**What Is Google BigQuery?**

Google BigQuery is a serverless, highly scalable data warehouse that comes with a built-in query engine. The query engine is capable of running SQL queries on terabytes of data in a matter of seconds, and petabytes fast.

**BigQuery: A Serverless, Distributed SQL Engine**

BigQuery is serverless, and you can run queries without the need to manage infrastructure.

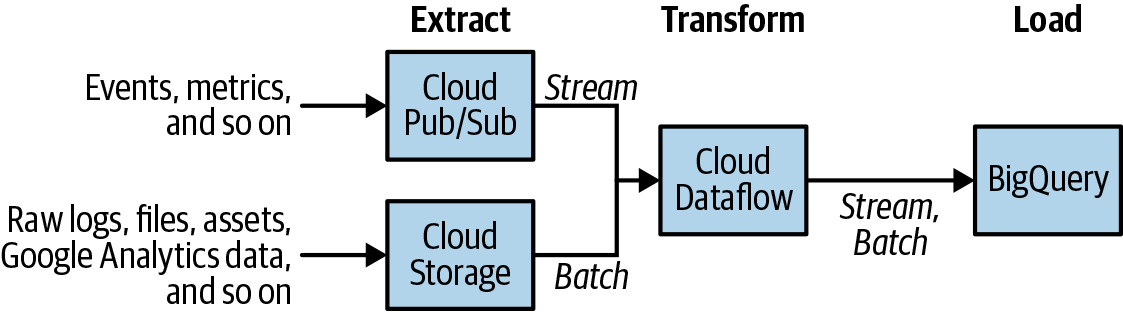
It enables you to carry out analyses that process aggregations over the entire dataset in seconds to minutes.

**BigQuery: Data Warehouse Access**

Note a few things about this. One is that you were able to run a query against a dataset that was already present in BigQuery. All that the owner of the project hosting the data had to do was to give you **“view”** access to this dataset. You didn’t need to start up a cluster or log in to one. Instead, you just submitted a query to the service and received your results. The query itself was written in SQL:2011, making the syntax familiar to data analysts everywhere. Although we demonstrated on gigabytes of data, the service scales well even when it does aggregations on terabytes to petabytes of data. This scalability is possible because the service distributes the query processing among thousands of workers almost instantaneously.

**Working with BigQuery**

BigQuery is a data warehouse, which means it has a high level of centralization and accessibility.



- BigQuery separates compute and storage, it is possible to run BigQuery SQL queries against CSV (or JSON or Avro) files that are stored as-is on Google Cloud Storage; this capability is called federated querying. You can take advantage of federated queries to extract the data using SQL queries against data stored in Google Cloud Storage, transform the data within those SQL queries, and then materialize the results into a BigQuery native table.

**Powerful Analytics**

The benefits of a warehouse derive from the kinds of analyses that you can do with the data held within it. The primary way you interact with BigQuery is via SQL, and because BigQuery is an SQL engine, you can use a wide variety of Business Intelligence (BI) tools such as Tableau, Looker, and Google Data Studio to create impactful analyses, visualizations, and reports on data held in BigQuery.

**Stream Data**

BigQuery supports the ingest both of batch data and of streaming data. You can stream data directly into BigQuery via a REST API. Often, users who want to transform the data—for example, by adding time-windowed computations—use Apache Beam pipelines executed by the Cloud Dataflow service. Even as the data is streaming into BigQuery, you can query it. Having common querying infrastructure for both historical (batch) data and current (streaming) data is extremely powerful and simplifies many workflows.

**BigQuery**

* Focus on building a compute engine and storage system in the cloud.
* SQL is “relationally complete,” meaning that any computation over the data can be done using SQL.
* Google had number of internal SQL engines that could operate over data, including some that were very popular. The most advanced engine was called Dremel; it was used heavily at Google and could process terabytes’ worth of logs in seconds.
* Dremel was able to process data so fast was that its query engine used distributed computing.
* The team renamed itself “BigQuery,” following the naming convention for “Bigtable,” Google’s NoSQL database.

**Summary**

BigQuery is a highly scalable data warehouse that provides fast SQL analytics over large datasets in a serverless way. Although users appreciate the scale and speed of BigQuery, company executives often appreciate the transformational benefits that come from being able to do ad hoc querying in a serverless way, opening up data-driven decision making to all parts of the company.

To ingest data into BigQuery, you can use an EL pipeline (commonly used for periodic loads of log files), an ETL pipeline (commonly used when data needs to be enriched or quality controlled), or an ELT pipeline (commonly used for exploratory work).

BigQuery is designed for data analytics (OLAP) workloads and provides full-featured support for SQL:2011. BigQuery can achieve its scale and speed because it is built on innovative engineering ideas such as the use of **columnar storage,** support for nested and repeated fields, and **separation of compute and storage,** about which Google went on to publish papers. BigQuery is part of the GCP ecosystem of big data analytics tools and integrates tightly with both the infrastructure pieces (such as security, monitoring, and logging) and the data processing and machine learning pieces (such as streaming, Cloud DLP, and AutoML) of the platform.

Eray-Erdoğan

**REFERENCES**

[1] Lakshmanan and Tigani ,Google BigQuery: The Definitive Guide (2022)