

# M6. 데이터수집 시스템 만들기

## 1. 시스템 설계 및 윈도우 스케줄러 소개

# 커리큘럼

## 과정 목표

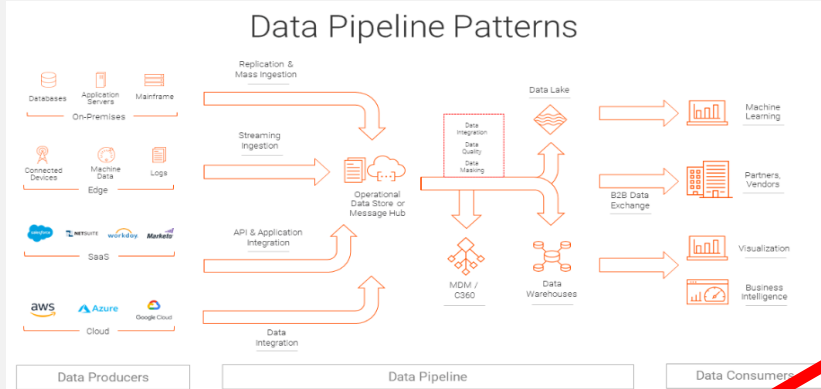
- 01 온라인(웹) 상의 다양한 데이터를 수집하기 위해 데이터의 형태와 웹의 기본구조를 이해한다.
- 02 크롤링을 위해 파이썬의 다양한 라이브러리의 기능과 역할을 학습한다.
- 03 실제 온라인에서 제공되는 다양한 형태의 데이터들을 원하는 방식으로 수집할 수 있도록 코드를 작성한다.

| 모듈 명                        | 주요 내용  |
|-----------------------------|--|
| M1.<br>데이터의 형태              | <ul style="list-style-type: none"> <li>데이터를 정의하고, 강의에서 활용할 데이터에 대한 유형을 소개</li> <li>JSON, CSV, TSV, XLSX, 등 데이터의 형태를 소개하고 간단한 활용법 학습</li> <li>Database에 대한 소개 및 SQLite를 활용한 간단한 활용법 숙지</li> </ul> |
| M2.<br>웹의 기본구조              | <ul style="list-style-type: none"> <li>웹의 개념과 간단한 역사 및 작동원리에 대한 내용을 파악</li> <li>실제 다양한 웹페이지를 통해 웹을 구성하고 있는 요소(HTML, Javascript, JSON 등)들을 학습</li> </ul>  |
| M3.<br>크롤링을 위한<br>파이썬 라이브러리 | <ul style="list-style-type: none"> <li>크롤링을 위한 다양한 파이썬 라이브러리에 대한 소개</li> <li>주피터노트북을 활용하여 각 라이브러리에 대한 실습 진행</li> </ul>   |
| M4.<br>웹페이지를 활용한<br>데이터 수집  | <ul style="list-style-type: none"> <li>실제 동작하는 웹페이지를 분석하고 수집하여, 원하는 형태의 데이터로 변환하는 실습 진행</li> <li>대상 : 네이버뉴스, Dart, CoinmarketCap 등</li> </ul>  |
| M5.<br>API를 활용한<br>데이터 수집   | <ul style="list-style-type: none"> <li>공공기관이나 민간기업에서 제공하고있는 OpenAPI를 활용하여 데이터를 수집하는 실습 진행</li> <li>대상 : 공공데이터, SNS API, FinHub 등</li> </ul>  |
| M6.<br>데이터 수집<br>시스템 만들기    | <ul style="list-style-type: none"> <li>상기 과정에서 작성된 파이썬 코드를 활용하여, 정기적인 수집 시스템을 만드는 실습 진행</li> </ul>   |

