빅쿼리를 활용한 빅데이터 분석 (22)

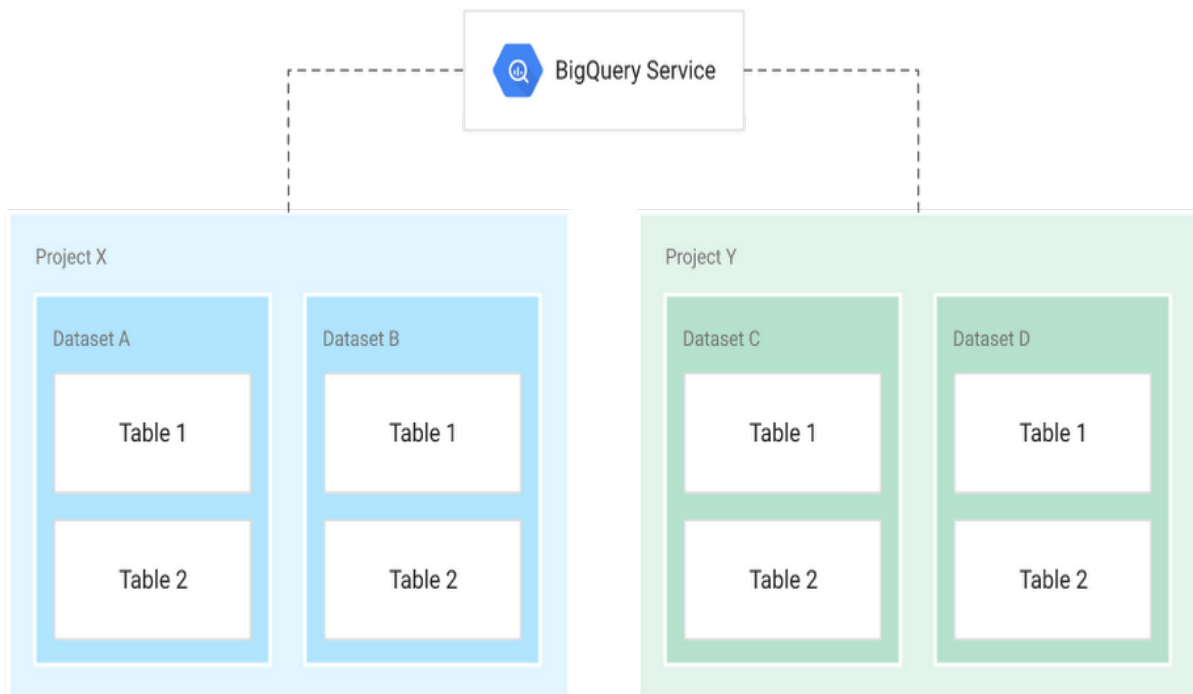서진호
자료 다운로드: https://www.github.com/synabreu/BigQuery

제 22 강 목표
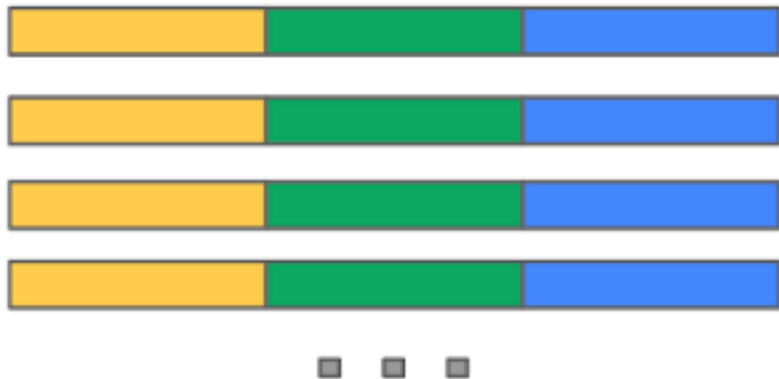
1. 빅쿼리 리소스 모델
2. 빅쿼리에서 스토리지 관리
3. 빅쿼리에서 테이블 파티셔닝
4. 빅쿼리에서 테이블 클러스터링

