안녕하세요 저희는 이장연 이진우 전수영으로 구성된 팀입니다.

저희가 이번 프로젝트로 선정한 것은 유라시아 지역의 코로나 현황을 csv파일로 다운받아 전처리 과정을 통해 시각화하는 작업을 하기로 했습니다.

저희 조의 환경 소개

1. 엘라스틱 서치(Elasticsearch)

2. 키바나(Kibana)

3. 로그스태시(Logstash)

4. 파일비트(Filebeat)

* Filebeat는 csv 파일을 Logstash에 메시지 형식으로 보내는 역할을 해줍니다.
* Logstash는 데이터를 받아 Elasticsearch에 인덱싱하기 적절하게 필터링을 해줍니다.
* Elasticsearch는 데이터를 저장하고 인덱싱 하는 역할을 해줍니다.
* Kibana는 Elasticsearch의 데이터를 가져와 시각화를 해줍니다.

장점 소개

1. Elasticsearch, logstash 및 Beats, Kibana는 각각 데이터의 쿼리(검색), 수집, 시각화를 담당합니다. 용도별로 분리하여 발전하는 솔루션이기에 구조적 안정성은 물론 다른 시스템과도 유연한 호환성을 가집니다.
2. 처음부터 확장을 고려하여 만들어졌기 때문에, 여러 대의 서버를 엮어서 성능 향상을 기대할 수 있는 클러스터 방식을 구성할 때 관계형 데이터베이스보다 상대적으로 관련 정보에 쉽게 접근이 가능합니다.
3. 데이터 처리 절차를 프로그래밍 언어를 이용한 코딩으로 명시한다면, 향후 유지보수가 용이하지 않고 불필요한 종속성이 높아지는 문제가 있지만 Logstash를 포함하는 ELK 스택 구성 시 유연성 확보가 가능합니다.

코로나 실태를 한눈에 알아볼 수 있게 아키텍처를 활용했기 때문에 자기가 원하는 정보들을 선택해 원하는 형식으로 정리한 정보들을 받아들일 수 있습니다.

저희가 프로젝트를 진행하면서 만들거나 사용한 파일들은 주석처리를 통해 설명이 나와있으니 참조하시면 좋을 것 같습니다.

이상입니다. 감사합니다.