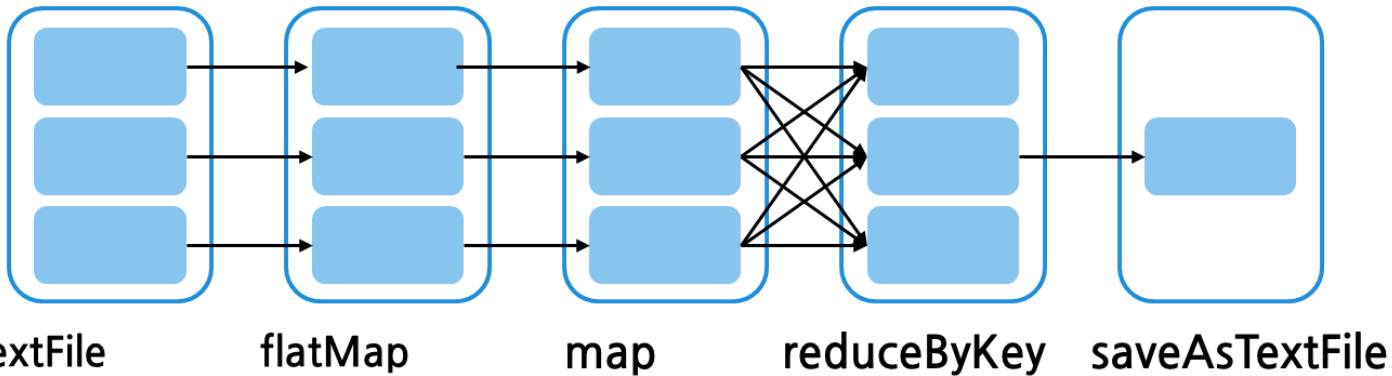


스파크 응용 프로그램

1 Word Count를 위한 RDD 연산



◀ RDD 연산의 예시 ▶

2 파이썬을 사용한 Word count 예제

```
text_file = sc.textFile("hdfs://...")
counts = text_file.flatMap(lambda line: line.split(" ")) \
    .map(lambda word: (word, 1)) \
    .reduceByKey(lambda a, b: a + b)
Counts.saveAsTextFile("hdfs://...")
```

스파크 응용 프로그램

3 스칼라를 사용한 Word count 예제

```
val textFile = sc.textFile("hdfs://...")
val counts = textFile.flatMap(line => line.split(" "))
                      .map(word => (word, 1))
                      .reduceByKey(_ + _)
Counts.saveAsTextFile("hdfs://...")
```

4 자바를 사용한 Word count 예제

```
JavaRDD<String> textFile = sc.textFile("hdfs://...");
JavaPairRDD<String, Integer> counts = textFile
    .flatMap(s -> Arrays.asList(s.split(" ")).iterator())
    .mapToPair(word -> new Tuple2<>(word, 1))
    .reduceByKey((a, b) -> a + b);
Counts.saveAsTextFile("hdfs://...")
```

스파크 응용 프로그램

5 spark-submit을 통한 실행

① 파이썬일 경우

1

텍스트 에디터를 이용해서
파이썬 코드(SimpleApp.py) 작성

2

스파크 홈 디렉토리로 이동

어플리케이션 실행

```
$ YOUR_SPARK_HOME/bin/spark-submit --master local[4]  
SimpleApp.py
```

② 스칼라 혹은 자바일 경우

스칼라
sbt

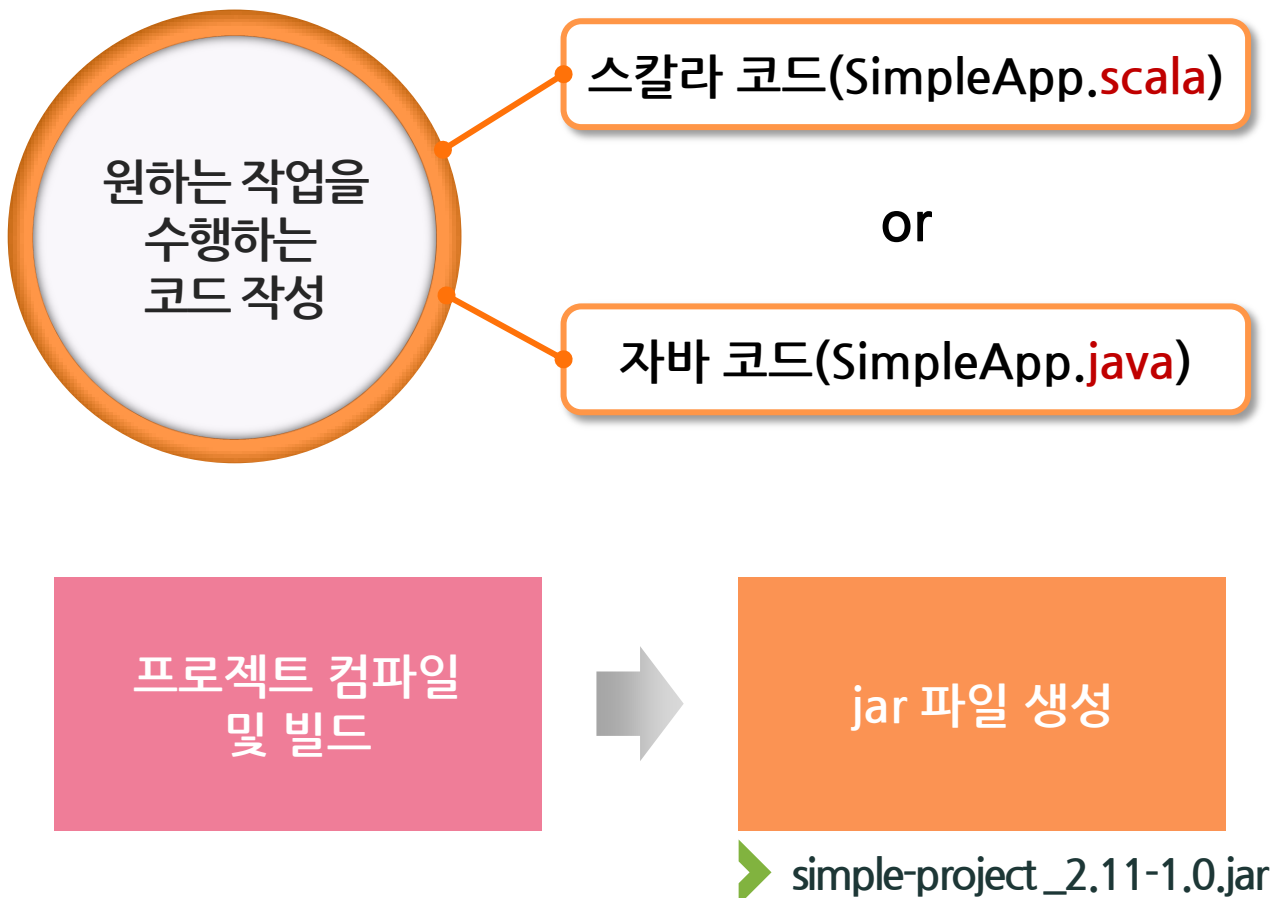
자바
eclipse Maven

프로젝트 생성에 이용

스파크 응용 프로그램

5 spark-submit을 통한 실행

② 스칼라 혹은 자바일 경우



어플리케이션 실행

```
$ YOUR_SPARK_HOME/bin/spark-submit ₩
--class "SimpleApp" ₩
--master local[4] ₩
simple-project_2.11-1.0.jar
```