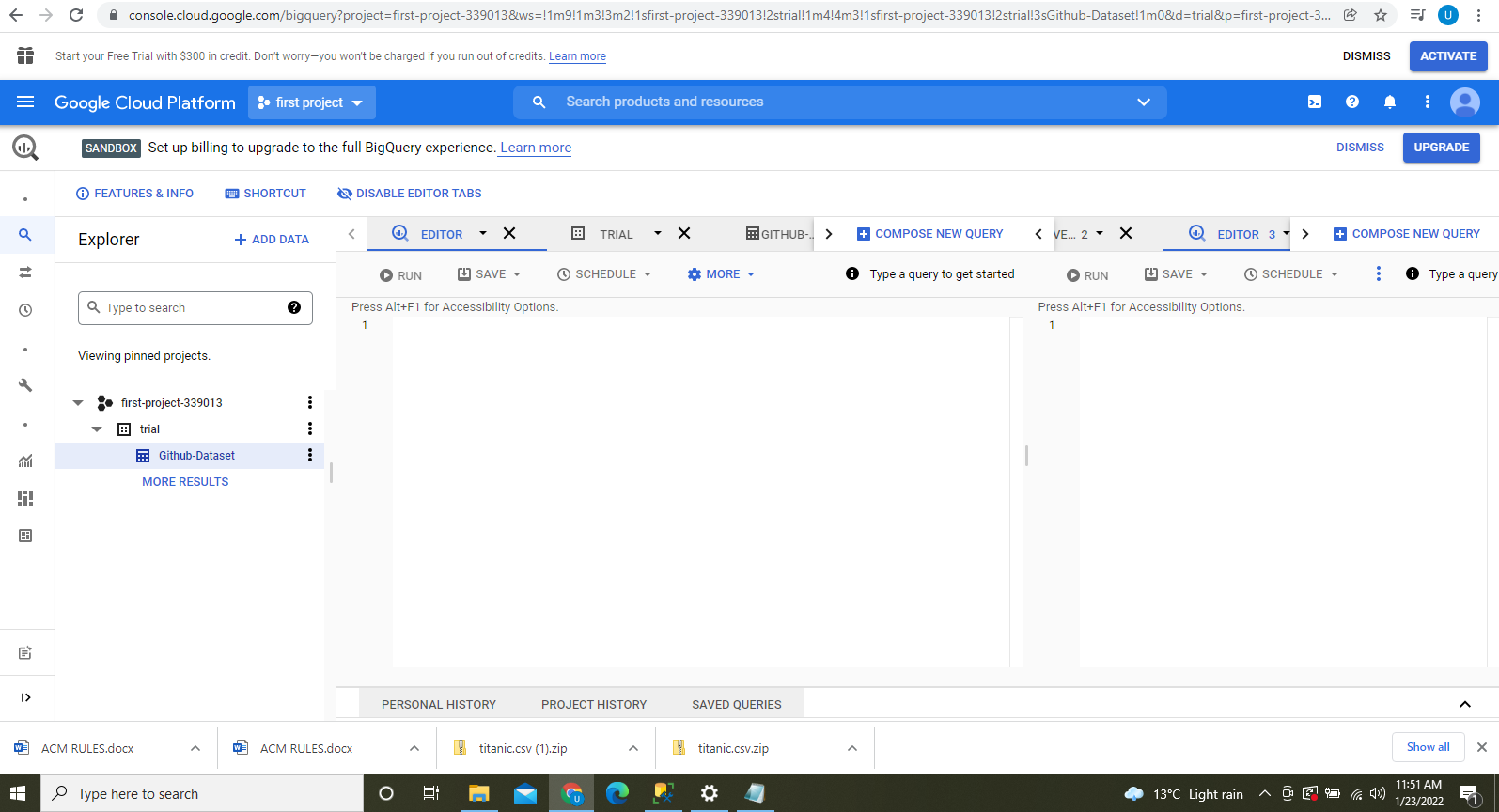
**What is BigQuery?**

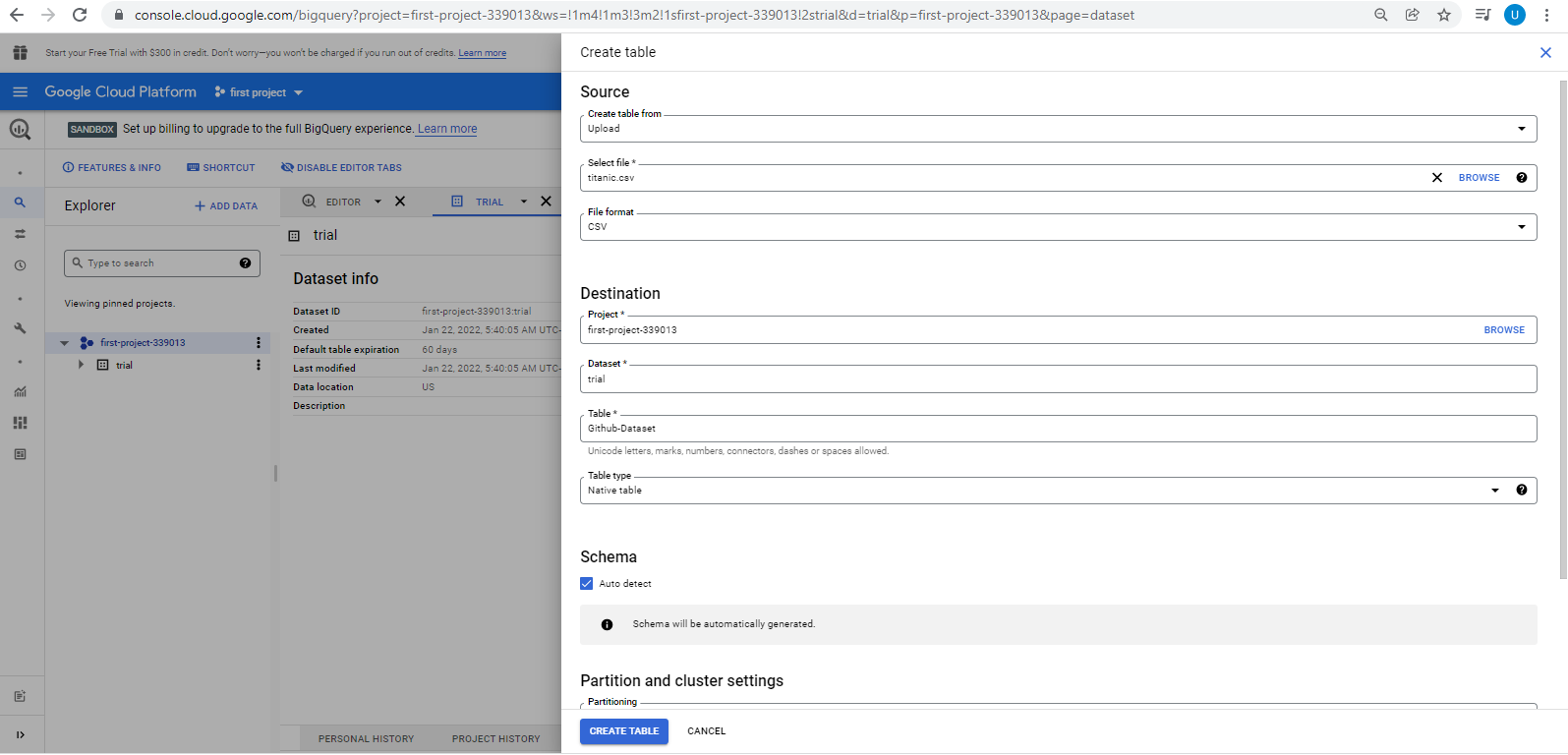
BigQuery is a Google-managed data warehouse platform that is extremely scalable, quick, and optimised for data analytics. It also includes rudimentary machine learning capabilities. It's also one of Google's several serverless offerings. This means you aren't in charge of the infrastructure assets or the overhead obligations and expenditures. It is only utilised to address the business use case, and it does it in a very efficient manner.

