1차년도 주요 결과물

(과제명) 대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW 프레임워크 개발 (과제번호) 2021-0-00077

> • 결과물명 : 에너지 인프라 응용별 안전 AI 분석 연계 SW 프레임워크 설계서

• 작성일자 : 2021년 12월 1일

과학기술정보통신부 SW컴퓨팅산업원천기술개발사업 "1차년도 주요 결과물"로 제출합니다.

수행기관	성명/직위	확인
한국전자기술연구원	최효섭/책임연구원	U/S

정보통신기획평가원장 귀하



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

사 용 권 한

본 문서에 대한 서명은 한국전자기술연구원 내부에서 본 문서에 대하여 수행 및 유지관리의 책임이 있음을 인정하는 것임.

본 문서는 작성, 검토, 승인하여 승인된 원본을 보관한다.

작성자 :	최우정	일자 :	2021. 09. 30
검토자 :	김창우	일자 :	2021. 11. 01
 숭인자 :	최京선	일자 :	2021, 12, 01

제・개정 이력

버전	변경일자	제.개정 내용	작성자
1.0	2021-09-30	최초 등록	최우정



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

목 차

1. 개요	1
1.1 목적1.2 범위1.3 시스템 개요1.4 관련 계획 및 표준	
2. 요구사항 추적 매트리스	4
3. SW 아키텍처 설계	5
3.1 SW 아키텍처 개요 3.2 SW 아키텍처 구성도 3.3 SW 아키텍처 구성 요소	
4. 시스템 아키텍처 설계	9
5. 용어	12
6 Reference	13



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

1. 개요

1.1 목적

본 명세서의 목적은 대규모 분산에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 안전SW 프레임워크에 다양한 기능이 추가됨에 따라 관리가 용이해야 하며 개발 자가 프레임워크의 기능부터 설계내용까지 쉽게 이해할 수 있도록 한다.

- 안전SW 프레임워크의 상세 설계를 위한 명세를 제공함
- 안전SW 프레임워크 검증을 위한 평가 기준을 제공함
- 안전SW 프레임워크 사용 매뉴얼 작성 시 기초자료로 활용함

1.2 범위

본 프레임워크 설계서는 지능형 안전SW 프레임워크의 개발에 필요한 소프트웨어와 시스템설계 명세를 기록한다. 3장에서는 데이터의 이동부터 결과의 시각화까지의 SW 아키텍처 설계에 대해 명세하며, 4장에서는 가상 환경과 실제 환경의 시스템 아키텍처 구성에 대해서 기술한다. 본 설계서는 프레임워크의 시스템과 아키텍처의 설계에 대한 내용을 담으며 프레임워크 내에서 제공하는 안전 솔루션 컨테이너 패키지의 기능에 대해서는 간략하게만 소개한다.

1.3 시스템 개요

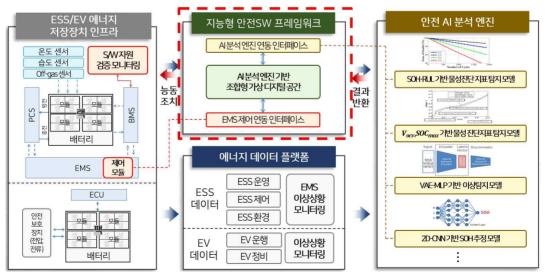


그림 1 대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW 프레임워크 구조



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

지능형 안전SW 프레임워크는 개발 솔루션을 실제 ESS 운영 사이트와 가상 디지털 공간에 실증하여 안전 소프트웨어를 구조화하고, 에너지 인프라 운영 소프트웨어를 최적화한다. 안전SW 솔루션 활용의 손쉬운 확장 및 배포를 위해 솔루션 컨테이너 패키지를 통해 프레임워크 사용자로 하여금 솔루션을 선택 활용하도록 개발하고자 한다.

