

Module 7

Ingesting New Datasets into Google BigQuery

In this module we will:

- Query from External Data Sources
- Avoid Data Ingesting Pitfalls
- Ingest New Data into Permanent Tables
- Discuss Streaming Inserts

Ingest data permanently into BigQuery from a variety of formats



Cloud
Storage



Google
Drive



Cloud
Dataflow



Cloud
Dataprep

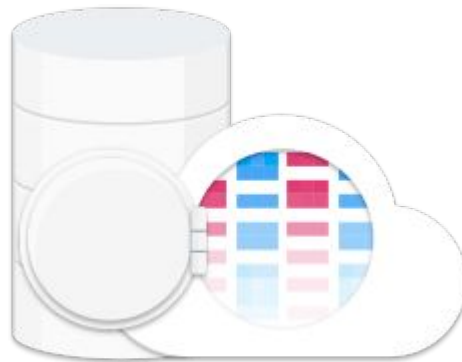
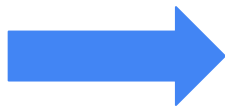


Cloud
Bigtable



CSV,
JSON,
Avro

... and more formats



BigQuery Managed Storage

Data ingested into BigQuery is stored in permanent tables and this storage is scalable and fully-managed



BigQuery
Query
Engine

BigQuery can query external data sources in GCS and Drive directly



Cloud
Storage



Google
Drive



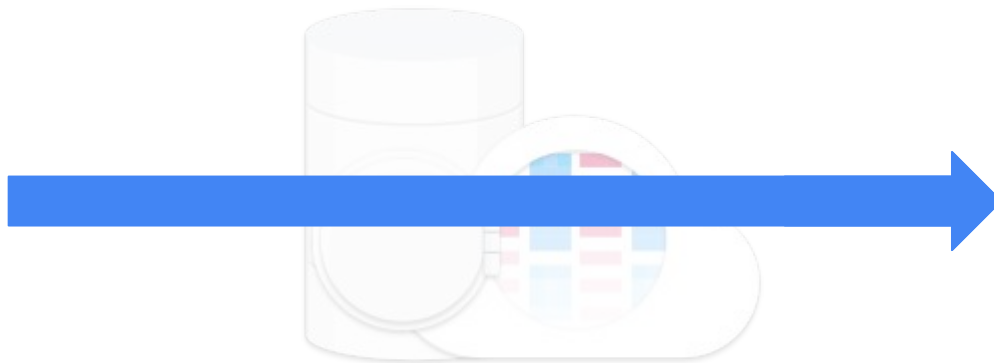
Cloud
Dataflow



Cloud
Dataprep



Cloud
Bigtable



BigQuery Managed Storage



BigQuery
Query
Engine

You can query external data sources directly from BigQuery which bypasses managed storage

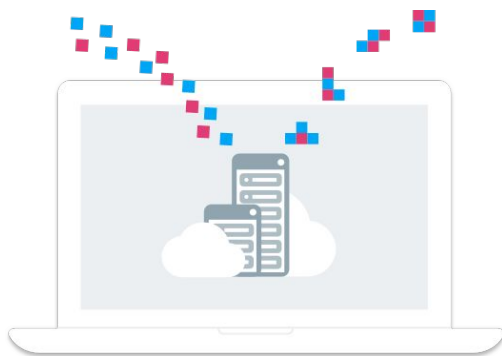
Pitfalls: Querying from External Data Sources Directly



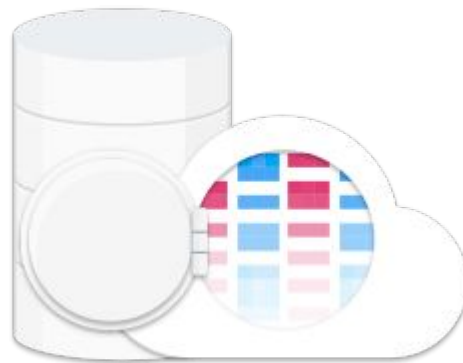
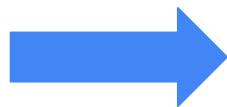
Limitations

- Strong Performance Disadvantages
- Data Consistency not Guaranteed
- Can't Use Table Wildcards (cool feature we will introduce shortly)

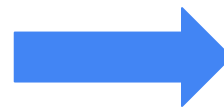
Streaming records into BigQuery through the API



Streaming Record Inserts



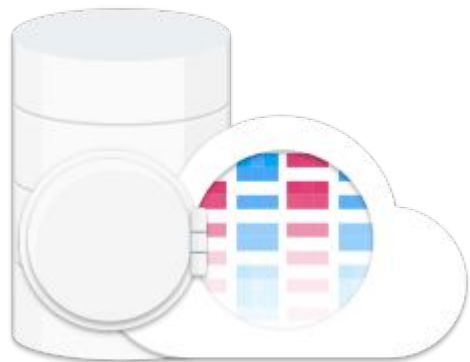
BigQuery Managed Storage



**BigQuery
Query
Engine**

Streaming data allows you to query data without waiting for a full batch load

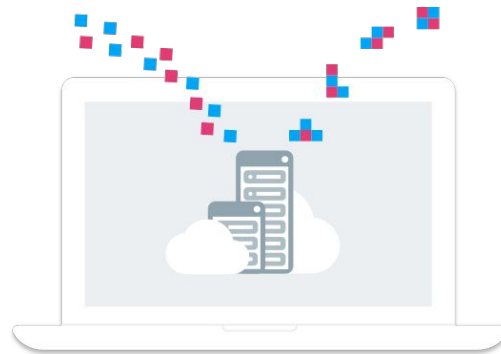
Summary: Ingest new datasets into BigQuery managed storage



Data stored permanently in BigQuery is fully-managed (performance, backups, redundancy).



Load new data from a variety of formats



Setup streaming ingestion into BigQuery through APIs

Lab 6

Ingesting and Querying New Datasets

Ingesting and Querying New Datasets

In this lab, you will ingest new data sources into Google BigQuery and learn how to query external data sources directly.

