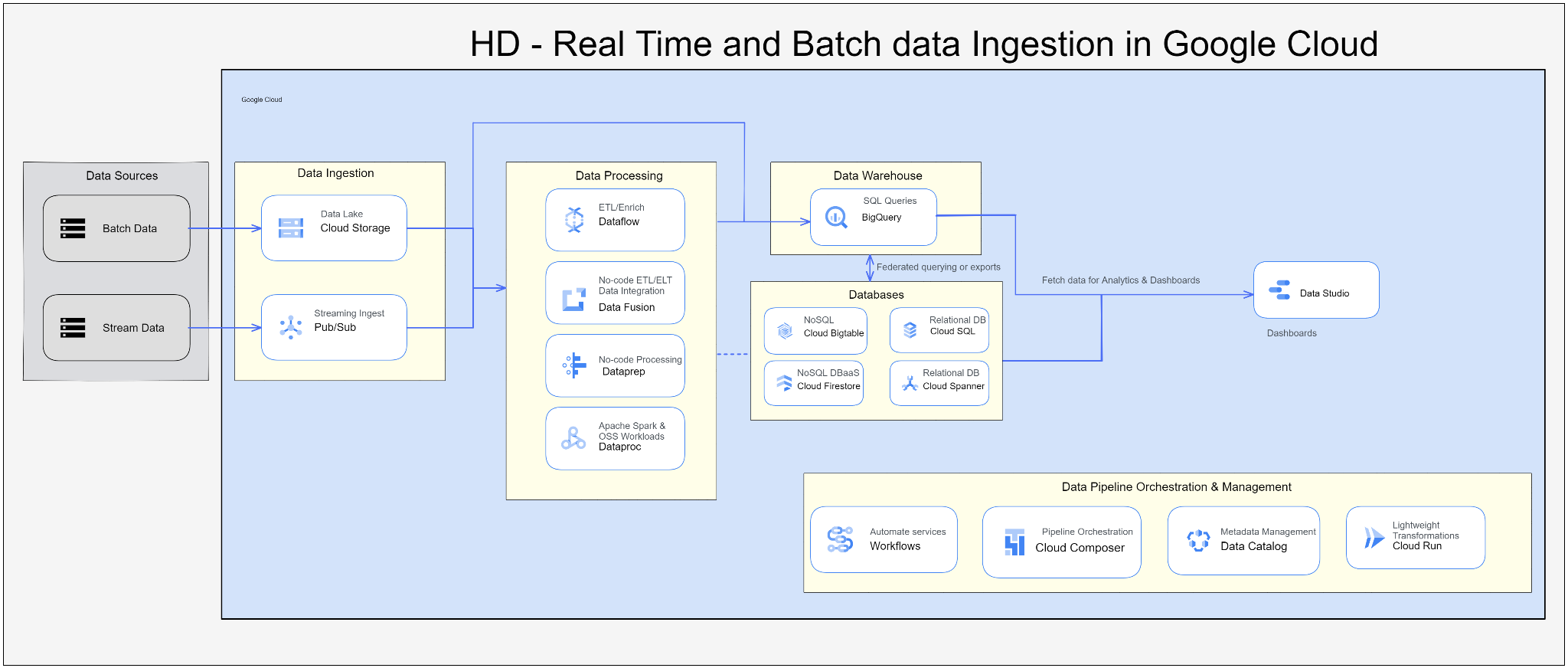
1. **Provide the architecture for an automated process to ingest and store the data in real-time and batch on Google Cloud.**

****

**Data Sources :-** Real Time and BatchData

**Ingestion :**

* **Batch data ingestion :** Batch ingestion can be done in many ways i.e. 3rd party ingestion tool, using APIs, and web UI.
* **Real-time data ingestion :** We should configure source in a such a way that it will send event message to pub/sub, and in pub/sub we should create Topic and subscription to read the event messages**.**

**ETL :**

* If we want to perform wide range of transformations in gcp across a large bounded dataset, then we have 2 options i.e. Dataflow or Spark via Dataproc.
* If we want to deduct anomales the dataprep is ideal for that to explore, clean and prepare the data for analysis.
* Data Fusion is not a replacement for Dataflow but rather a complementary. It enables Hybrid integration because it is based on an open-source alternative called CDAP. It also has additional metadata and lineage features that are not currently available in Dataflow.

**Data warehouse :**

* BigQuery is a datawarehouse solution in GCP, it supports structured and semi-structured data and it is a MPP so we will process the queries very quickly.

**SQL and NOSQL DB:** We have BigTable, it is a noSQL database and for relation data we have Cloud SQL. we can store the structured,Semi and Unstructured data in BigTable.

