This case involves a media company that has decided to move their in-house data processing into BigQuery. This example is focused on security and compliance.

As part of the migration, they have been moving their data centers from on-prem to BigQuery in the cloud. They have a lot of concerns about security. Who has access to the data they are migrating into the cloud? How is access audited and logged? What kind of controls can be placed on top of that? And they are very concerned about data exfiltration. They are worried about potential bad actors within the company, who, as part of their role have access to certain data. They want to make sure that employees who have access to that data cannot then take the data, load it onto their own computer, or load it onto another cloud project, and from there, perhaps, take that data somewhere else.

A customer had this interesting business requirement...

- Capture data reading and update events to know who, what, when, and where.
- Separation of who manages the data and who can read the data.
- Allocate costs appropriately; costs to read/process vs. costs to store.
- Prevent exfiltration of data to other Google Cloud projects and to external systems.

We worked together to understand these business requirements and to help turn them into more technical requirements. We wanted to focus the technologies on the capabilities already available in BigQuery. So we introduced them to the concept of audit logs on Google Cloud, and specifically the default logs available from BigQuery. We presented them with Admin Logs that record creating and deleting datasets, and then the more detailed Access Logs that identify when people are reading datasets or perhaps even reading or accessing parts of the BigQuery UI.

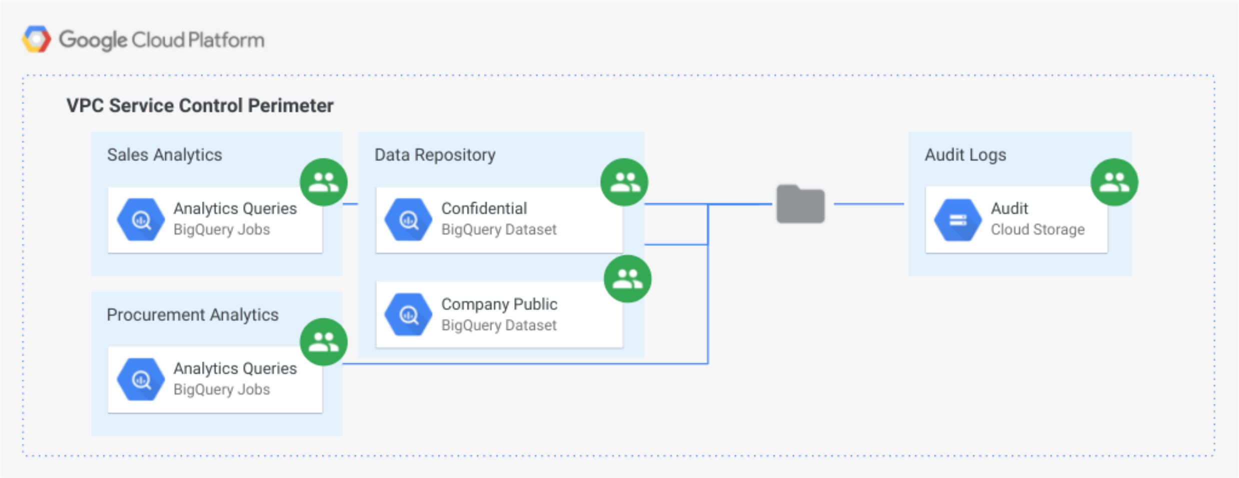
We encouraged them to have everything managed by IAM. We developed groups based on role. Then assigned members to groups. And established permissions and applied those to the groups based on role.

We mapped that to technical requirements like this…

Requirements

- All access to data should be captured in Audit Logs.
- All access to data should be managed via IAM.
- Configure service perimeters with VPC service controls.

And this is how we implemented that technical requirement.



Each group was isolated in separate projects and allowed limited access between them using VPC Service Controls. BigQuery allows separation of access by role, so we were able to limit some roles to only loading data and other to only running queries. Some groups were able to run queries in their own project using datasets for which they only had read access, and the data was stored in a separate repository. We made sure that at the folder level of the resource hierarchy, we had aggregated log exports enabled. That ensured that even if you were the owner of a project and had the ability to redirect exports, you wouldn't be able to do so with specific exports, because those rights were set at the folder level, where most team members didn't have access. So by using aggregated log exports we were able to scoop up all the logs, store them in cloud storage, and create a record of who is running what query at what time against what dataset.

The VPC perimeter enabled us to allow APIs within the perimeter to run and only talk with other APIs belonging to other projects within the same perimeter.

So if someone had a separate project and started a BigQuery job that was to read from a dataset within the perimeter, even though they have credentials and access to the dataset, they would not be able to use run the queries, because the APIs would not allow it at the perimeter.