**4차 멘토링 피드백**

2024/06/01 2팀 김건호

1. **공통**

**백업**

자동 백업이 아닌 수동 백업으로 주기적으로 할 것인지 이야기를 해줘야 함

자동 백업의 경우 시스템적으로 정확히 언제 하는지 할 수 없음.

즉, 정확하게 원하는 타이밍에 맞춰 돌릴 수 없다는 것임

백업은 사후 대처 측면에서 하는 것이기 때문에 이 부분에 대한 고민도 들어가야 함

→ 최대한 빨리 복구하고, 최대한 최근의 데이터를 유지할 수 있도록 하는 관점에서 접근해야 함

그래서 시나리오 만들기 나름인 것임

백업에서 RTO, RPO와 같은 사후대처 관련을 언급하면 좋음

오로라 같은 경우는 시간이 명확하게 정해져 있음. 자동 복구가 됨

이러한 고민이 없으니까 중간 평가 때 백업에 대한 피드백이 있었던 것임

아키텍처를 선정했을 때 고객은 잘 모르니까 이 부분 이유를 잘 언급해야 함

ex) RDS를 굳이 안쓰고 EC2에 설치하면 안되나? 이런 질문이 나올 수 있음

DB의 경우 암호화를 고려함. RDS는 커스텀식임. 설치형 인스턴스라 보면 됨. 서버 설치를 하는 것처럼 보임

패치, 백업, 업데이트와 관련한 관리 부분을 AWS에 위임함.

즉, 운영에 신경 쓰는 것이 아닌 비즈니스에 집중할 수 있게 함.

→ 아키텍처를 제안할 때 명확한 이유 하나 확실하게 찾아올 것

1. **개인**

* parquet에 대해 정리할 것
* OLAP성 데이터 -> 단일 레코드가 아닌 모든 데이터를 컬럼 단위로 저장
* columner 방식 중복된 데이터를 압축하는 효과. 하지만 실제 압축은 아님

crawler를 실행하면 partition이 생성되는 것임.

s3에 구분돼 있는 것은 파티셔닝 됐다 표현하는 것이 아님.

그것은 키 값을 기준으로 디렉토리로 만들어진 것임

**kinesis 데이터 재처리**

data streams – firehose의 경우 이 구간에서 장애가 발생한다는 것은 AWS 자체가 장애가 난다는 것임.

그래서 이 부분에 대한 장애는 고민을 할 필요가 없음

**발표**

심사위원보다 장표를 많이 봄. 발표시에는 가능하다면 몸 방향을 심사위원 쪽으로 고정해야 함

심사위원 한 명씩 눈을 마주보며 진행해야 함. 발표를 듣는 청중의 눈을 보면서 잘 이해했는지 확인하기

발표 연습할때는 스크립트를 짜놓고 연습하는 것이 좋음.

짜놓고 발표를 진행할 때는 그 스크립트를 외우는 것이 아니라 그것을 기반으로 말하면 장표를 봤을 때 잘 나올 것임

**멘토님 피드백**

로그 파이프라인, SQL, crawler 활용 부분에 힘을 많이 줄 것.

이 부분에서 한 것이 많으니 알기 쉽게 잘 설명하면 완성도 높아 보일 것임

1. **소감**

데이터 분석과 인프라가 결합 되어서 효과적인 아키텍처를 만들 수 있다는 것이 새롭고 즐거웠다. 컨테이너와 CI/CD가 개발자 입장에서는 빠르고 효율적으로 서비스를 배포할 수 있어서 좋듯이, 데이터 분석가 입장에서는 AWS가 관리해주는 서비스를 이용해서 데이터 분석에 집중할 수 있다는 것이 좋을 것이다.

그러한 입장에서 생각해보면서 관점을 바꿔보니 즐거운 경험이었다. 데이터 분석 부분은 끝냈고 원래의 인프라 부분에 집중할 시기이지만, PPT 장표에 내가 한 것을 확실하고, 효과적으로 나타낼 수 있게 추상화하는 작업에 시간을 좀 써야 겠다는 생각이 든다.

멘토링 또한 도움이 많이 되었다. 이때까지 데이터 분석을 대체 왜 하는 건지, 어떻게 하는 건지도 잘 알지 못하였는데 직접 하면서 경험해보니 효과적인 인사이트를 도출해낼 수 있는 매력이 있는 분야인 것 같다. 많이 도와 주신 멘토님께 진심으로 감사드리고 도와주신 것을 바탕으로 도움이 되는 나의 경험으로 정리해야겠다.