

Part. 01

데이터 아키텍처

자동화의 이해

FASTCAMPUS
ONLINE
데이터 엔지니어링

강사. 한승수

I 데이터 프로세싱 자동화란?

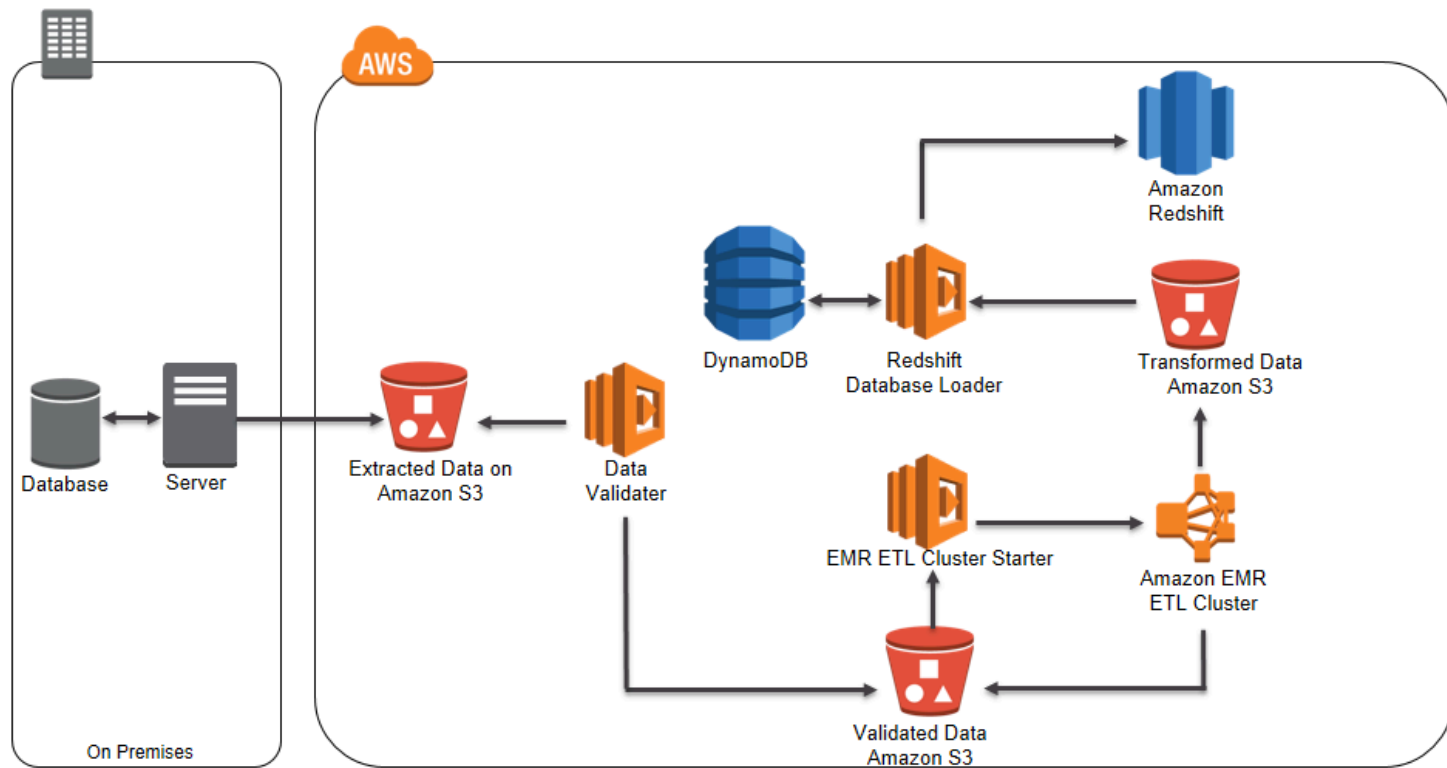
데이터 프로세싱 자동화란 필요한 데이터를 **추출, 수집, 정제**하는 프로세싱을 최소한의 사람 인풋으로 머신이 운영하는 것을 의미

예) Spotify 데이터를 하루에 한번 API를 통해서 클라우드 데이터베이스로 가져온다고 했을 때 매번 사람이 데이터 파이프라인을 작동하는 것이 아니라 크론탭 등 머신 스케줄링을 통해 자동화

I 자동화를 위해 고려할 사항

- 데이터 프로세싱 스텝들
- 에러 핸들링 및 모니터링
- 트리거 / 스케줄링

I 데이터 프로세싱 스텝들



AWS Data Pipeline, AWS Step Functions 등

I 에러 핸들링 및 모니터링

Python Logging Package

```
import logging
logging.basicConfig(filename='example.log',level=logging.DEBUG)
logging.debug('This message should go to the log file')
logging.info('So should this')
logging.warning('And this, too')
```

Cloud Logging Systems

- AWS Cloudwatch
- AWS Data Pipeline Errors

I 트리거 & 스케줄링