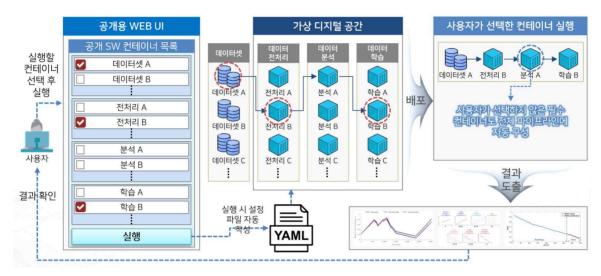
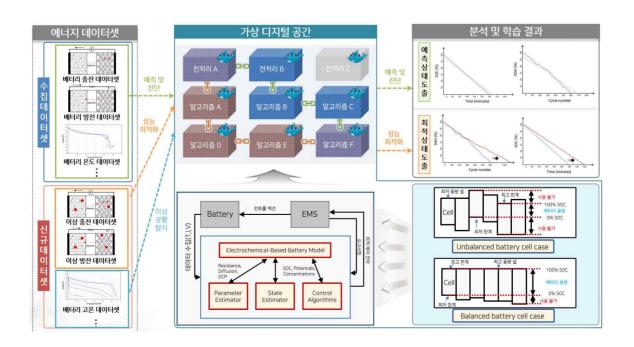


그림 2 안전 소프트웨어 솔루션의 컨테이너 기반 패키지 제공 구조

안전 소프트웨어 솔루션에서 제공하는 패키지란 ESS 시스템에서 나오는 데이터의 전처리부터 안전에 관한 이상감지, 진단, 예측 등을 수행하며 지표를 생성하는 분석 엔진들을 각각의 도커 이미지들을 말한다. ESS 데이터를 가지고 있는 연구개발자는 지능형 안전SW 프레임워크를 통해 위에서 설명한 분석 모델들을 가상 공간에서 실행시킬 수 있다.





대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

1.4 관련 계획 및 표준

본 설계서는 아래 계획 및 표준을 참고한다.

구분	식별자	세부 내용	설명
설계서	ISO/IEC 9126	9126-1 (품질 모델) 9126-2 (외부 품질) 9126-3 (내부 품질) 9126-4 (사용 품질)	품질 특성 및 측정기준을 제시 소프트웨어의 기능성, 신뢰성, 사용성, 효율성, 유지보수 용이성, 이식성
	ISO/IEC 14598	14598-1 (개요) 14598-2 (계획과 관리) 14598-3 (개발자용 프로세스) 14598-4 (구매자용 프로세스) 14598-5 (평가자용 프로세스) 14598-6 (평가 모듈)	ISO 9126에 따른 제품 평가 표준: 반복성, 공정성, 객관성, 재생산성
	ISO/IEC 12119	소프트웨어 패키지 -제품설명서 -사용자문서 -프로그램과 데이터	패키지 SW 품질 요구사항 및 테스트
계획서	-	지능형 안전 SW 프레임워크 요구사항 정의서	대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위해 개발된 솔루션을 가상 환경에서 시뮬레이션 분석을 도와주는 프 레임워크 요구사항을 정의한 문서



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

2. 요구사항 추적 매트릭스

ESS 지능형 안전SW 프레임워크 요구사항정의서에 기반한 요구사항 목록을 기록한다.

NO	요구사항 명
1	가상 시뮬레이션 기술 내 안전AI분석 엔진 연동
2	에너지 데이터의 고속 이벤트 복합 분석
3	가상화 컨테이너 기반 데이터 분석 파이프라인 기능
4	가상화 컨테이너 간 선택형 복합 연계
5	안전SW 프레임워크 및 테스트/실증사이트 간 연동
6	에너지 인프라 테스트베드 검증 수행 및 결과 확인
7	안전SW 검증단계 검증항목 검증절차의 규격
8	안전사고 유발 SW 취약점 기반 검증 규격
9	ESS의 안전 SW 검증 프로세스, 검증 항목 및 지표 규격
10	분석 및 학습을 위한 데이터 처리 크기
11	수집 데이터셋 호환 및 데이터 연동 플로우 처리율
12	데이터 입력 규격화
13	안전SW 패키지의 배포 자동화
14	가상환경에 ESS의 설계가 가능
15	웹 상에서 파이프라인의 시각화
16	안전SW 패키지 분석 결과의 시각화
17	기준 데이터의 제공
18	안전 SW 분석 패키지의 조합이 가능



