Setting up a HIPAA-Aligned workload using Data Protection Toolkit

Solution Guide

Contents

Disclaimer

- This solution guide and accompanying templates, including the Data Protection Toolkit (DPT), provide a reference architecture, leading practices, and recommendations for Google Cloud customers. Please note that this guide does not constitute advice on implementing the appropriate administrative, technical, and physical safeguards required to implement Health Insurance Portability and Accountability Act (HIPAA) requirements.
- The customer is responsible for managing data and applications, including configuration and maintenance of services hosted utilizing Google Cloud. These responsibilities are further enumerated in section A.4 Customer Responsibilities.
- The scope of this solution guide is limited to providing security guidance for protecting and monitoring data within the in-scope resources defined as part of the Reference Architecture for a HIPAA-aligned Analytics and Artificial Intelligence and Machine Learning (AI/ML) Platform in Section 3.
- Implementation of the solution guide or reference architecture does not automatically cover any data assets that are stored or processed by other Google Cloud Storage services. Similar protective measures must be applied to all other data stored across the environment.
- The information mentioned in the solution guide is illustrative and does not constitute exhaustive guidance on HIPAA. The information must be read alongside the official documentation.
- The implementation of this solution may vary in customer environments based on the choice of products and configuration options.
- This solution guide can be used as an accelerator or framework and will need to be customized to include additional specific requirements and use cases to deploy HIPAA-aligned workloads.

1. Overview

1.1 Solution Guide

This solution guide covers the process and the guidelines to deploy a typical HIPAA-aligned workloadon Google Cloud (hereinafter, the Analytics and AI/ML Platform). This workload contains various recommended security configurations related to role-based access control, data protection and retention, audit logging, and monitoring, amongst others aligned to HIPAA Privacy and Security Rule requirements. This guide explains security configurations related to Google Cloud products and services which can be used to provision a minimum viable product (MVP) environment, which can be customized and expanded upon as a part of other workloads or use-cases. Furthermore, this guide also covers post-deployment verification steps in section 3.8 Post-Deployment Verification using the Google Cloud console to verify resources deployed and their corresponding security parameters.

The Analytics and AI/ML Platform example used in this guide covers a subset of Google Cloud resources used to build and deploy a HIPAA-aligned workload. While this solution guide will help accelerate the deployment of Google Cloud environments and applications in alignment with HIPAA, please customize the guidance to include specific additional requirements and other use cases as applicable to ensure that you continue to implement appropriate technical and administrative safeguards for your organization.

Note: The HIPAA-aligned Analytics and AI/ML platform template discussed in the following sections deploy the following resources: Cloud Storage buckets, BigQuery datasets, Cloud SQL Instances, Healthcare datasets (that interact with Healthcare API), a Deep Learning VM, Cloud Dataproc job, and Cloud Datalab module. It also enables the following APIs which can be used to assist in analytics and AI/ML functions: Cloud AutoML, Data Loss Prevention, Life Sciences API, and Cloud Speech-to-text. Additional services covered in the architecture can be integrated into the final solution as needed to build an entire workflow around the analytics and AI/ML workflow processing.

1.2 Data Protection Toolkit

The Data Protection Toolkit (further referred to as DPT) is an open-source utility, released by Google, for provisioning and managing Google Cloud projects. DPT combines infrastructure-as-code best practices, security configurations, and best practices for provisioning Google Cloud products into a comprehensive end-to-end framework. DPT's easy-to-use and declarative "deployment templates" (written in YAML or JSON) makes it intuitive to understand and easy to validate the deployment workflow even before its implementation. These templates help accelerate the provisioning and configuration of projects, resources, network infrastructure, access management, and monitoring, by leveraging tools like Terraform and gcloud. DPT can also be used to set up Forseti, an open source tool for continuous configuration monitoring of the deployed projects and their resources. To learn more about Forseti, visit its website. Due to its expressive deployment capabilities and integrated monitoring tools, DPT is a powerful tool for privacy, security, and compliance focused use cases.

