

# BigQuery

Atin Pothiraj

January 2021

## 1 Introduction

BigQuery encompasses the majority of the functions for which Cloud Computing was designed for and deals especially with large amounts of data in multiple server-less warehouses. BigQuery is produced solely by Google and available on the Google Cloud, but Amazon has a similar product called Redshift. For the purpose of simplicity and accessibility, this lecture will primarily focus on BigQuery, but feel free to explore Redshift in your own time. This lecture will be the building block for more complex projects due to the ML-capabilities directly engineered into BigQuery.

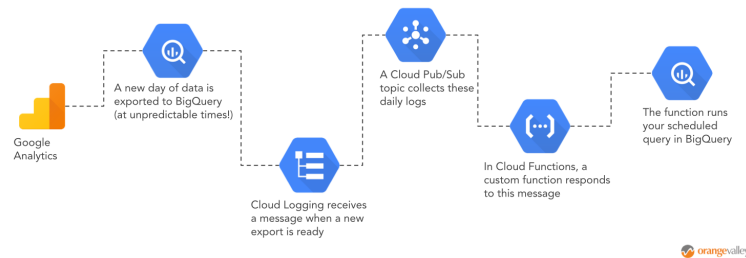


Figure 1: Graphic of BigQuery Migration

## 2 Benefits

- No Servers: Google deals with all the resourcing behind the scenes, meaning the data provided is secure and can be accessed efficiently.
- Limitless: BigQuery has multiple locations and can be accessed relatively anywhere because of the duplicated storage it has.
- Follows Standard SQL: BigQuery uses standard SQL dialect, meaning no extra code for applications that can only use SQL.

- Cost: Flexible pricing depending on the amount of storage needed. "Pay for what you use" pricing models also available.
- Auto-backups: BigQuery keeps track of previous data for up to seven days - allows the user to restore previous backups if needed

### 3 Notable Features

- Natural Language Processing: Google allows the user to analyze petabytes of data with embedded chatbots and UIs.
- Data Science Workloads: BigQuery allows open source data science workloads to be run - like Tensorflow, Dataflow, and Pandas - directly in its storage API.
- Geospatial Data Capabilities: BigQuery can optimize geospatial data in order to be more efficient with location intelligent algorithms.
- AI Capabilities: Users can integrate ML and train models on data with SQL inside of BigQuery.
- BI Capabilities: BigQuery allows companies to use business intelligence and run various statistical models on their data.
- Programming Capabilities: Its possible to integrate various applications or libraries that use Java, Python, Node.js, C, and Ruby to work with the data efficiently.