JSON 기반 OLAP DB 비교

작성일: 2024-08-9
작성자: 이서연

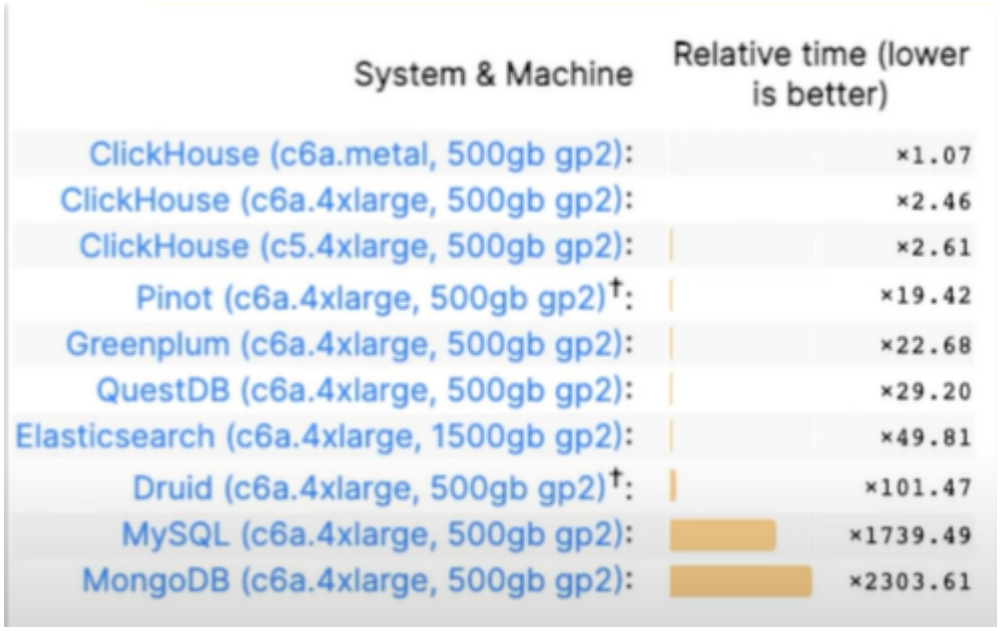
OLAP 개념 설명

온라인 분석 처리.
데이터 웨어하우스, 데이터 레이크 또는 기타 데이터 저장소에 있는 대량의 데이터의 복잡한 쿼리 또는 다차원 분석을 고속으로 수행하기 위한 기술입니다. OLAP는 비즈니스 인텔리전스(BI), 의사 결정 지원, 다양한 비즈니스 예측 및 보고 응용 프로그램 등에 이용됩니다.

1. ClickHouse

(참조 URL : What Is ClickHouse?)

- 특징
  : 대규모 데이터 분석에 최적화. 고성능 컬럼 기반 데이터베이스. 실시간 처리에 강점(천 만, 억 단위의 행을 처리 가능) 다양한 내장 함수와 기능을 통해 복잡한 쿼리를 실행할 수 있음. OLAP 워크로드에 적합
  - 속도 : 굉장히 빠름. 억 단위 행을 처리 가능. 집계 쿼리를 실행했을 시 클릭하우스의 빠르기는 하단과 같음.



  - **HA** : 애플리케이션 비동기형. 멀티 마스터 지원. zookeeper와 호환. 메타 데이터 저장용 DB가 필요하지 않음
  - **SQL** 문법 사용

    - 기타 :
        - Real-Time Analytics 대쉬보드 제공
        - SQL 문법 사용
        - 북미와 서유럽에서 많이 사용
        - 분산형 아키텍처, 확장성이 좋음

2. Apache Druid

(참조 영상 : Apache Druid in 5 Minutes)
: 실시간 데이터 분석과 대규모 데이터 처리를 위한 오픈소스 분산 데이터베이스

- 특징: 실시간 데이터 분석과 대규모 데이터 처리를 위한 오픈소스 분산 데이터베이스로, JSON과 같은 비정형 데이터의 저장과 처리가 가능합니다. OLAP 쿼리를 빠르게 처리할 수 있는 능력이 있으며, 실시간 분석에도 강점을 가지고 있습니다.
    - 실시간 시계열 데이터 저장 및 분석 : clickstream 데이터 수집을 위해 만들어진 DB. 실시간 데이터 수집 및 조회 성능은 우수하나 트랜잭션이 따로 존재하지 않음(데이터 삭제를 원하는 경우 저장된 세그먼트를 삭제)
    - **HA** : zookeeper를 통한 HA, 메타데이터 스토리지로 RDBMS 따로 필요
    - 실시간 인덱싱 지원
    - 고속 쿼리 성능
    - 분산형 아키텍처
    - 대규모 병렬처리 지원
    - 다차원 분석 지원
    -

3. AWS DynamoDB

: AWS에서 지원하는 NoSQL 기반 데이터베이스 서비스.

- 특징: AWS에서 지원하는 서비스인 만큼 S3와 연동이 쉬우며 key-value 값으로 스키마가 따로 정해져 있지 않음. 수평확장과 같은 프로비저닝이 쉬운 편. 질의 속도도 빠른 편이다.
    - 빠른 질의 속도
    - 자동 프로비저닝
    - 대량의 비정형 데이터 처리에 유리
    - 조인이 불가능. 연산이 다양하지 못 하다는 단점이 있음
    - AWS 대쉬보드 지원. SaaS.
    -