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

3. 프레임워크 아키텍처 설계

3.1 프레임워크 아키텍처 구성도

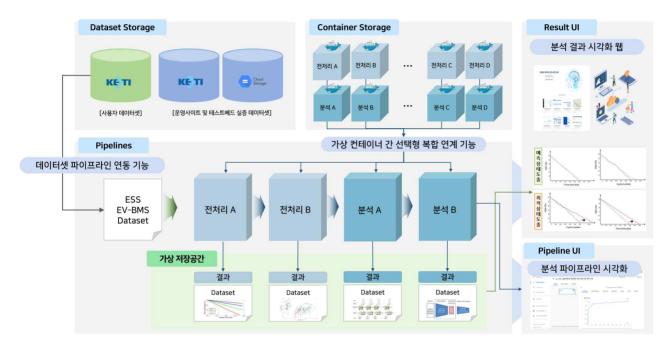


그림 4 지능형 안전 SW 프레임워크 아키텍처

3.2 프레임워크 아키텍처 구성 요소

1) 데이터 저장소

데이터셋 저장소는 분석에 사용할 데이터를 저장하는 공간으로 기본 데이터 저장소와 사용자데이터 저장소가 있다.

(1) 기본 데이터 저장소

기본 데이터 저장소에는 에너지 데이터 플랫폼에서 제공하는 실제 운영사이트의 ESS 데이터와 데 스트베드에서 실험을 통해 생성한 다양한 ESS 위험 상황의 데이터셋이 존재한다.

(2) 사용자 데이터 저장소

사용자 데이터 저장소는 지능형 안전SW 프레임워크의 사용자가 분석을 원하는 임의의 데이터를 이용하여 시뮬레이션을 진행할 수 있게 데이터를 업로드하는 공간입니다. 분석에 필요로 하는 필수 속성, 데이터 포맷, 파일명 등의 양식을 제공한다.

2) 컴포넌트 저장소

컴포넌트 저장소에는 에너지 데이터 플랫폼에서 제공하는 다양한 전처리 엔진과 분석 엔진이 저장되어 있다. 컴포넌트는 가상 컨테이너 이미지의 형태로 패키징된 독립적인 코드 조각으로



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

공개 SW의 도커 허브에 존재한다. 가상화 컨테이너는 Json 포맷의 데이터셋을 읽고 반환할 수 있으며, 동일한 데이터 입출력 방법과 실행 방법을 컴포넌트 스펙 YAML 파일에 정의하여 'Kubeflow Pipeline'을 통해 유연하게 실행시킬 수 있다. 이에 프레임워크 사용자는 개발된 컴포넌트를 선택하여 추가적인 개발 없이 그대로 사용한다.

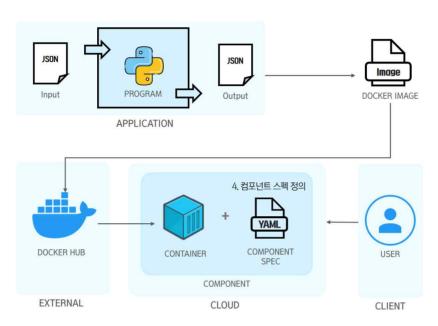


그림 2. 분석 컴포넌트 생성 구성도

그림 6 입출력 데이터셋, 컴포넌트 스펙 규격 샘플



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

3) Kubeflow

Kubeflow는 파이프라인이라는 ML 워크플로우를 구축하고 배포하기 위해 제공되는 플랫폼이다. Kubeflow를 사용하여 가상 환경에서의 분석이 용이해지며, 데이터 분석과 모델 개발에 집중하여 복잡성을 줄인다. 이를 클라우드나 On-Prem 환경으로 제공하여 개인 개발 환경에 상관 없이 동일한 머신러닝 플랫폼을 이용할 수 있게 한다.

