A.1 Basic Ethereum Data Store

Return to DIDO CLI Background

A.1.1 The Ethereum Blockchain Data Flow

Return to Top

Ethereum, as well as most DIDO Platforms, use a Representational State Transfer (REST) model as the basis of their Application Programming Interface (API). Ethereum Node (i.e.,Clients) formulate a request in eXtensible Markup Language (XML) or JavaScript Object Notation (JSON) and send it to a RESTful API service that processes the request and calls the Ethereum Virtual Machine (EVM) in the Ethereum Node using a Remote Procedure Call (RPC). The EVM processes the request and returns any required data back to the API Service as part of the RPC. The API Service then creates an appropriate XML or JSON response and returns it to the client over Hypertext Transfer Protocol (HTTP) or Hypertext Transport Protocol Secure (HTTPS).

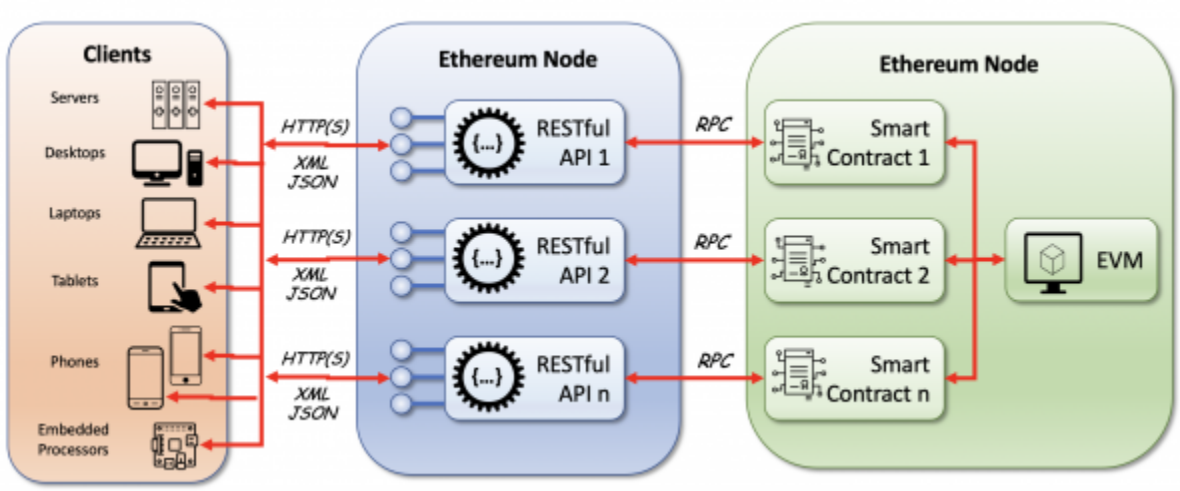

Figure 1: The Flow of Data from Clients to Ethereum Blockchain

A.1.2 Overview of Ethereum Blockchain Objects

Return to Top

There are a few basic building blocks included by Ethereum in the infrastructure to support the Ethereum Ecosystem: Blocks, Transactions (TX), Log Entries, Calls, and Trace.

There are two objects that support the development of domain-specific applications: Token and Contract.

There are also a couple of generic data structures that define important concepts within the environment using JavaScript Object Notation (JSON) (i.e., JSON Support: Arguments, Links and Traces.

