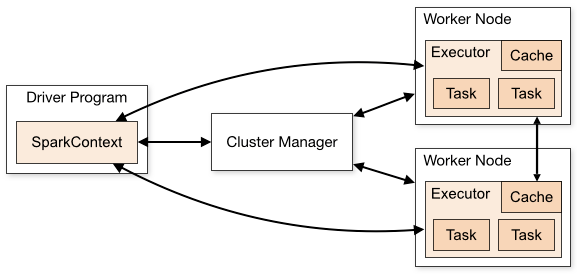
Spark Components

* Driver Program, SparkContext, Cluster Manager, Worker Node, Executor, Task, Cache
* 

Driver Program

* Main 함수를 갖고 있는 프로세스
* Spark-submit을 통해 구현한 코드 제출 🡪 코드에서 SparkContext 객체 생성 🡪 RDD 생성
* 제출한 Application을 task(실제 수행 단위)로 변환 🡪 Task를 묶어 Executor로 전달
* Executor는 받은 task를 RDD에 저장하고 처리
* 연산들의 관계를 DAG(Directed Acyclic Graph)를 생성 🡪 DAG를 물리적인 실행 계획으로 변환
* 최적화를 거쳐 여러 개의 stage로 변환하고 각 stage는 여러 개의 task로 구성

SparkContext

* Driver Program에 의해 생성되고 Cluster Manager와 연결된다.
* Spark는 클러스터내부에 있는 각 Worker Node로부터 Executor를 얻는다.

Executor

* Computation과 data를 저장하는 역할을 하는 process
* Application의 lifecycle과 동일하게 수행된다.
* Multi threads에서 task를 수행하고 수행 결과를 Driver Program으로 전송하는 역할
* application에서 cache하는 RDD를 저장하기 위한 메모리 공간을 제공
* 여러 tasks를 스케쥴링하는 역할을 한다.

Spark Application 실행 동작 순서 흐름

1. 사용자가 Spark-submit을 사용해 작성한 Application 실행
2. Spark-submit은 Driver Program을 실행하여 main method 실행
3. Driver에서 생성된 SparkContext는 Cluster Manager로부터 Executor실행을 위한 리소스 요청
4. Cluster Manager는 Executor를 실행
5. Driver Program은 Application을 Task 단위로 나누어 Executor에게 전송
6. Executor는 Task를 실행
7. Executor는 Application이 종료되면 결과를 Driver Program에게 전달하고, Cluster Manager에 리소스 반납

