Querying the Ethereum Blockchain

with Google BigQuery

Wow...this is a game changer!

- Google have recently released a public dataset of the Ethereum blockchain.
- It is currently updated daily and provides access to easily digestible tables containing data that can be queried using standard SQL.

Why is this exciting?

Previously...

- Open source tools
- Analytics dashboards
- Ethereum explorers - i.e. etherscan
- Data Analytics Companies
- ETL code available for retrieving full blockchain dataset

Challenges

- ➔ Storage, cost and maintenance of full blockchain
- ➔ Complexity of raw data
- ➔ Difficult for non-dev users to get traction
- ➔ APIs limited use cases
- ➔ Difficult to obtain and classify required data for analysis - web scraping, reliance on third-party sites, reliability and quality, etc.

Now...

- **Aggregated data** - Token transfers, Blocks, Contracts, Transactions, Logs, Tokens
- **Query using SQL**
- **Data stored, maintained & publicly available by Google**
- **Daily update**
- **Partitioned data** by datetime (block timestamp)
- **Fast & free** (the first 1 TB per month is free)
- **Kaggle** also has public dataset available for querying via Kernels (a browser coding environment for Python)
- **It’s all at your fingertips!**

What is Google BigQuery?

- Data Warehouse & big data querying tool offering within the Google Cloud Platform (GCP).
- GCP analogous to other Cloud solutions, eg. AWS solution
  - cloud computing
  - data storage
  - ETL, querying,
  - identity & access management (IAM)
  - Analytics
  - machine learning

How to access Google BigQuery

There are 4 ways to access and interact with data in BigQuery:

1. Web UI (Beta)
2. Classic Web UI
3. Command-Line tools
4. Client libraries

Demos

1. Demo 1 - Web UI (Beta)
2. Demo 2 - BigQuery API with Client Libraries