빅쿼리 리소스 모델

- 프로젝트(Project)
- 데이터셋(Dataset)
- 테이블(Table)
- 뷰(View)
- 작업(Job)
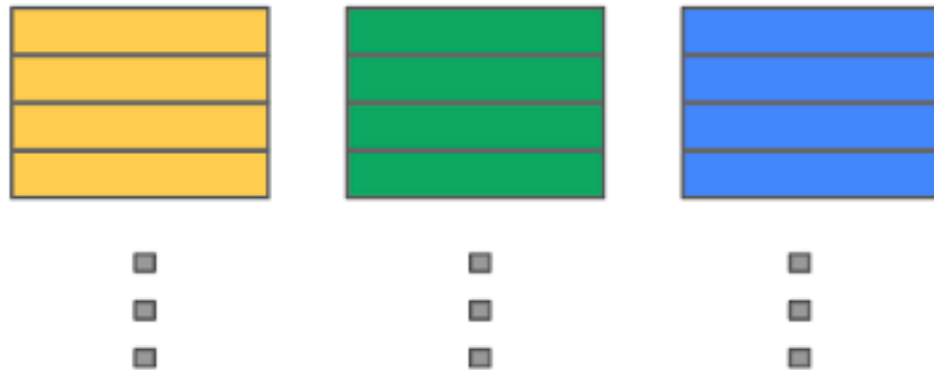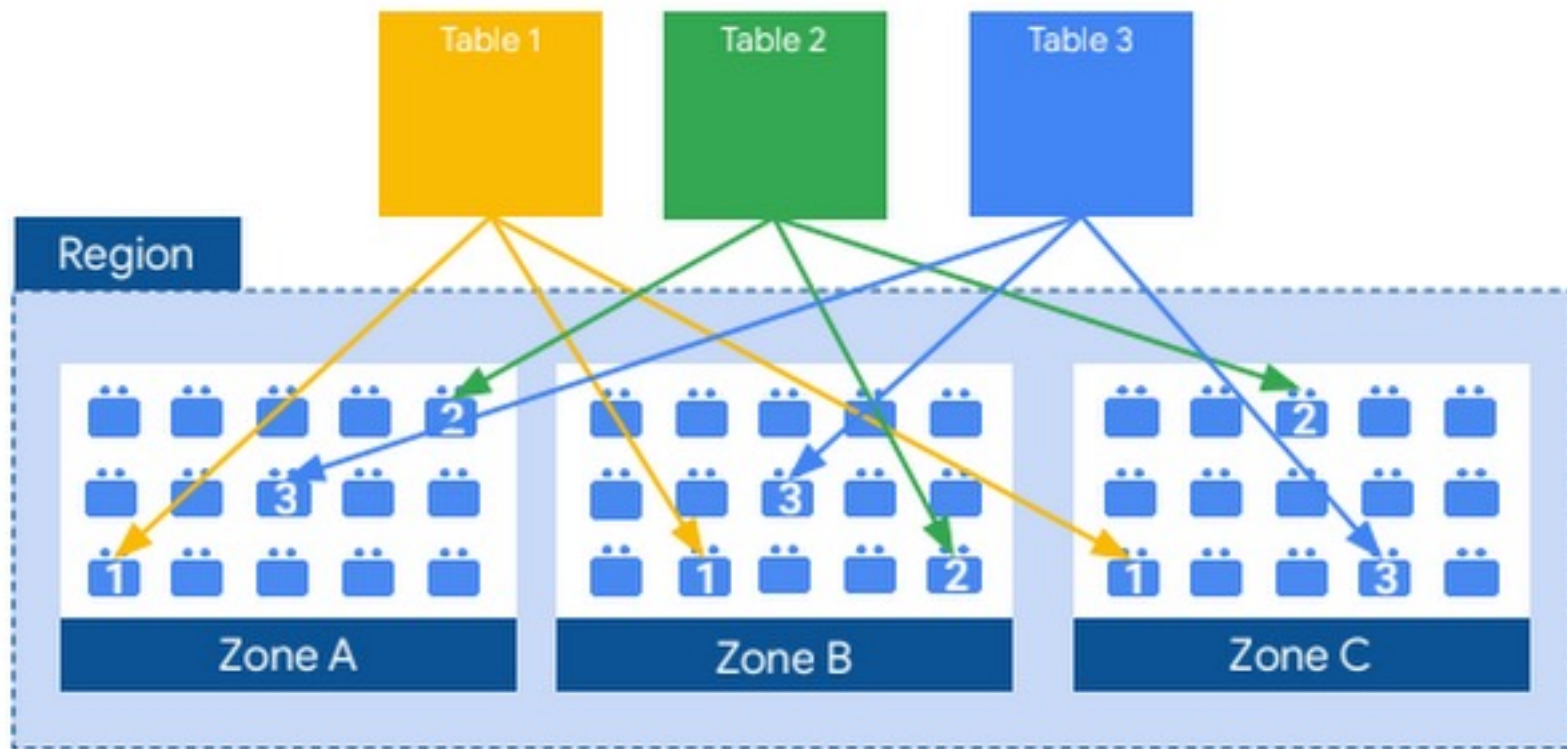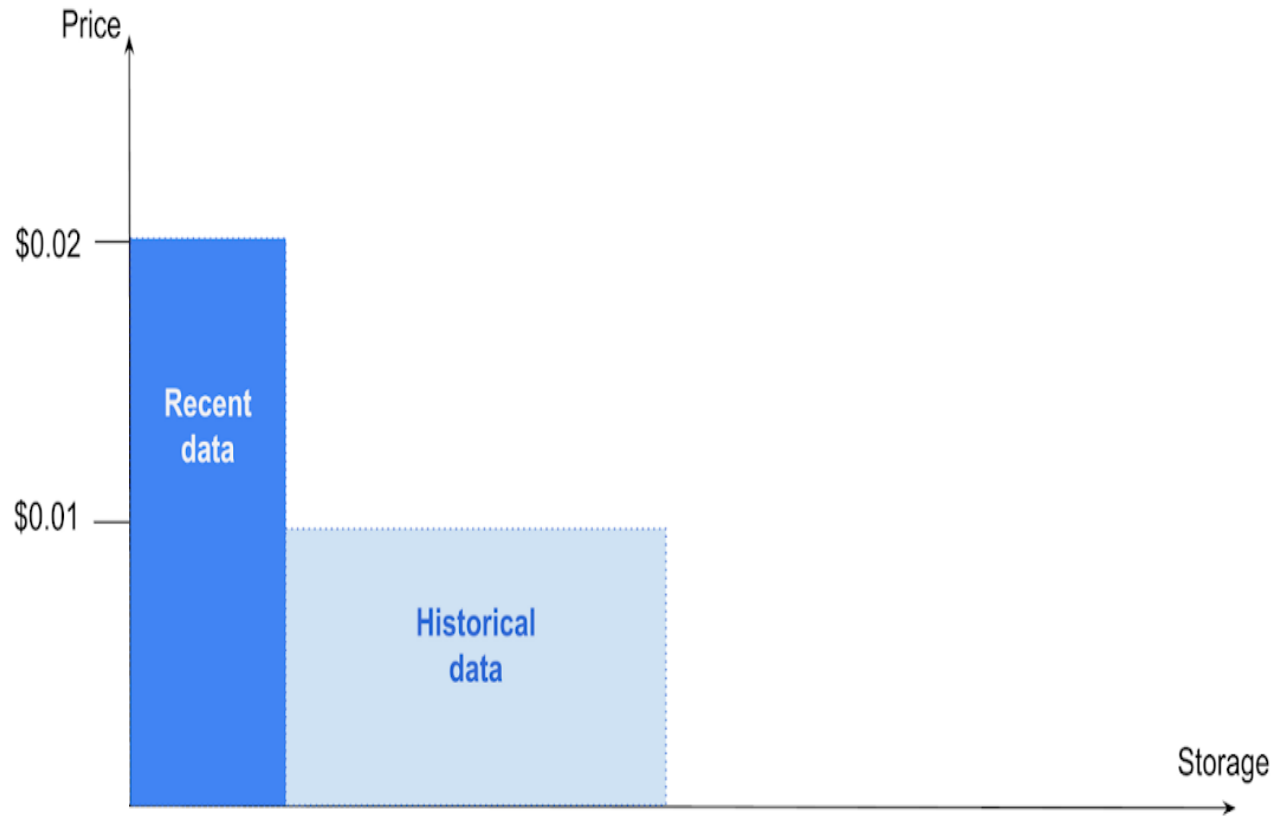
스토리지 관리 (1)