DPT templates help in

- Deploying identical environments (e.g. development, test, and production) with minimal manual intervention.
- Minimizing build and deployment errors in comparison to manual builds.
- zero-downtime deployment, testing, and validation of Google Cloud workloads.
- Disaster recovery by enabling rapid deployment of failed workloads.
- Deploying infrastructure-related auditing and monitoring tools in parallel with workload deployment.
- Reducing maintenance costs by automating removal of unused resources in conjunction with capacity monitoring.

DPT templates can update or restore the deployments to the required state driving development efficiency. Also, changes to the DPT template can be tracked by maintaining it in a code repository. This drives accountability and maintains discipline and quality control.

Google has published DPT as an open-source repository, which can be cloned and used to deploy the templates. To learn more about DPT, refer to the DPT Repository on GitHub. DPT, while currently targeted to the healthcare industry, can also be used to support use-cases related to banking and finance, gaming, marketing and education.

2. HIPAA and Healthcare

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) set standards in the United States to protect individually identifiable health information. HIPAA applies to health plans, most healthcare providers, and healthcare clearinghouses - collectively known as “covered entities” - that manage protected health information (PHI) electronically and to persons or entities that perform certain functions on their behalf, known as “business associates”. The HIPAA Privacy Rule requires covered entities and their business associates to safeguard the privacy of PHI handled in any medium, while the HIPAA Security Rule obligates them to protect the confidentiality, integrity, and availability of PHI they create, receive, maintain, or transmit electronically with administrative, physical, and technical measures.

For additional details and to learn more about HIPAA on Google Cloud, please refer to A.1 HIPAA and Google Cloud.

3. Deploying a HIPAA-aligned Analytics and AI/ML Platform using DPT

This section elaborates on the Analytics and AI/ML platform example and describes the steps for deploying it using DPT in alignment with the HIPAA Privacy Rule and the HIPAA Security Rule

To support requirements from the above HIPAA Rules, an Analytics and AI/ML platform should be protected through the implementation of technical controls including, but not limited to the following features available through DPT :

- identity and access management
- infrastructure security
- data security
- audit logging
- labelling

The sections below describe how to implement the above mentioned technical security controls for a specific use-case using DPT. It includes the identified in-scope Google Cloud products and services described in the reference architecture below.

To learn more on best practices for setting up HIPAA-aligned workloads for other use cases, refer to this solution.

3.1 HIPAA-aligned Analytics and AI/ML Platform - Reference Architecture

This solution guide leverages the following security reference architecture for setting up the example Analytics and AI/ML platform in alignment with HIPAA requirements. Please note that:

- The platform processes and stores PHI and helps in carrying out analytics and AI/ML tasks using pre-installed tools and packages within Datalab and Deep Learning VM.

- This architecture is designed to provide a quick start environment that can be further customized for deploying more complex analytics and AI/ML use cases or additional Google Cloud products and services.

Though most of the components in the architecture below can be implemented using DPT templates, some of the capabilities, such as Cloud IAM and multi-factor authentication will require additional custom configuration apart from DPT for integration with existing internal and external systems.

