스파크 대화용 프로그램

1 스파크 특징

인-메모리 기반의 대용량 데이터 고속 처리 엔진

자바, 파이썬, 스칼라, R 인터페이스 제공

스파크 Standalone(단일 노드), 하둡 YARN 및 mesos 등의 클러스터 리소스 관리자를 통해 다양한 환경에서 구동

2 스파크 장점 및 활용

① 스파크의 장점

하둡의 맵리듀스 작업처리 속도보다 100배 빠른 성능

8000개 이상의 노드 추가 가능한 확장성 확보

HDFS, 카산드라, Hbase, S3 등 다양한 데이터의 활용 가능

스파크 대화용 프로그램

2 스파크 장점 및 활용

② 스파크의 활용

전자상거래 데이터 수집을 통한 고객 매출 최적화

웹사이트의 추천 엔진 지원 모델 구축

사용자와 상호작용할 수 있는 예측 모델 구성

스파크 대화용 프로그램

3 스파크 대화용 프로그램

Scala, pyspark 사용

스파크 프로그램을
대화형으로 실행할 수 있음

1 스파크 프로그램 실행

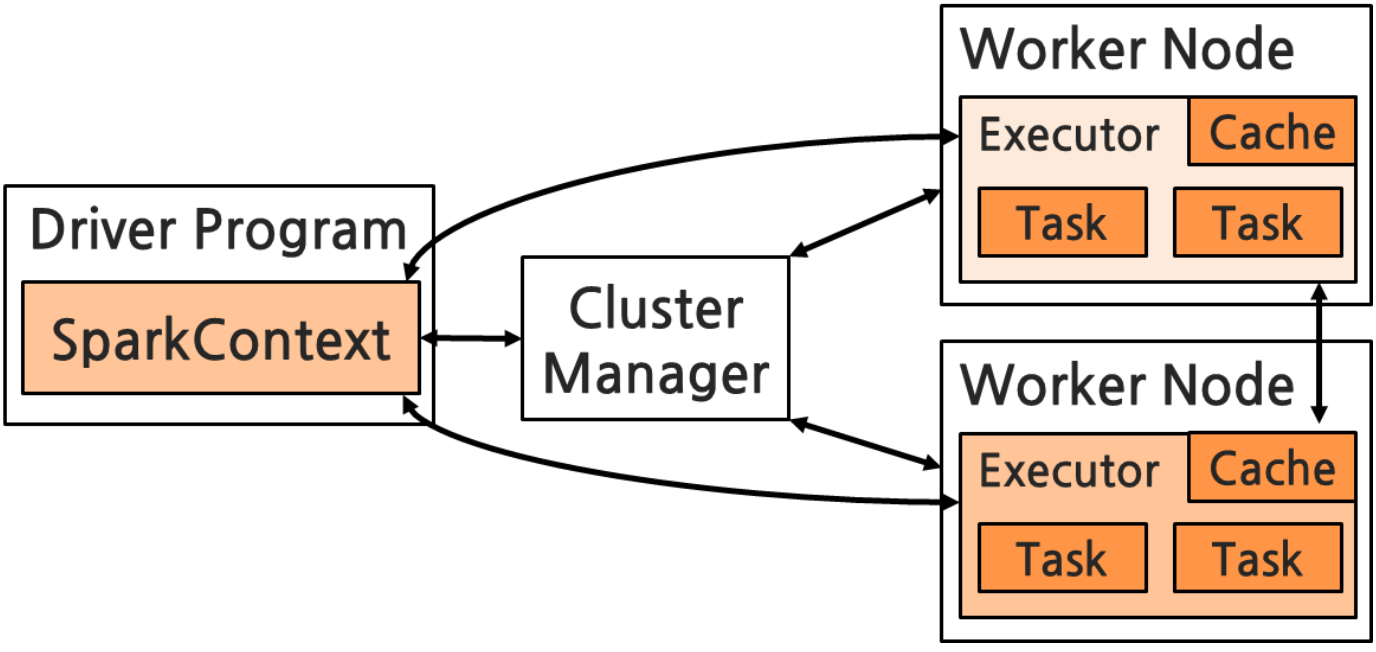
드라이버 프로그램
(Spark Driver Program)