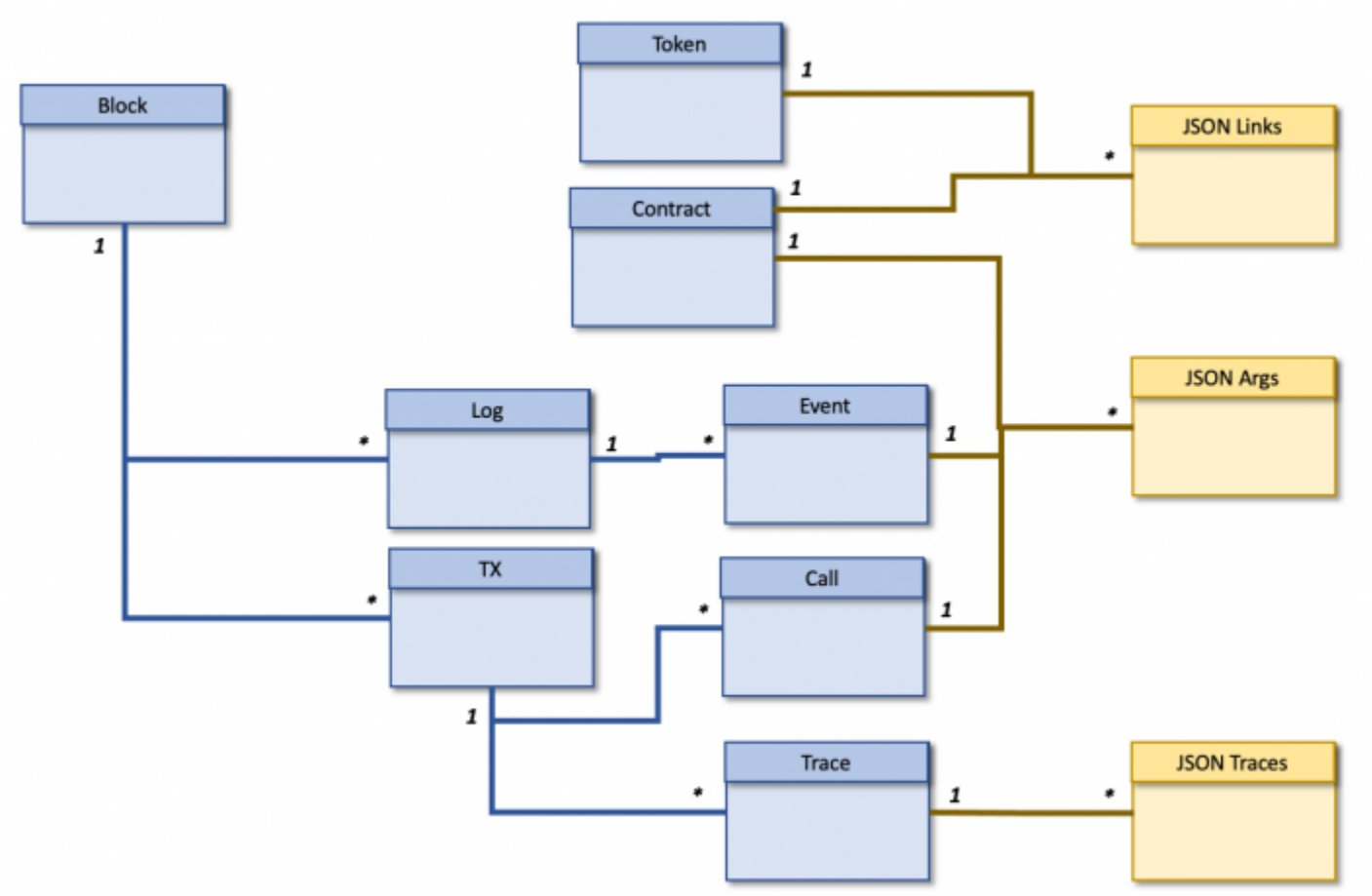

Figure 2: The Ethereum Blockchain Objects

- *Keep in mind, Ethereum is the database, smart contracts are the data tables, and transactions from wallets are the rows in each table.*[1)]

A.1.3 Details of Ethereum Blockchain Objects

Return to Top

Ethereum uses a Document Object Model (DOM) Data Model (DM) rather than a Relational Model (RM), however, the documents are shallow[2)] and lend themselves to mapping to the relational model well.

The use of Google's BigQuery was designed for analyzing data on the order of billions of rows, using a Structured Query Language (SQL)-like syntax. It runs on the Google Cloud Storage infrastructure and can be accessed with a REST-oriented application program interface (API).

Trying to understand and use BigQuery developers have been able to tab into the Ethereum Distributed Ledger Technology (DLT) (i.e., Blockchain) data using very close to standard SQL[3)].

Figure 3 provides a high-level flow of converting Ethereum Blockchain data into Google's BigQuery and as

a consequence accessible using SQL. In this workflow, an Ethereum Node is deployed within a container orchestrated by a Kubernetes Engine. there are two paths through the data flow: a Real-Time path and a Daily Path.