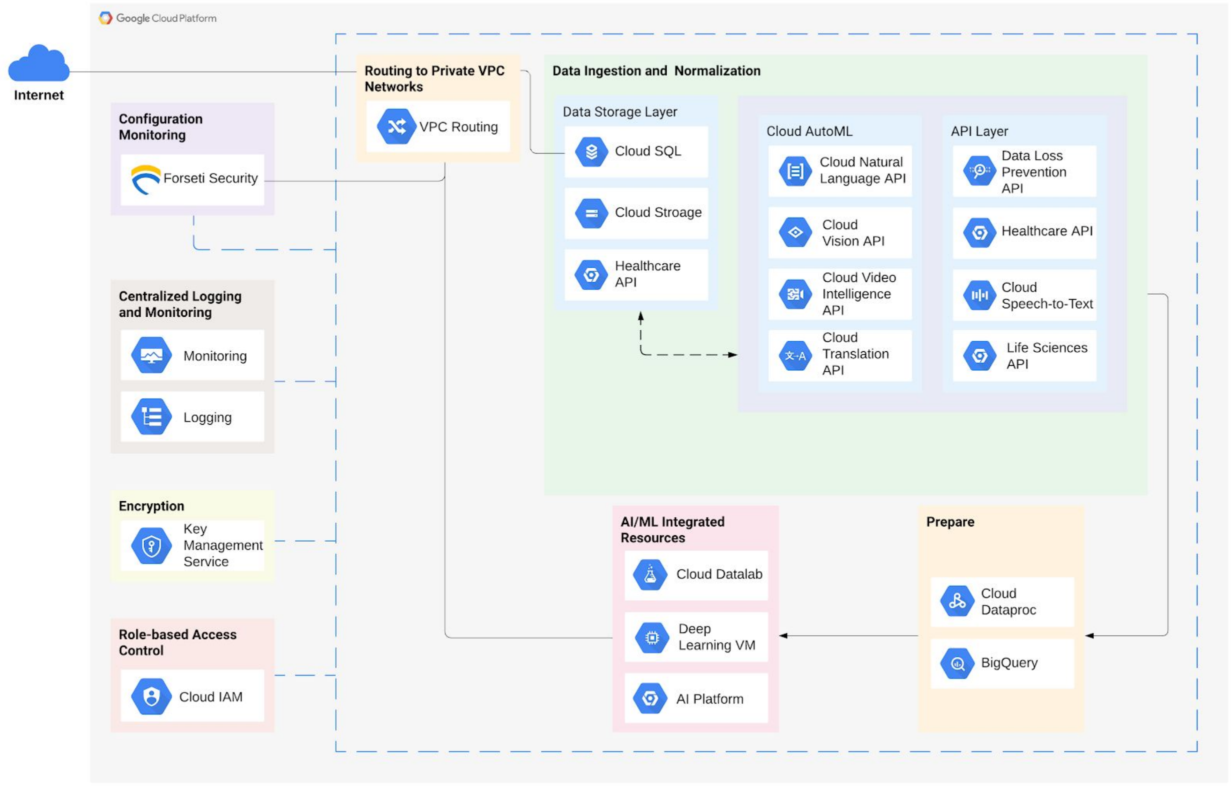


Figure 1: Sample - HIPAA-aligned Analytics and AI/ML Platform Architecture

In Figure 1, "Internet" does not mean that these resources are publicly accessible. This serves as a single point of entry for users. The entire architecture is deployed as two projects by the DPT template. The first project contains all the resources deployed for use by the Analytics and AI/ML platform. The second project contains resources used by Forseti to conduct configuration monitoring of resources deployed. The dashed lines box in the diagram encompasses the capabilities and resources of the Analytics and AI/ML platform while all the services linked to that box provide technical controls for the Analytic and AI/ML platform. And finally, the bi-directional dashed line in the Data Ingestion and Normalization section suggests that data hosted by GCP storage services can be modified using various APIs suggested in the diagram.

3.2 In-scope Product Guidance

The following section provides product details for each service and its role in the analytics and AI/ML architecture. Further, it demonstrates the security configurations for each product through DPT. The products deployed using DPT in this guide are covered by the Google Cloud BAA.

The Google Cloud resources deployed by the HIPAA-aligned Analytics and AI/ML Platform DPT template are:

- *Google Cloud Storage*

- *Google BigQuery*
- *Google Cloud SQL*
- *Google Cloud Healthcare API*
- *Deep Learning Virtual Machine (as a Google Compute Instance)*
- *Google Cloud Dataproc*
- *Google Cloud Datalab*

The DPT template also enables the following Google Cloud APIs which are HIPAA-enabled and can assist in delivering AI/ML services.

- *Cloud Natural Language API*
- *Cloud Vision API*
- *Cloud Video Intelligence API*
- *Cloud Translation API*
- *Data Loss Prevention API*
- *Cloud Speech-to-Text API*
- *Life Sciences API*

Resources such as the AI Platform Notebooks mentioned in the architecture diagram are not discussed in the solution guide as they are not currently supported by DPT. Refer to AI Platform's documentation to configure them after the template deployment.

