Data Ocean Architecture

Data within the ***Data Ocean*** is stored within two fully managed services on Google Cloud Platform.

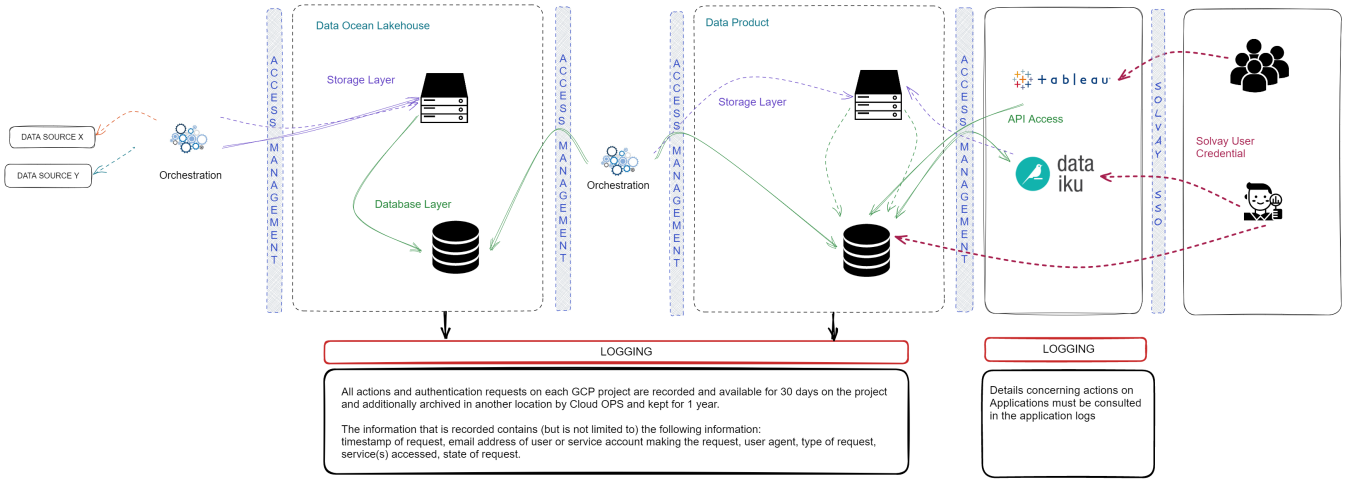
- Initial Raw Storage and Staging Area - Google Cloud Storage
- Database - Google BigQuery

Google Cloud Storage is used as the primary staging area for data being received from various source applications. Once data is loaded into Cloud Storage, it is then loaded into ***BigQuery*** in a raw state and then transformed and loaded into the ***Operational Data Store (ODS)*** dataset and eventually other datasets.

The data that transits into both Cloud Storage and BigQuery all passes over Google's Public API's. Information concerning the security can be referenced here: https://cloud.google.com/docs/security/encryption-in-transit#cio-level_summary

All data loaded into ***Data Ocean*** ultimately arrives over TCP (HTTPS) over TLS, and any other further actions or transformations performed on the data are under the same security restrictions. For any data stored on Google Cloud Storage or Google BigQuery, the data is encrypted at rest using Google's default encryption. At the moment Solvay does not used **Customer Supplied Encryption Keys.**

There is no ***end user*** (business user) access directly to data stored within the Data Ocean GCP Projects. Data within the Data Ocean that is exposed to business users and applications is done through ***Product*** GCP Projects. If possible the data will be exposed in the form of ***Authorized*** Views.



The current configuration of Google Cloud Platform resembles this

Network Traffic Legend

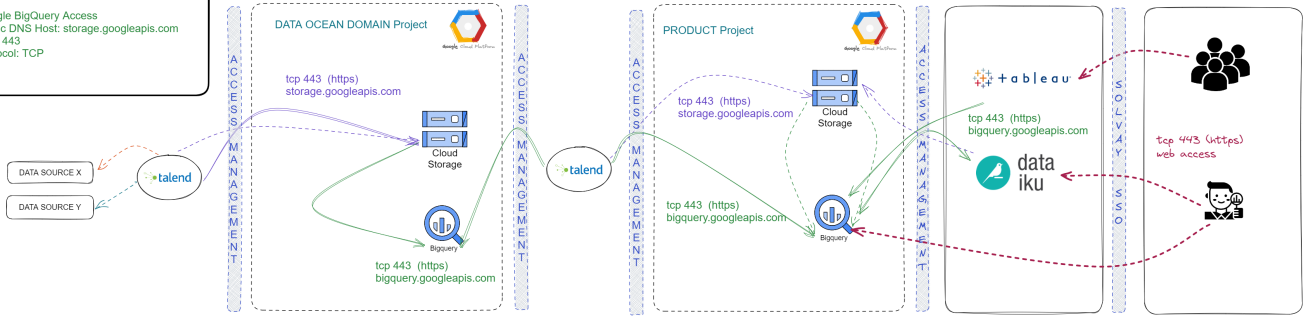
Google Cloud Storage Access
Public DNS Host: storage.googleapis.com
Port: 443
Protocol: TCP

Google BigQuery Access
Public DNS Host: storage.googleapis.com
Port: 443
Protocol: TCP

All data exchanges and data access within the GCP resources on the Data Ocean projects use Google's Cloud APIs that only accept secure requests using TLS encryption.

https://cloud.google.com/apis/docs/overview & https://cloud.google.com/docs/security/encryption-in-transit

Authentication credentials between Applications, such as Tableau, Dataiku, and BigQuery is done with GCP service account credentials that are rotated every year.



LOGGING AND REPORTING

All actions and authentication requests on each GCP project are recorded and available for 30 days on the project and additionally archived in another location by Cloud OPS and kept for 1 year.

The information that is recorded contains (but is not limited to) the following information:
timestamp of request, email address of user or service account making the request, user agent, type of request, service(s) accessed, state of request.