**Dashboard** : Datastudio is visualisation tool in GCP.

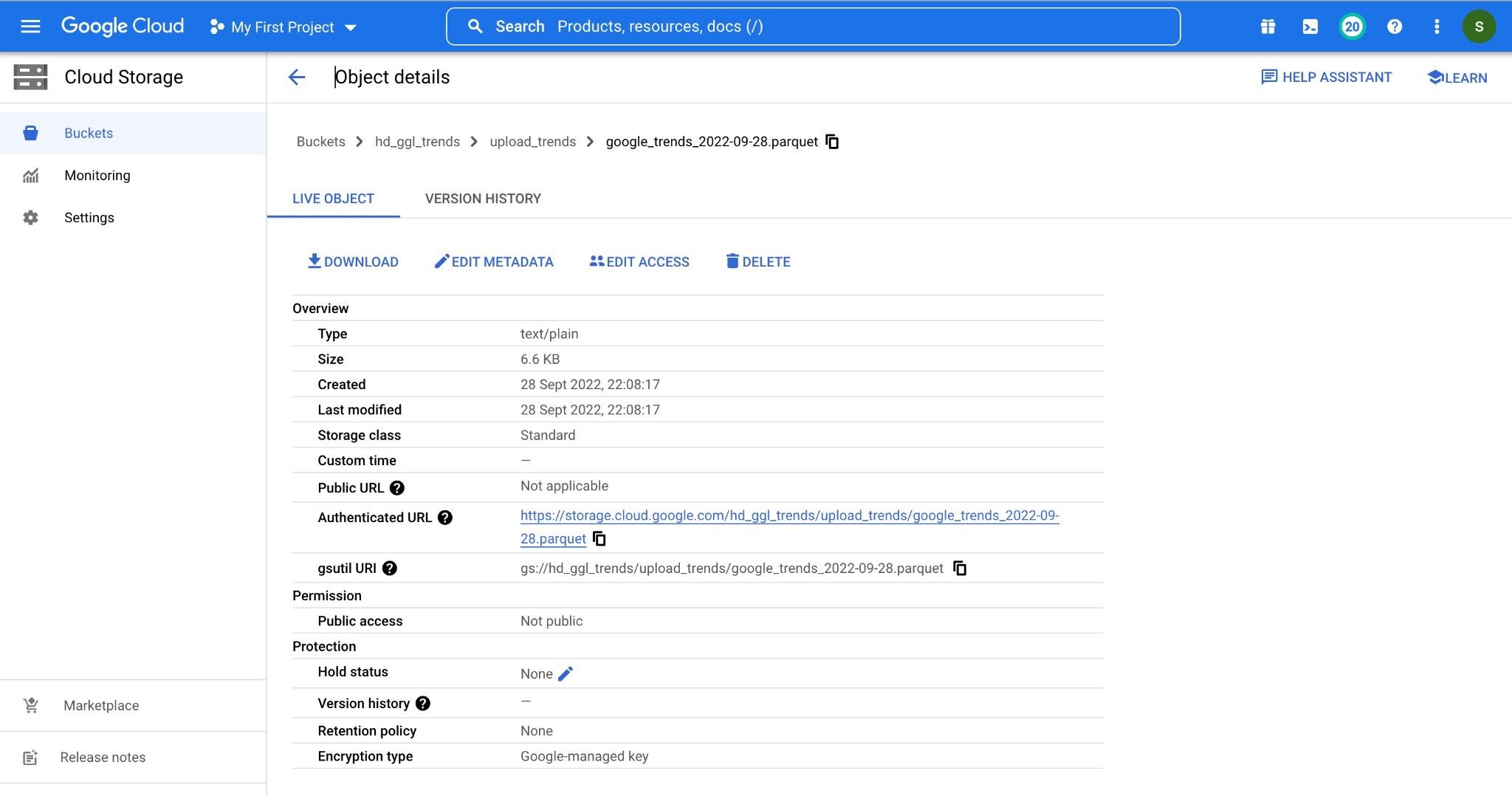
1. **Describe the process to enable BigQuery to efficiently query the Google Trends data from Cloud Storage ?**

Ans: We just need to create a external table in BigQuery either using webUI or DDL.

If we are using webUI then we must ensure that we change the default table type from “Native Table” to “External Table”. If we forget to do this will result in a static table that won’t load our files into the table.

Please refer below images: First image talks about googleTrendFile location in GCS, and In the second image, i have created an external table over gcs google trends file and queried it.

**Image - 1**



**Image - 2: To be continued.**