Product Guidance - Reading Instructions

The product guidance has been divided into three sections for easy reading:

1. Product description
2. HIPAA Alignment for each product
    a. The HIPAA alignment for each product includes a table describing Default Configurations and User-Controlled Configuration options.
    b. Default Configurations are applied by the GCP service and are meant to represent a reasonable default for most applications. These are enabled when a service is provisioned with default configurations through the Cloud Console, or through the API or CLI interfaces.
    c. User Controlled Configurations can be applied using DPT, through the Cloud Console, or through the API or CLI interfaces.
3. DPT configurations for each product
    a. This table contains modular blocks of code referenced to the HIPAA configuration guidance for each product.

3.2.1 Google Cloud Storage

Google Cloud Storage provides worldwide, highly durable object storage that can scale up to exabytes of data. There are four storage classes - Multi-regional, Regional, Nearline, and

Coldline. The appropriate storage class can be chosen based on the business purpose and other requirements (e.g., for availability).

As part of the solution architecture, Cloud Storage buckets store the raw data ingested from the healthcare systems before they are normalized and imported into BigQuery.

Cloud Storage offers features like attribute-level access control using Cloud IAM Conditions, admin activity and event logging, encryption, object lifecycle management and versioning, etc.

To learn more about Google Cloud Storage and the parameters discussed below, refer to the Cloud Storage documentation and resource configuration respectively.

HIPAA Alignment for Google Cloud Storage

Security Rule: Technical Safeguards	*Default Configurations*	*User-Controlled Configurations (ex. via DPT)*
Identity and Access Management	Users in the owners' group of the project are allowed to perform any action allowed by the project's organization policies.	Use G Suite to create users and groups. Additional role-member bindings can be added as mentioned in the template to control access to the storage bucket.
Data Security	Accidentally deleted or overwritten objects cannot be retrieved as versioning is disabled by default.	Versioning is always enabled by default when deployed using DPT. Note: DPT has versioning as a mandatory parameter for deployment.
	Data is encrypted at-rest with Google managed security keys. Refer to the documentation on default encryption at rest for more information.	Custom encryption keys can be used instead of Google Cloud managed keys. Refer to Customer-managed Encryption Keys for more information.
	Labels are not provided by default.	Label parameter values need to be specified per requirements in the template.

DPT Template Configuration for Google Cloud Storage

Note: For options for the customizable parameters in the template below, please refer to Cloud Storage Guidance for Terraform. The configurable values in the below template are indicative only. Please modify it to match specific requirements in the context of usage.

Template (Please refer to accompanying *.yaml template for detailed configuration)

```
storage_buckets:
 # appropriate name must be chosen based on its purpose
- name: example-storage-bucket
   # IAM Role Binding
   # role-member bindings can be added/removed as required
   # Code Block 3.2.1.a
  _iam_members:
  - role: roles/editor
    member: user:user@domain
  - role: roles/viewer
    member: user:user@domain
# The Cloud KMS key configuration. If not specified, Google's default
encryption at rest is used.
   # Code Block 3.2.1.b
  # encryption:
  #   default_kms_key_name: (ex.{google_kms_crypto_key.gcs.self_link})
  location: EU
   # Code Block 3.2.1.c
  lifecycle_rule:
  - condition:
      age: #(e.g. 90)
    action:
      type: SetStorageClass
      storage_class: #(e.g.
STANDARD/REGIONAL/MULTI_REGIONAL/COLDLINE/NEARLINE)
   # Code Block 3.2.1.d
  labels:
    data_criticality: #(e.g. low, medium, high)
    data_type: #(e.g. phi, pii, gcslogs, auditlogs, statefiles, and general)
```

Identity and Access Management	**User access control**	**Refer to Code Block 3.2.1.a** **_iam_members** - Configuration for assigning roles to members and granting appropriate level permissions to the services **role:** Role to be assigned to the user **member:** G Suite Users or Groups to which the above role is assigned
Data Security	**Encryption**	**Refer to Code Block 3.2.1.b** **default_kms_key_name** - A custom Google Cloud KMS key that is used to encrypt objects added to the

		bucket. If a custom KMS key is not specified, Google Cloud default encryption will be used for the data at rest. See this whitepaper for more information on how Google encrypts data at rest by default and to understand the key management options.
	Information Lifecycle Management	**Refer to Code Block 3.2.1.c** **lifecycle_rule** - An array of objects where each object is a rule consisting of an action and a set of conditions. If multiple conditions are specified in a rule, an object has to match all of the conditions for the action to be taken. If multiple rules are specified with the same action, the action is taken when an object matches the condition(s) in any of the rules. Each rule should contain only one action.
	Labelling	**Refer to Code Block 3.2.1.d** **labels** - They are used to identify assets based on their classification. This can be used to identify the appropriate types of data stored within the service.

