
Programming with RDDs

이 승우

More Information
Please Visit This [Link](#).

1. 왜 RDD를 사용하는가?

사실은 SlideShare 에서 발견한것을 요약하였다

1. RDD는 Resilient Distributed Datasets 의 약자.
 2. RDD는 RAM 에서 Read-Only로 동작하기 때문에 빠르다.
 3. Read-Only로 동작하다 보니 Storage → RDD, RDD → RDD 로 변환이 가능
 4. (고치고 싶은데)만들어진 이래 고쳐진 적이 없는가? → 그럼 계보를 등록하지 (Lineage) → 문제가 덜 생김.
 5. 코딩을 하는것은 실제 계산을 하는것이 아닌, Lineage(계보를 쌓아가는 작업) 을 하는것.
-

More Information
Please Visit This [Link](#).

2. Lazy Execution

어떻게 만들어지는지 대략의 Execution Plan 을 만들어서 실행되므로 , 자원이 배치될 혹은 배치될 상황을 미리 고려해서 최적의 코스를 돌 수 있다!

그리고 리니지 들만 생성되므로 Action 들에 해당하는 명령어가 붙으면 그제서야 실행됨.

3. Getting Started with Docker

1. Digital Ocean Server 하나를 판다 (1G RAM, 1C CPU)
 2. SSH Server 주소를 알아서 접속해서
 3. Docker 명령어를 입력한다
 4. `docker run -d -p 9999:8888 -e GRANT_SUDO=yes --name psy_spark jupyter/all-spark-notebook`
 5. 그리고 자신의 ip:9999 로 노트북 서버에 접속을 하면 됩니다.
-

Let's Code!