(1) Kubeflow Pipeline

파이프라인은 머신러닝 워크플로우의 코드화된 표현으로 각 노드에 컴포넌트(컨테이너화된 프로세스)가 있는 방향성 그래프입니다. 파이프라인을 실행하게 되면 각 컨테이너는 Kubernetes 스케줄링에 따라 종속성을 고려하여 클러스터 전체에서 실행됩니다. 각각의 컴포넌트들을 규격화하여 저장되어 간단하게 컴포넌트를 재사용, 공유, 교체할 수 있습니다.

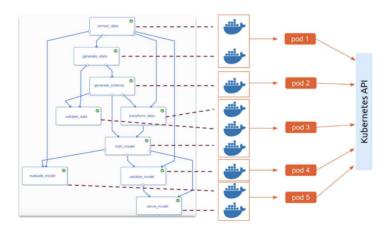


그림 7 Kubeflow Pipeline

(2) Pipeline UI

Kubeflow의 대시보드에서 파이프라인 UI를 통해 기본적인 분석 결과를 확인할 수 있으며, 매개 변수의 조정과 더 많은 실행을 생성할 수 있습니다.

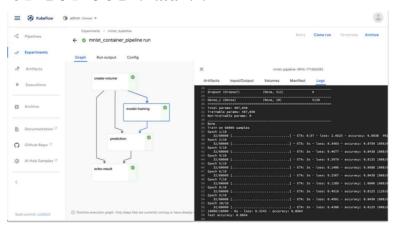


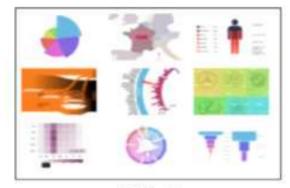
그림 8 Kubeflow Pipeline UI



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

4) 결과 시각화 웹

다양한 안전 AI 분석엔진의 출력결과를 웹 화면으로 시각화하여 전달합니다. 'AMCharts'는 Bar, Pie, Donut, Line, Area, Bubble, Scatter, Candle, Gant, Radar, Map 등의 다양한 유형의 데이터 시각화 차트를 제공합니다. 'DECK.GL'은 Map 특화 데이터 시각화 라이브러리이다. CSV, JSON, Array, Image, GeoJson, 3D Tile 등의 다양한 형식의 데이터를 로드할 수 있고, 국내 지역 시각화도 가능하다. API/Resource Server와 API Portal에서 API Gateway에서 제공하는 모니터링, 로그 데이터와 구축한 API를 결합하고, Javascript 데이터 시각화 라이브러리를 활용해데이터를 시각화한다.