3.2.2 Google BigQuery

Google BigQuery is a serverless, highly scalable, and cost-effective data warehouse that can help in analyzing big data at high speed with zero operational overhead. BigQuery offers features like custom retention periods and enhanced query filtering to manage and debug workloads.

As part of the solution architecture, BigQuery is used as the central data warehouse ("data lake") to store and process data ingested from various sources, including Cloud SQL, Cloud Storage buckets, and Healthcare API datasets. Using BigQuery as a central repository makes it easier to leverage Data Visualization tools, such as Looker or Data Studio, and Data Transformational tools, such as Dataprep, Data Fusion, and Dataflow.

To learn more about BigQuery and the parameters discussed below, refer to the BigQuery documentation and resource configuration respectively.

HIPAA Alignment for Google BigQuery

Security Rule: Technical Safeguards	*Default Configurations*	*User-Controlled Configurations (ex. via DPT)*
Identity and Access Management	Users in the owners' group of the project are allowed to perform any action allowed by the project's organization policies.	Use G Suite to create users and groups. Additional role-member bindings can be added as mentioned in the template using the access block to control access to the storage bucket.
Data Security	Data is encrypted at-rest and in-transit, with Google managed encryption keys.	Custom encryption keys can be used instead of Google Cloud managed keys. For more information, refer to the Customer-managed Encryption Keys (CMEK) documentation. See this whitepaper for more information on how Google encrypts data at rest by default and to understand the key management options.
	Labels are not provided by default.	Label parameter values can be specified to identify assets as per the required values in the template.

DPT Template Configurations for Google BigQuery

Note: For options for the customizable parameters in the template below, please refer to BigQuery guidance for Terraform. The configurable values in the below template are indicative only. Please modify it to match specific requirements in the context of usage.

Template (Please refer to accompanying *.yaml template for detailed configuration)

```
bigquery_datasets:
  # appropriate id must be chosen based on its purpose
- dataset_id: example_dataset
   # Code Block 3.2.2.a
  access:
  - special_group: #(e.g. allAuthenticatedUsers, projectOwners,
projectWriters or projectReaders)
    role: READER
  - user_by_email: user@domain
  role: roles/bigquery.dataViewer
# The Cloud KMS key configuration. If not uncommented and specified, default
keys are used.
 # Code Block 3.2.2.b
 # default_encryption_configuration:
 # Code Block 3.2.2.c
 #   kms_key_name:  (ex.{google_kms_crypto_key.gcs.self_link})
 location: EU
 # Code Block 3.2.2.d
 labels:
  data_criticality: #(e.g. low, medium, high)
  datatype: #(e.g. phi, pii, gcslogs, auditlogs, statefiles, and general)
```

Identity and Access Management	**User access control**	**Refer to Code Block 3.2.2.a** **access** - An array of objects that define dataset access for one or more entities. Each object has a role and a user entity to which the role must be assigned. These user entities can be domain, group_by_email, user_by_email or special groups.
Data Security	**Encryption & Key Management**	**Refer to Code Block 3.2.2.b** **default_encryption_configuration** - The default encryption key for all tables in the dataset. Once this property is set, all newly created partitioned tables in the dataset will have an encryption key set to this value, unless table creation request (or query) overrides the key. It has the following parameter: **Refer to Code Block 3.2.2.c** **kms_key_name -** Google Cloud KMS encryption key that will be used to protect the destination BigQuery table. If a custom KMS key is not specified, Google Cloud default encryption will be used for the data at rest. See this whitepaper for more information on

		how Google encrypts data at rest by default and to understand the key management options.
	Labelling	**Refer to Code Block 3.2.2.d** **labels** - Labels are used to identify assets based on their classification. This can be used to identify the appropriate types of data stored within the service.

