

본 과제에서는 AWS 에서 제공되는 Elastic Map Reduce (EMR) 서비스를 활용하여 하둡/스파크 클러스터를 동작시킨 후 HDFS 실습을 진행하고자 합니다. 아래 요구사항대로 진행해주세요.

1. 이메일로 초대된 AWS Educate 서비스를 통해서 본 수업을 위해서 개설된 클래스에 참여해주세요. (슬랙에 공유된 문서 참조하기 - AWSEducateBigDataCourse.pdf). **만약 awseducate 서비스에 처음으로 가입을 해야 한다면 생년 월일 및 졸업년도를 제대로 써주세요. 유효하지 않은 값으로 설정시에 가입이 거부 될 수 있습니다.**
2. EMR 서비스를 통하여 하둡 클러스터를 생성후에 마스터 노드에 SSH 를 활용하여 접근하세요. (슬랙에 공유된 문서 참조하기 - EMR-AWS-Educate.pdf)
3. HDFS 실습 문서를 (HDFS-Exercise.pdf) 따라서 실행하면서 각 페이지에 나와 있는 HDFS 명령어를 실행한 화면을 캡처해서 보고서를 작성해주세요. 실습 페이지마다 보고서에 작성할 중요 내용이 있을 경우 해당 내용을 보고서에 추가로 작성해주세요.
4. 본 수업에서 HDFS 의 블록 사이즈 설정에 대한 이야기가 있었습니다. 블록사이즈 설정에 따른 성능 특성을 HDFS 의 NameNode 오버헤드 관점 과 MapReduce 작업에서 성능 특성과 연관해서 보고서에 서술해주세요.
5. 모든 실습을 종료하고 (AWS EMR 클러스터 종료) AWS Educate 페이지를 통해서 현재 남은 Credit 을 확인하세요. \$50 의 크레딧중 해당 Credit 이 어떤 방식으로 어떤 서비스에 소모 되었는지 계산해서 보고서에 작성해주세요.