Traditional RDBMS storage

BigQuery storage

Record-oriented storage

Columnar storage

스토리지 관리 (2)

장기간 스토리지

파티셔닝

stackoverflow.questions_2018

Creation_date	Title	Tags
2018-03-01	How do I??	Android
2018-03-01	When Should?	Linux
2018-03-02	This is great!	Linux
2018-03-03	Can this?	C++
2018-03-02	Help!!	Android
2018-03-01	What does?	Android
2018-03-02	When does?	Android
2018-03-02	Can you help?	Linux
2018-03-02	What now?	Android
2018-03-03	Just learned!	SQL
2018-03-01	How does!	SQL

Partition

stackoverflow.questions_2018_partitioned

20180301

Creation_date	Title	Tags
2018-03-01	How do I??	Android
2018-03-01	When Should?	Linux
2018-03-01	What does?	Android
2018-03-01	How does!	SQL

20180302

Creation_date	Title	Tags
2018-03-02	This is great!	Linux
2018-03-02	Help!!	Android
2018-03-02	When does?	Android
2018-03-02	Can you help?	Linux
2018-03-02	What now?	Android

20180303

Creation_date	Title	Tags
2018-03-03	Can this?	C++
2018-03-03	Just learned!	SQL

Google

빅쿼리에서 파티셔닝하는 테이블

- 데이터 주입할 때 시간으로 파티션을 나눈 테이블: _PARTITIONTIME, _PARTITIONDATE

- DATE/TIMESTAMP 컬럼으로 파티션을 나눈 테이블: __NULL__ , __UNPARTITIONED__

- INTEGER 범위 파티셔닝 테이블

실습1 – 파티셔닝 사용 예제

빅쿼리에서 클러스팅

실습2 – 클러스터링 사용 예제