3.2.3 Google Cloud SQL

Google Cloud SQL is a fully managed relational database service compatible with applications using MySQL, PostgreSQL, and SQL Server. Cloud SQL ensures reliability and security through regular security updates and by providing options for high availability, automated backups, etc.

SQL Instances in this reference architecture can be used to store raw data imported from various on-prem sources and also as application datastores handling large amounts of data transactions that require consistency. Cloud SQL Instances are also used as one of the inputs from the aggregated data lake hosted in BigQuery. Compared to BigQuery, which is designed for large-scale analyses, Cloud SQL is primarily used as an application datastore handling large amounts of data transactions requiring consistency.

To learn more about Cloud SQL and the parameters discussed below, refer to the Cloud SQL documentation and resource configuration respectively.

HIPAA Alignment for Google Cloud SQL

Security Rule: Technical Safeguards	*Default Configurations*	*User-Controlled Configurations (ex. via DPT)*
Data Security	Data is encrypted at-rest and in-transit, with Google managed encryption keys.	Custom encryption keys can be used instead of Google Cloud managed keys. Refer Customer-managed Encryption Keys for more information. See this whitepaper for more information on how Google encrypts data at rest by default and to understand the key management options.

Infrastructure Security	Instances should have public and private network access defined. These are not enabled by default.	Instances can either be confined to a private network or made publicly accessible by modifying the network configurations in the template. Cloud SQL instance and its replica are created in a private VPC network as configured in the template.
Resilience	A read/read-only replica is not created by default for Cloud SQL.	A failover replica is configured for the main SQL instance which ensures the availability of data in case of a physical or/and a technical incident.

DPT Template Configuration for Google Cloud SQL

Note: For options for the customizable parameters in the template below, please refer to Cloud SQL guidance for Terraform. The configurable values in the below template are indicative only. Please modify it to match specific requirements in the context of usage.

Template (Please refer to accompanying *.yaml template for detailed configuration)

```
terraform_deployments:
  resources:
    config:
      resource:
      - google_sql_database_instance:
          instance:
            name: sql-instance-name
            region: europe-west3
            depends_on:
            - google_service_networking_connection.private_vpc_connection
        # Code Block 3.2.3.a
            #  encryption_key_name: # {full path to the encryption key used
for the CMEK disk encryption}
            settings:
              availability_type: ZONAL
              tier: db-f1-micro
        # Code Block 3.2.3.b
              ip_configuration:
                ipv4_enabled: false
                private_network:
${google_compute_network.private_network.self_link}
        # Code Block 3.2.3.c
              backup_configuration:
                binary_log_enabled: true
                enabled: true
          replica-instance:
            name: sql-replica-instance-name
            region: europe-west3
            master_instance_name:
${google_sql_database_instance.instance.name}
            depends_on:
            - google_service_networking_connection.private_vpc_connection
            settings:
              availability_type: ZONAL
              tier: db-f1-micro
              ip_configuration:
                ipv4_enabled: false
                private_network:
${google_compute_network.private_network.self_link}
        # Code Block 3.2.3.d
            replica_configuration:
              failover_target: true
              master_heartbeat_period: 60000
```

Data Security	**Encryption and Key Management**	**Refer to Code Block 3.2.3.a** **encryption_key_name** - The full path to the encryption key used for the CMEK disk encryption. Refer to this link for more information. If a custom KMS key is not specified, Google Cloud default encryption will be used for the data at rest. See this whitepaper for more information on how Google encrypts data at rest by default and to understand the key management options.
Infrastructure Security	**Network Security**	**Refer to Code Block 3.2.3.b** **settings.ip_configuration -** Configuration for the IP traffic directed to the database instance. It is critical to have the best settings for the traffic to ensure security. It has the following parameters: ● **ipv4_enabled** - Whether this Cloud SQL instance should be assigned a public IPV4