**Public Datasets**

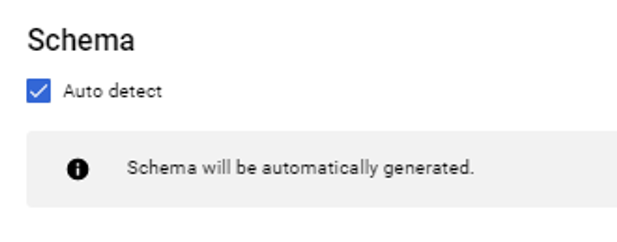
BigQuery isn't just a tool; it's a tool that's only as good as the data it's powered by. The good news is that if one doesn't have their own dataset to play with, BigQuery offers a variety of public datasets to choose from, including sample queries. There are many public datasets available online, one of them being Github Datasets. Below is the BigQuery Web UI. On the left hand side we see our project, if you're already logged in there. Here’s another link for public datasets we can use and explore from [https://cloud.google.com/bigquery/public-data/.](https://cloud.google.com/bigquery/public-data/)



**Figure 1: Bigquery Web UI**

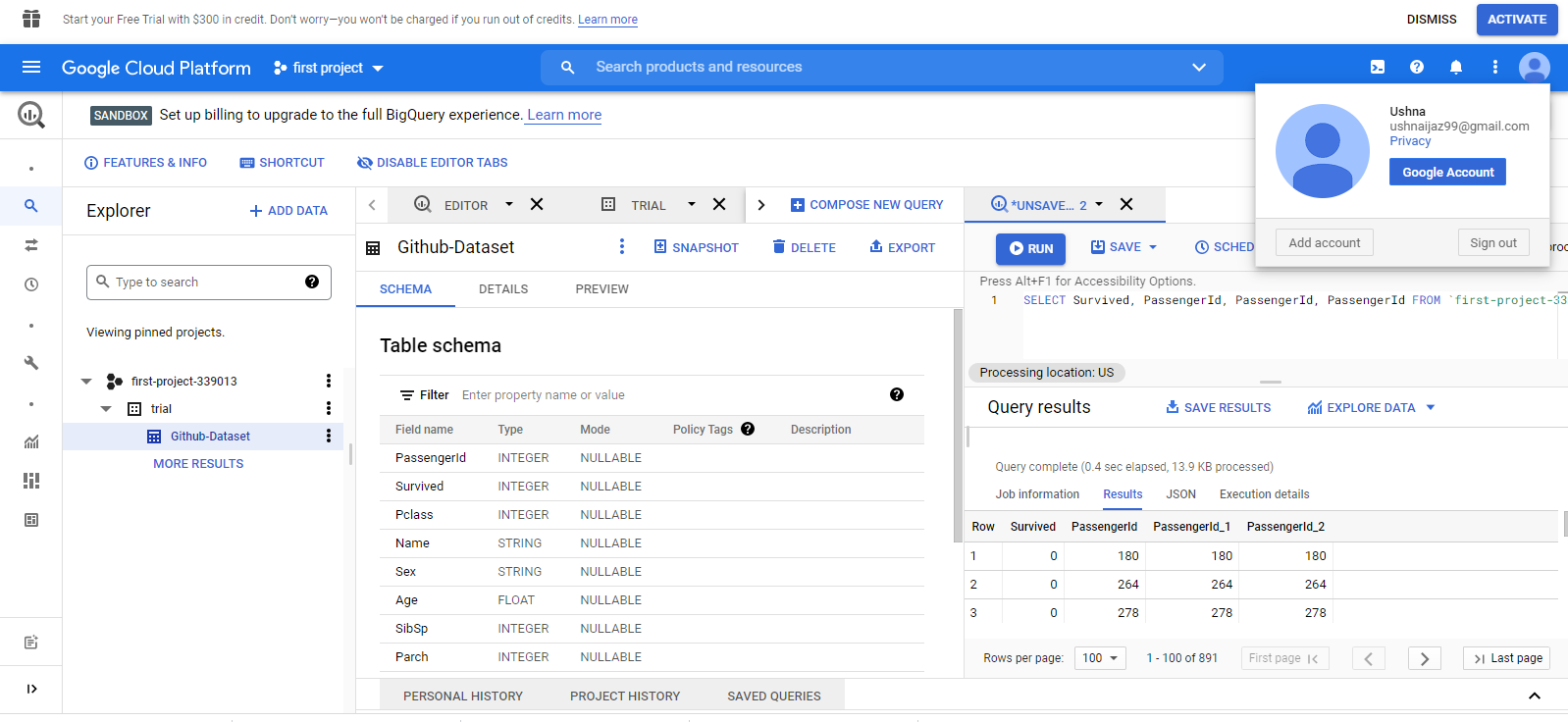
The first step is to establish a database, which you can then connect to. You can construct datasets in the cloud by starting a BigQuery session, going to a data directory, and creating a new dataset. When you're in the cloud, you may connect to the freshly formed dataset and wait for the BigQuery server to start. That means your information is saved locally on your computer. The dataset I used for this trial project was: <https://github.com/awesomedata/awesome-public-datasets>