- 어플리케이션 마스터에 의해 여러 개의 작업으로 나누어짐

작업자 노드에 있는
실행 프로세스
(Spark Executor)에서
실행

스파크 대화용 프로그램

3 스파크 대화용 프로그램

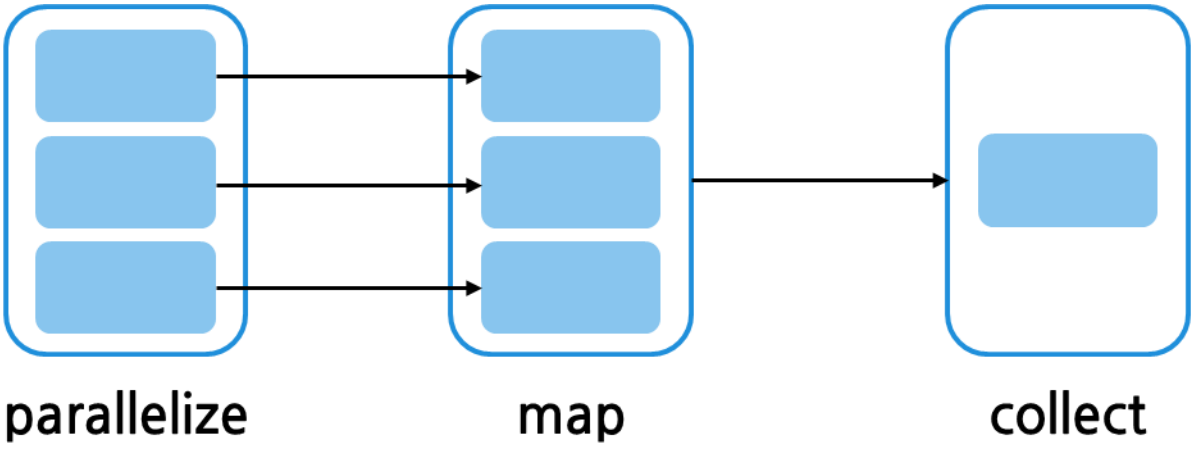


스파크 대화용 프로그램 예시

스파크 대화용 프로그램

4 스칼라를 사용한 맵 예제

① RDD 연산



RDD 연산의 예시

② 스칼라에서의 기본적인 맵 예제

```
// Basic map example in scala

scala> val x = sc.parallelize(List("spark", "rdd", "example", "sample", "example"), 3)
scala> val y = x.map(x => (x, 1))
scala> y.collect
res0: Array[(String, Int)] = Array((spark,1), (rdd,1), (example,1), (sample,1), (example,1))
```

스파크 대화용 프로그램

4 스칼라를 사용한 맵 예제

3 스칼라에서의 짧은 syntax 표현

```
// rddy can be re writen with shorter syntax in scala as

scala> val y = x.map(_, 1))
scala> y.collect
res0: Array[(String, Int)] = Array((spark,1), (rdd,1), (example,1), (sample,1), (example,1))
```

4 문자열 길이 구하기

```
// Another example of making tuple with string and it's length

scala> val y = x.map((x, x.length))
scala> y.collect
res0: Array[(String, Int)] = Array((spark,5), (rdd,3), (example,7), (sample,6), (example,7))
```

스파크 대화용 프로그램

5 파이썬을 사용한 맵 예제

① 파이썬에서의 기본적인 맵 예제

- ✓ 런타임에 생성해서 일시적으로 사용하고 버릴 수 있는 익명 함수
- ✓ Lambda 함수에 대한 설명과 형식

```
// Bazic map example in python

>>> x = sc.parallelize(["spark", "rdd", "example", "sample", "example"], 2)
>>> x = x.map(lambda x: (x,1))
>>> y.collect()
[('spark', 1), ('rdd', 1), ('example', 1), ('sample', 1), ('example', 1)]
```

② 문자열 구하기

```
// Another example of making tuple with string and it's length

>>> y = x.map(lambda x: (x,len(x)))
>>> y.collect()
[('spark', 5), ('rdd', 3), ('example', 7), ('sample', 6), ('example', 7)]
```