		address. Either ipv4_enabled must be enabled, or a private_network must be configured. ● **private_network** - The VPC network from which the Cloud SQL instance is accessible for private IP. For example, projects/myProject/global/networks/default. Specifying a network enables private IP. Either ipv4_enabled must be enabled, or a private_network must be configured. This setting can be updated, but it cannot be removed after it is set.
Resilience	**Disaster Recovery**	**Refer to Code Block 3.2.3.c** **settings.backup_configuration** - Configuration for backup of the database for added reliability. It has the following parameters: ● **binary_log_enabled** - True if binary logging is enabled. ● **enabled** - (Optional) True if backup configuration is enabled. ● **start_time** - HH:MM format time indicating when backup configuration starts.
		Refer to Code Block 3.2.3.d **replica_configuration -** Configuration for the replication process and associated operations for the database. It has the following parameters:- ● **failover_target -** Specifies if the replica is the failover target, this means if the master database instance

		fails due to some reason, the replica will take over as the new master instance. ● **master_heartbeat_period -** Time in milliseconds between replication heartbeats.

3.2.4 Google Cloud Healthcare API

Google Cloud Healthcare API bridges the gap between healthcare systems and applications on Google Cloud. The purpose of the Cloud Healthcare API is to ingest resources in healthcare-specific data formats, such as HL7v2, DICOM, and FHIR. The Cloud Healthcare API also provides de-identification capabilities which can be used to redact or obfuscate personal information in healthcare data before aggregation in BigQuery, enabling an extra layer of data privacy.

To learn more about Cloud Healthcare API and the parameters discussed below, refer to the Cloud Healthcare API documentation and resource configuration respectively.

HIPAA Alignment for Google Cloud Healthcare API

Security Rule: Technical Safeguards	*Default Configurations*	*User-Controlled Configurations (ex. via DPT)*
Identity and Access Management	Users in the owners' group of the project are allowed to perform any action allowed by the project's organization policies.	The template here provides guidance for adding additional role member bindings to the healthcare dataset and its datastores.
Privacy Rule: De-identification of PHI	*Default Configurations*	*User-Controlled Configurations (ex. via DPT)*
Data Security	The methods are not applied by default and need to be chosen based on the type of data.	Healthcare API's methods for de-identification must be used when appropriate based on the data residing in the healthcare datasets. See the Healthcare API Data De-Identification Guide for more information.

DPT Template Configuration for Google Cloud Healthcare API

Note: For options for the customizable parameters in the template below, please refer to Google Cloud documentation for [Cloud Healthcare API](). The configurable values in the below template are indicative only. Please modify it to match specific requirements in the context of usage.

Template (Please refer to accompanying *.yaml template for detailed configuration)

```
healthcare_datasets:
  # appropriate name must be chosen based on its purpose
- name: healthcare-datasets
  location: us-central1
    # IAM Role Binding
    # role-member bindings can be added/removed as required
   # Code Block 3.2.4.a
  _iam_members:
  - role: roles/healthcare.datasetViewer
    member: user:user@domain
  _dicom_stores:
      # appropriate name must be chosen based on its purpose
  - name: dicom-store
        # role-member bindings can be added/removed as required
   # Code Block 3.2.4.a
    _iam_members:
    - role: roles/healthcare.dicomEditor
      member: user:user@domain
    - role: roles/healthcare.dicomStoreAdmin
      member: user:user@domain
  _fhir_stores:
      # appropriate name must be chosen based on its purpose
  - name: fhir-store
        # role-member bindings can be added/removed as required
   # Code Block 3.2.4.a
    _iam_members:
    - role: roles/healthcare.fhirResourceReader
      member: user:user@domain
    - role: roles/healthcare.fhirResourceEditor
      member: user:user@domain
  _hl7_v2_stores:
      # appropriate name must be chosen based on its purpose
  - name: hl7-v2-store
        # role-member bindings can be added/removed as required
   # Code Block 3.2.4.a
    _iam_members:
    - role: roles/healthcare.hl7V2StoreAdmin
      member: user:user@domain
    - role: roles/healthcare.hl7V2Ingest
      member: user:user@domain
    - role: roles/healthcare.hl7V2Editor
```

`member: user:user@domain`		
Identity and Access Management	**User Access Control**	**Refer to Code Blocks 3.2.4.a** **_iam_members** - Configuration for assigning roles to members and granting appropriate level permissions to the services **role:** Role to be assigned to the user **member:** G Suite Users or Groups to which the above role is assigned
Data Security	**De-identification/Anonymization**	Applications can call specific methods within the Healthcare API to de-identify data based on specific datastores. This is a capability which can be access by applications dealing with sensitive ePHI by methods such as dicomStores.deidentify

3.2.5 Deep Learning VM

Deep learning VMs mentioned in this architecture are a set of preconfigured Debian 9-based Compute Engine virtual machine images optimized for data science and machine learning tasks. All images come with key ML frameworks and tools pre-installed with integrated support for JupyterLab, and can be used out of the box on instances with GPUs to accelerate the data processing tasks.

