

스파크 드라이버(driver)란, 프로그램의 main() 메소드가 실행되는 프로세스임.

드라이버는 ^①Spark Context 를 생성하고 ^②RDD를 만들고 ^③transformation , action 등을 실행하는 사용자 코드를 실행한다.

driver process = Spark Session. 즉 Driver 란 Spark Session 이다.

df,ds,udf 를 생성하고
사용자 입력에 응답 하며
애플리케이션 정보 유지 관리한다.

① 하나의 스파크 애플리케이션 처리

② 프로그램이나 입력에 ^(= Job) 대한 응답

③ 익스큐터 작업과 관련된 분석, 배포 및 스케줄링 역할 수행

SparkSession 이라고 하며 스파크 애플리케이션의 "엔트리 역할" 맡은 object

↑ 'YARN'의 'Resource Manager'와 비슷함

Cluster 모드란?

'YARN'의 Application Master.

드라이버 프로세스가 Cluster 내의 Application Master 에서 실행되는 모드
driver가 YARN 컨테이너에서 동작하기 때문에 driver에 장애가 발생할 경우 다른 노드에서 driver가 재실행
주로 Production 환경에 적합

Client 모드란?

드라이버 프로세스가 Cluster 와 관련 없는 다른 외부 서버(ex, 내 방 데탑)에서 실행되는 모드
주로 개발과정에서 대화형 디버깅을 할 때 사용(스파크 쉘)
따로 지정하지 않으면 기본으로 선택되는 모드임

	yarn cluster	yarn client	spark standalone
Driver runs in	Application master	client	client
Who request resources?	Application master	Application master	client
Who starts executor processes?	YARN ResourceManager and NodeManager	YARN ResourceManager and NodeManager	Spark Master and Workers
Supports Spark Shell	No	Yes	Yes