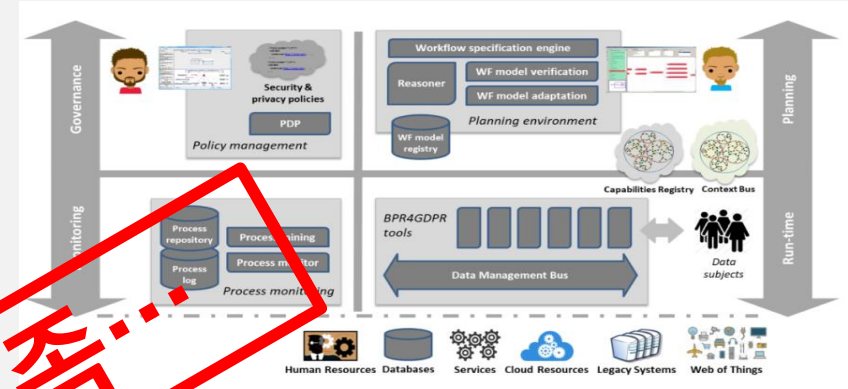
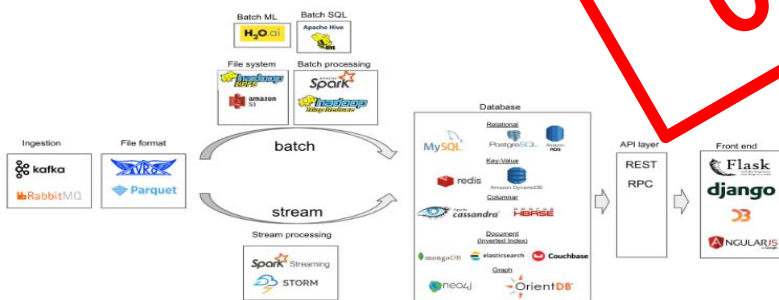
\*상기 커리큘럼의 내용 및 교육시간은 기관 및 교육과정의 사정에 따라 변동이 가능합니다.



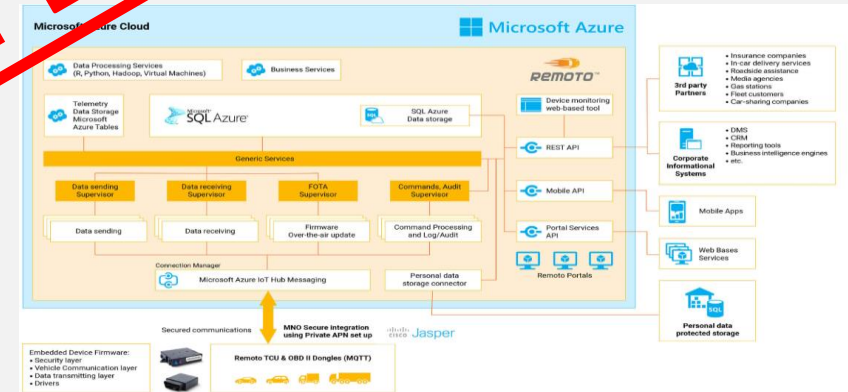
## 시스템설계? 내가?



### Data engineering pipeline with choices of technologies for each stage



이건 좀...





설계한 시스템 도면을 따라 레고블럭을 쌓듯이!

Fail!



내가 해야하는 것!