DECK.GL

그림 9 웹 시각화 도구

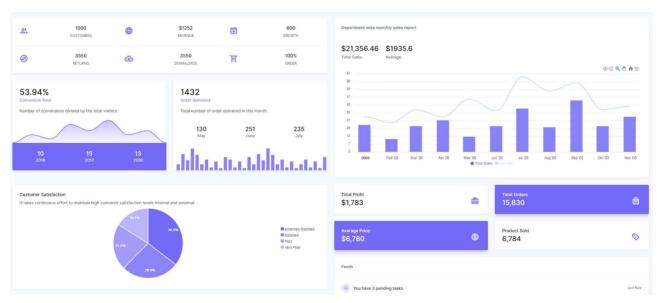


그림 10 웹 대시보드 화면



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

4. 시스템 아키텍처 설계

4.1 네트워크 구성도

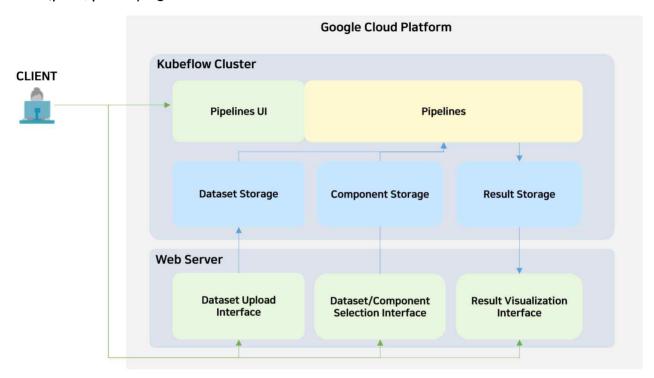


그림 11 프레임워크 네트워크 구성도

5. 주요 기능 설계

1) 데이터셋 업로드 서버

프레임워크 사용자는 분석을 원하는 임의의 데이터를 업로드할 수 있는 기능을 구현한다. 파이썬 http 서버를 통해 데이터를 저장소로 전달받는다.

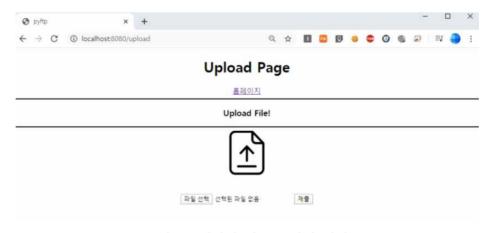


그림 12 데이터 업로드 파일 서버



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

2) 데이터셋 선택 인터페이스

분석에 사용할 데이터셋을 선택할 수 있다. 데이터셋 선택 인터페이스를 웹 환경에서 구축한다.

- (1) 테스트베드 데이터셋
- (2) 운영사이트 데이터셋
- (3) 사용자 임의 데이터셋

3) 컴포넌트 선택 인터페이스

분석에 사용할 컴포넌트를 선택할 수 있다. 분석을 진행할 여러 컴포넌트를 선택하는 인터페이스를 웹 환경에서 구축한다. 선택이 완료되면 파이썬 서버를 통해 큐브플로우 파이프라인 파일을 자동으로 생성하여 클러스터에서 실행시킨다.

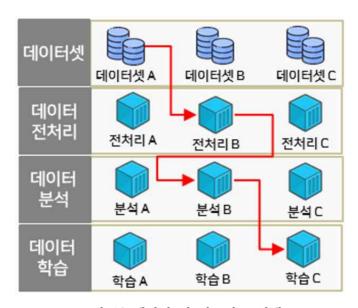


그림 13 데이터 및 컴포넌트 선택

4) Kubeflow 파이프라인 생성 기능 개발

앞서 선택한 데이터셋과 컴포넌트들을 kubeflow 서버 위에서 실행시키기 위해 'pipeline.py'파일을 생성한다. 파이프라인 이름 구성, 함수 생성, 각 파이프라인에서 사용되는 name과 image 그리고 각 컨테이너 환경에서 사용되는 argument 내용을 작성한다. 'pipeline.py'파일을 kubeflow 클러스터 위에 저장하여 분석을 진행한다.



대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

```
load_data.py
                  pipeline.py ×
kubeflow > iris > ♥ pipeline.py > ♡ soojin_pipeline
       import kfp
      import kfp.components as comp
      from kfp import dsl
      @dsl.pipeline(
          name='soojin-iris',
           description='soojin iris test'
      def soojin_pipeline():
          add_p = dsl.ContainerOp(
               name="load iris data pipeline",
               image="lsjsj92/soojin-iris-preprocessing:0.5",
               arguments=[
                   '--data_path', './iris.csv'
               file_outputs={'iris' : '/iris.csv'}
          ml = dsl.Container0p(
              name="training pipeline",
              image="lsjsj92/soojin-iris-train:0.5",
               arguments=[
                   '--data', add_p.outputs['iris']
          ml.after(add_p)
       if __name__ == "__main__":
          import kfp.compiler as compiler
           compiler.Compiler().compile(soojin_pipeline, __file__ + ".tar.gz")
```