**Figure 2: Creating a table**



Schema is very important because we need to model after what we have in the columns and what data type it actually is.

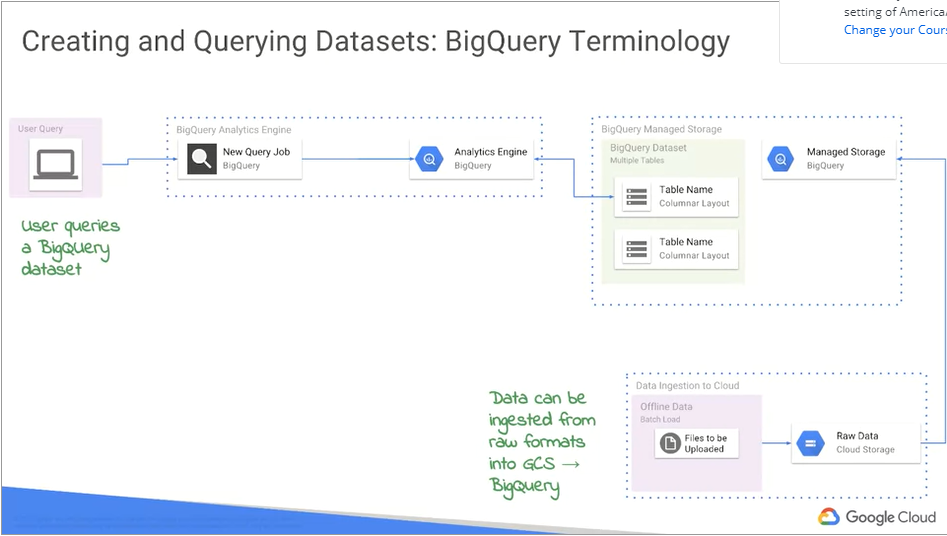
You can write SQL queries to retrieve data from ISB-CGC BigQuery tables directly in the Google BigQuery console.

**Figure 3: Using queries** 

We can use the “Preview” feature in the BigQuery web UI.

The query I used was:

SELECT Survived, PassengerId, PassengerId, PassengerId FROM `first-project-339013.trial.Github-Dataset` LIMIT 1000

**Bigquery terminology** 

Starting with the left. Usual query BigQuery. A unit of work in BigQuery itself is called a job. Jobs are conducted on a lightning-fast analytics engine developed by Google and made available as a service through BigQuery. The underlying data, which is completely maintained behind the scenes in those tables, is then mapped to those query jobs. And then, walking back the other way, all the way at the bottom there, you can ingest data into something like Google Cloud storage if you wanted to. Or directly into BigQuery if you wanted to, and then have that be available for analysis. So, Google BigQuery is the scalable managed storage component, and it's the same technology that Google uses to store a lot of their product data, right? Consider advertisements with Google's e-mail service, Gmail. But it's also that blazingly fast analytics engine, SQL engine, and it's built on Google's huge technological growth over time. The relentless march, if you will, to keep performing better and better. because Google is naturally incentivized because of the massive amounts of data that it has.

**Conclusion**

BigQuery is a fantastic tool for quickly going from zero to hero in terms of data exploration and analysis. Tools like BigQuery help extract value from data in a world where data is increasing at an unbelievable rate. BigQuery isn't a silver bullet, despite its unique advantages and tremendous features. It's best not to use it for data that changes frequently, and it's best not to use it as a primary data storage because of its storage location tied to Google's own services and processing constraints.