

Data engineering challenges

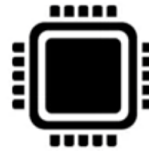
Common challenges encountered by data engineers



Access to data



Data accuracy and quality



Availability of computational resources



Query performance

1. Access to data

: 서로 다른 데이터세트를 통합하기 위해 다양한 데이터 형식 및 대규모 데이터에 대한 액세스를 관리합니다. 모든 데이터가 다른 조직에서 왔으며, 그것은 다른 도구에서 온다, 그것은 다른 스키마를 가지고, 그리고 아마도 그 데이터 중 일부는 구조화되지 않았을 수도 있습니다.

2. Data Accuracy and Quality

: 데이터는 정리해야 유용하기 때문에 수집한 모든 원시 데이터는 소스 시스템은 청소하고 변형해야 합니다. 쿼리를 효율적으로 만드는 형식입니다. 즉, ETL 데이터를 저장하고 데이터 웨어하우스에 저장합니다.

3. Availability of Computational Resources

: 너무 많은 통합 및 정리 작업을 수행해야 하는 경우 발생하는 일반적인 문제는 계산 자원의 가용성 그 자체가 어려워 될 수 있습니다. 온프레미스 시스템에 있는 경우 데이터 엔지니어가 관리해야 할 서버 및 클러스터 용량을 확인하고 수행하기에 충분한 능력이 존재해야 하고, 그 사람이 ETL 작업을 완료해야 합니다. 문제는 필요한 계산이 일정하지 않습니다. 즉, 트래픽이 적을 때 당신은 당신이 있기 때문에 돈을 낭비하게 됩니다.

4. Query performance

: 온프레미스 데이터 분석 클러스터를 관리하는 경우 모든 사람이 선택에 대한 책임이 있기

때문에 쿼리 엔진 및 설치 쿼리 엔진 소프트웨어 및 유지 뿐만 아니라 최신 추가 용량을 위해 더 많은 서버를 프로비저닝합니다.