그림 14 pipeline.py 예시

5) 결과 시각화 기능

Kubeflow 컴포넌트 내 콘솔로 출력되는 결과 이외의 파일로 저장된 데이터를 웹페이지를 통해 보여준다. 분석 결과 데이터를 'AMCHART'나 'DECK.GL'을 이용하여 기본적인 차트로 나타내결과 해석을 용이하게 도와준다. 나아가 Redash를 이용하여 사용자가 원하는 차트 및 그래프로 시각화하여 결과를 들여다볼 수 있게 한다. Redash는 데이터 조회/시각화를 편하게 해주는 툴로써데이터를 다양한 그래프로 보여줄 수 있다.

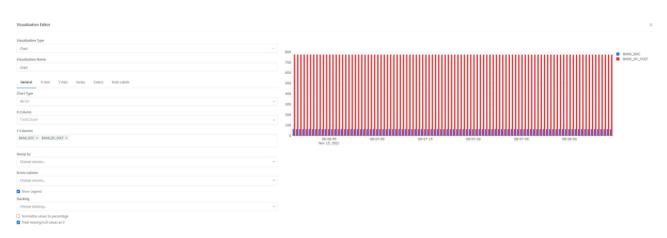


그림 15 Redash 웹기반 데이터 시각화 도구

에너지 인프라 응용별 안전 AI 분석 연계 SW 프레임워크 설계서



과제명

대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

6. 용어

- ESS: 에너지를 효율적으로 사용할 수 있도록 저장, 관리하는 시스템이다. 주로 발전소, 송배전 시설, 가정, 공장, 기업 등에서 활용된다. 주로 생산된 전기를 '전력 계통'에 저장해놓고 전기가 가장 필요한 시기에 공급해 에너지 효율을 높이는 데에 사용되는 시스템이다.
- EMS: EMS(Energy Management System)는 에너지 매니지먼트 시스템 혹은 에너지 관리 시스템이라 불리는 것으로 전기, 열, 가스 등의 에너지의 시각화나 설비의 최적 운용 등을 실현하는 시스템이다.
- 도커: 도커(Docker)는 리눅스의 응용 프로그램들을 프로세스 격리 기술들을 사용해 컨테이 너로 실행하고 관리하는 오픈 소스 프로젝트이다.
- 도커 허브: 용량이 큰 공개 도커 이미지를 서버에 저장하고 관리해주는 웹 플랫폼
- Kubeflow: Kubernetes + ML flow를 합한 의미로, 파이프라인이라는 ML 워크플로우를 구축하고 배포하기 위해 제공되는 플랫폼으로 개발, 테스트 및 프로덕션 수준 서비스를 위해 다양한 환경에 ML 시스템을 배포하려는 ML 엔지니어 및 운영 팀을 위한 것
- YAML: XML, C, 파이썬, 펄, RFC2822에서 정의된 e-mail 양식에서 개념을 얻어 만들어진 '사람이 쉽게 읽을 수 있는' 데이터 직렬화 양식이다. 2001년에 클라크 에반스가 고안했고, Ingy dot Net 및 Oren Ben-Kiki와 함께 디자인했다.
- JSON: JSON(제이슨, JavaScript Object Notation)은 속성-값 쌍(attribute-value pairs and array data types (or any other serializable value)) 또는 "키-값 쌍"으로 이루어진 데이터 오브젝트를 전달하기 위해 인간이 읽을 수 있는 텍스트를 사용하는 개방형 표준 포맷이다.





대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW프레임워크 개발

7. Reference

- Google Cloud Storage 문서 https://cloud.google.com/storage/docs
- 큐브플로우 문서 https://www.kubeflow.org/docs/
- python htttp file server docs https://docs.python.org/ko/3/library/http.server.html
- deck gl 문서 https://deck.gl/docs
- amCharts4 문서 https://www.amcharts.com/docs/v4/