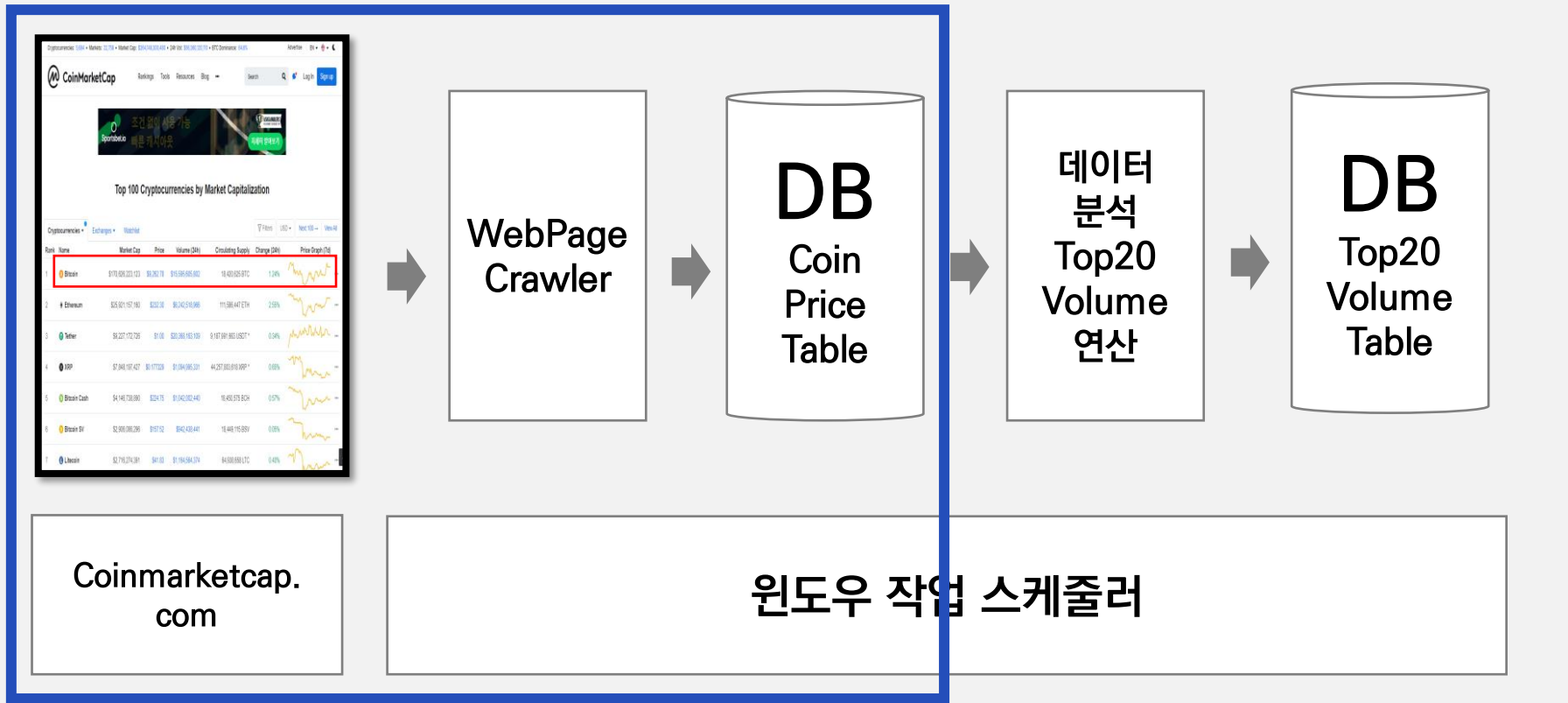
Success!



내가 할 수 있는 것!



# 우리가 만들 시스템은?





준비물?

Window10 Python3.X SQLite

그리고...

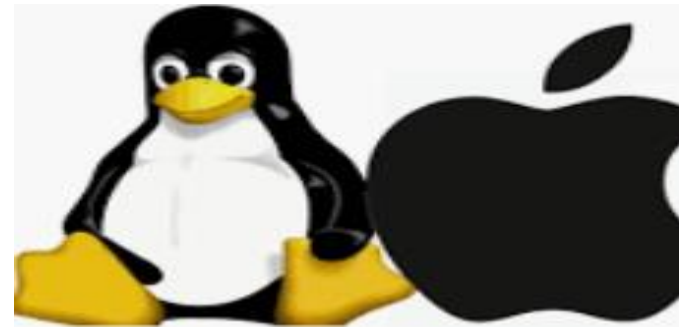
윈도우 작업 스케줄러



## 작업 스케줄러는 무엇?



작업 스케줄러



<https://ko.wikipedia.org/wiki/Cron>



# 작업 스케줄러는 어떻게 쓸까?

작업 스케줄러 > 작업만들기 > 한번 > 작업반복 간격 : 무제한 > pythonw.exe 실행파일.py > 확인

