**2차 멘토링 피드백**

2024/05/18 2팀 김건호

1. **방향성**

각자 공통 주제로 테스트했던 시나리오를 하나로 모아야 함

물론, 주제를 특정하기보다는 적절히 섞어서 진행하면 좋을 것 같음

S3와 같이 하나의 데이터 레이크로 모아서 데이터를 처리하는 작업을 세분화해야 함

중요한 부분은 S3로 수집한 데이터를 어떻게 전처리하고 어떻게 분석해 어떤 시각화를 할 것인지

어떤 결과를 시각화 할 것인지가 미리 정해져 있어야 함

WBS를 작성해서 현실적으로 할 수 있는 것 체크

R&R을 나눠서 진행. task로 나눌 것인지, 아니면 서비스로 나눌 것인지 생각

detailed task를 나누고 status, issue 사항도 체크할 것

**테스트 – 전처리 – 분석 – 시각화 까지는 필수**

**3개의 순서로** 가이드라인 제시 받음

1순위로 데이터 마트 구성, Visualization까지의 과정을 완성할 것

* S3 – glue crawler – Data Catalog - Quicksight까지의 과정
* 데이터 분석을 우선 완수할 것. 시각화는 무조건 필수임
* 앞에 과정이 완성됐다고 가정하고 분석부터 진행

2순위로 데이터 수집 방법을 구현할 것

* 기본 인프라에서 데이터를 어떻게 가져올것이냐의 문제를 생각

3순위로 데이터 workflow를 설정할 것

* 워크플로우가 돌아갈 스크립트가 없는데 만드는 것은 의미 없음
* 제일 후순위로 미룰 것

**Q. 데이터 분석에서 시나리오가 먼저? 구축이 먼저?**

* 큰 틀을 잡아놓고 분석 및 인사이트를 도출 후 시나리오를 구체화
  + 고객사가 다루어야 하기 때문에 고객사의 스킬셋과 같은 부가적인 요소들을 고려해야 함.
  + 데이터쪽은 시나리오에서는 상대적으로 자유로움
  + 그래서 프로젝트 진행 시 사용한 기술 및 언어나 결과에 따라 시나리오를 구체화 시킬 것

**Q. 시나리오에서 고객이 무엇을 중점적으로 보는지를 어떤 기준으로 설정해야 하나?**

* 데이터 쪽은 일반적인 운영 개념과는 다르게 생각해야 함
  + 고객들은 우선 성능은 당연한 관점으로 생각.

여기서 성능의 종류는 대시보드를 의미할 수도, 얼마나 데이터를 빨리 처리해 결과를 낼 수 있는지를 의미할 수도 있음

* + 부가적으로 비용을 줄일 수 있으면 좋다 라고 생각

이 모든 것을 다 고려해서 얼마나 합당한 솔루션을 제시했냐를 시나리오에 입히면 됨. 인사이트 결과를 내고 구체화 시키는 방향으로 진행

1. **공통 과제 – 데이터 분석**
2. **설문 조사**

많은 기업에서 구글 폼과 같은 설문조사 툴을 사용

굳이 설문조사를 위해 추가 인프라 필요 X

아키텍처를 기반으로 시나리오를 변경 혹은 수집 방법을 달리 해서 가져갈 필요 있음

1. **데이터베이스**

**Q. DMS를 사용하여 RDS 데이터를 S3로 옮기는 것이 맞는 것인가?**

-> 가능은 하지만, 외부에 데이터베이스가 있는 상황에서 DMS를 고려함. DB 인스턴스보다 사양이 높아야 할 수 있음. 그래서 비용적 문제가 발생

-> DMS 이외에 s3로 스트리밍 하는 방법을 추천

**Q. 크롤러를 돌려야 하는 이유?**

-> 데이터 카탈로그를 만들어야 하기 때문임.

Athena에서 효율적으로 스캔할 수 있도록 파티셔닝이 필요함

**Q. OLAP, OLTP를 고려했나?**

-> OLAP는 분석을 위한 데이터, OLTP는 트랜잭션을 위한 데이터

각 데이터의 특성에 맞춘 관리가 필요

민경, 혜성은 시나리오를 적절히 섞으면 좋을 듯함

DB에 있는 데이터를 가지고 와야 할 경우가 있기 때문에 활용하면 좋을 것 같음

1. **클릭 스트림**

**Q. 액세스 로그가 의미 있는 데이터인가?**

-> 크게 의미 없는 데이터임

액세스 로그 자체로는 볼 수 있는 것이 없고, 크게 할 수 있는 것도 없음

그래서 차라리 수집 서버를 만들고 로그를 쌓도록 스크립트를 구성할 것

더 쉽게 하려면 자바스크립트로 만들어서 현재 세션 id 정보 등을 넘겨 서버에서 받아와 저장하게끔 구성