- In the Daily Path, a snapshot of the blockchain data is exported once a day, converted to Comma Separated Values (CSV) files, and then loaded into BigQuery where the data can be accessed using the BigQuery Console.
- In the Real-Time Path that has a lagtime delay to prevent orphaned blocks, the data is streamed to a publish/subscribe utility that can create either
  - An ETL Dataflow is loaded into BigQuery where data can be accessed using the BigQuery Console
  - An Application subscribes directly to a topic

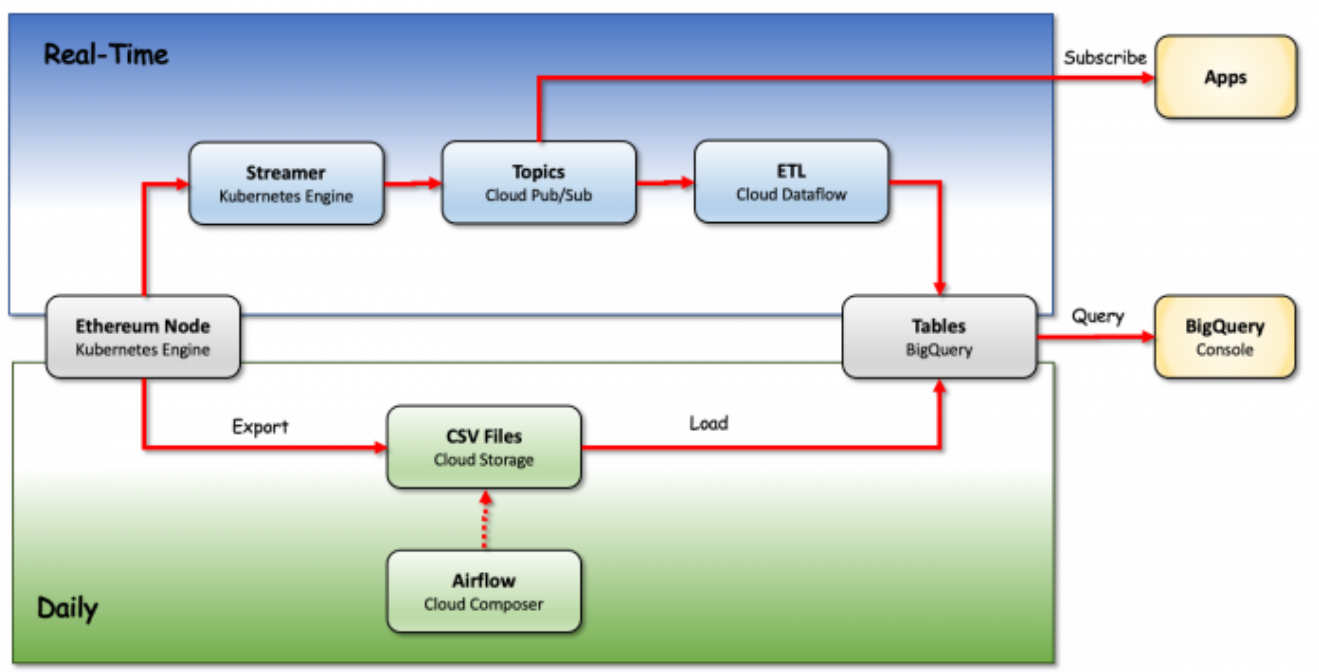

Figure 3: Blockchain ETL Architecture[4)]

A.1.4 BigQuery Data Objects

Return to Top

The following is a list of Ethereum Data Objects (i.e., tables) that are available using BigQuery using Structured Query Language (SQL).

1)

Andrew Hong, Hands-On Tutorials, Your guide to basic SQL while learning Ethereum at the same time, 17 April 2021, Accessed: 11 June 2021, https://towardsdatascience.com/your-guide-to-basic-sql-while-learning-ethereum-at-the-same-time-9eac17a05929

2)

Shallow documents do not have a lot of nesting. Almost all of the attributes occur and immediately under the document encapsulation. As the depth of nesting increases, the mapping of the documents to a relational model becomes more difficult, especially when trying to enforce normalization rules.

3)

Rif Kiamil, 22 February 2021, Medium, Accessed: 13 June 2021, https://medium.com/google-cloud/full-relational-diagram-for-ethereum-public-data-on-google-bigquery-2825fdf0fb0b

4)

Evgeny Medvedev, Medium, 7 October 2019, Accessed“ 14 June 2021, https://medium.com/google-cloud/live-ethereum-and-bitcoin-data-in-google-bigquery-and-pub-sub-765b71cd57b5