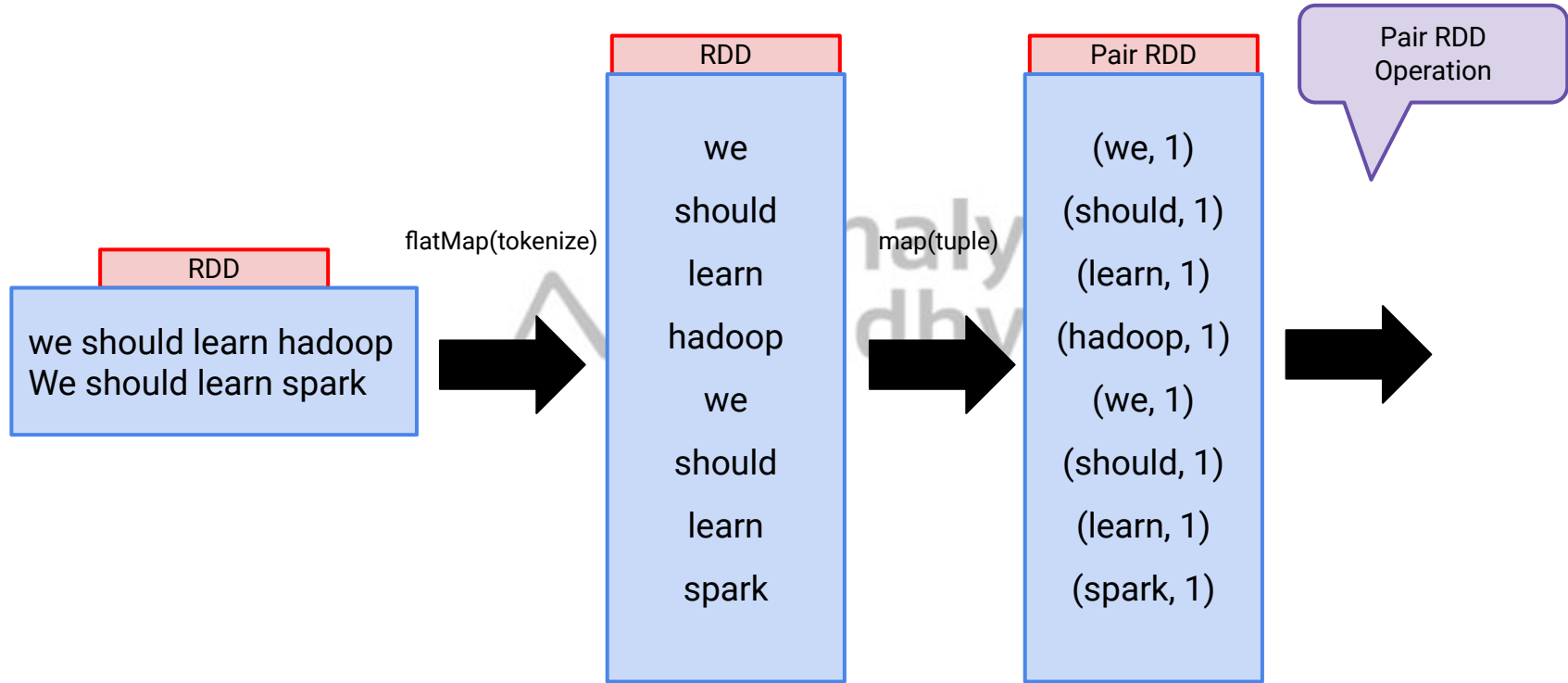
Pair RDD Concept: Word Count



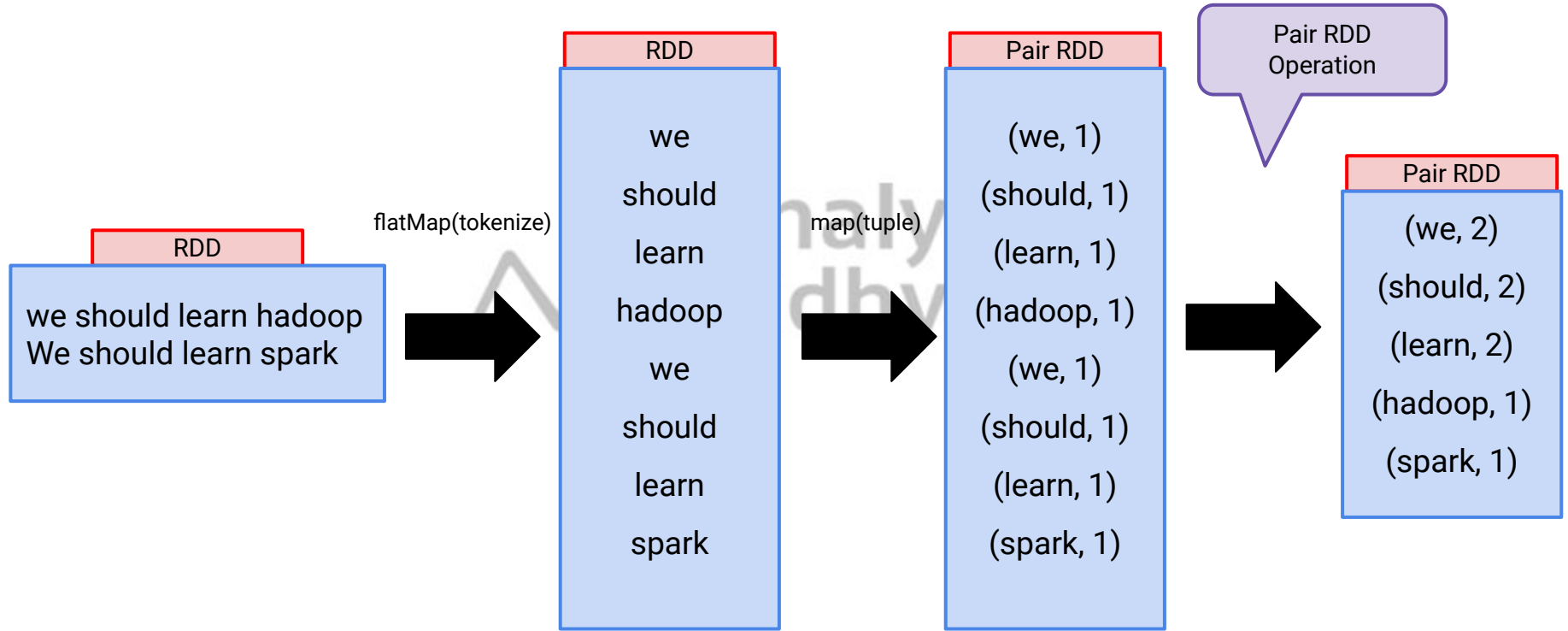
Pair RDD Concept: Word Count



Pair RDD Concept: Word Count



Pair RDD Concept: Word Count





Thank You!!