

# 캡스톤디자인 면담 확인서

팀원	2018102238 조인화			
주제	MLOps 를 위한 클라우드 환경에서의 데이터파이프라인 구축 및 ETL 작업 자동화 구축			
면담일시	2023. 5. 22.	지도교수	허의남 (인)	
면담내용	<p>현재의 구조는 데이터 ETL 과정에서 데이터를 분할하고 각각 분할한 데이터를 전처리하기 위해 Airflow Scheduler 가 Worker 를 Pod 형태로 생성하고, Worker 가 Hadoop/Spark Pod 를 생성하여 작업을 지시하게 된다. 해당 과정에서 Harbor 에서 Hadoop/Spark 이미지를 가져와 Pod 를 생성할 때 많은 시간이 소요될 것으로 보인다. Worker 컨테이너의 Layer 를 변경하여 실행하는 방법을 고민해본다.</p>			

팀원	2018102238 조인화		
주제	MLOps 를 위한 클라우드 환경에서의 데이터파이프라인 구축 및 ETL 작업 자동화 구축		
면담일시	2023. 5. 22.	지도교수	허의남 (인)
면 담 내 용	<p>현재의 구조에서 Airflow Scheduler 가 각각의 노드에 데이터를 뿌려줄 수 있는지 조사를 해보는 것이 필요하다. 만약 불가능하다면 Kafka 와 같은 플랫폼 도입을 고려한다.</p> <p>현재의 구조에서 분산 학습까지 진행하게 된다면, Tensorflow 의 병렬 처리와 해당 구조를 비교한다. 그리고 Spark Pod 에서 GPU 가 필요한 점을 고려한다.</p>		