Access to the VM instance should be restricted to secure shell (SSH) or remote desktop protocol (RDP) via secure key-based access. To support resilience, snapshots of the Compute Engine instances should be enabled to back up data periodically.

To learn more about Compute Engine, Deep Learning VMs and the parameters discussed below, refer to the Compute Engine documentation, Deep Learning VM documentation, and resource configuration respectively.

HIPAA Alignment for Deep Learning VM

Security Rule: Technical Safeguards	*Default Configurations*	*User-Controlled Configurations (ex. via DPT)*
Identity and Access Management	Users in the owners' group of the project are allowed to perform any action allowed by the project's	Use G Suite to create users and groups. Additional role-member bindings can be added as mentioned

	organization policies.	in the template to control access to the compute instance.
Resilience	A default disk is attached to the compute instance. The compute instance can use any number of disks as required.	The template can be configured to attach a compute disk to the compute instance and enable its backup by a compute snapshot to recover data in case of a technical incident.
Data Security	Data is encrypted at-rest and in-transit, with Google managing the security keys.	Custom encryption keys can be used for encryption of data stored on Google Compute Engine disks. See Customer-managed Encryption Keys for more information. See this whitepaper for more information on how Google encrypts data at rest by default and to understand the key management options.
	Labels are not provided by default.	Label parameter values need to be specified as per the given values in the example as guided by the template.
Infrastructure Security	The instance can either be confined to use a private network or just a subnet of that network. Also, the instance may be assigned a public IP.	The VM instance is created under a subnet of a user-defined VPC network as configured in the template and public access can be restricted as required.

DPT Template Configuration for Google Cloud Instance

Note: For options for the customizable parameters in the Deep Learning VM Instance template below, please refer to Compute Instance for Terraform, since it is a compute instance with a Deep Learning VM image. The configurable values in the below template are indicative only. Please modify it to match specific requirements in the context of usage.

Template (Please refer to accompanying *.yaml template for detailed configuration)

```
compute_instances:
- name: deep-learning-vm-name
  zone: europe-west3-c
  machine_type: n1-standard-1
   # Code Block 4.5.a
  _iam_members:
  - role: roles/editor
    member: user:user@domain
  - role: roles/viewer
    member: user:user@domain
   # Code Block 4.5.b
  attached_disk:
    source: ${google_compute_disk.deep-learning-vm-disk.self_link}
    mode: READ_ONLY
  # A 256-bit customer supported key stored with them.
    # disk_encryption_key_raw:
  # A key stored on Google Cloud KMS.
    # kms_key_self_link:  (ex.{google_kms_crypto_key.gcs.self_link})
   # Code Block 4.5.c
  deletion_protection: false
  boot_disk:
    auto_delete: false
    mode: READ_WRITE           # can be changed to READ_WRITE or READ_ONLY
    initialize_params:
    # Deep Learning VM containing tensorflow pre-installed as example. Use
"gcloud compute images list --project deeplearning-platform-release
--no-standard-images" to find a list of more VMs users can use.
    # using VMs with GPU requires specific configurations. Check the Deep
Learning VM documentation for more info. Image link working as of 4/1/2020
      image:
https://www.googleapis.com/compute/v1/projects/deeplearning-platform-release
/global/images/tf2-latest-cpu-20200227
  # Updated link as on 29-Apr-2020
   # Code Block 4.5.d
  # A 256-bit customer supported key stored with them.
    # disk_encryption_key_raw:
  # A key stored on Google Cloud KMS.
    # kms_key_self_link: (ex.{google_kms_crypto_key.gcs.self_link})
   # Code Block 4.5.e
  labels:
    data_criticality: #(e.g. low, medium, high)
    datatype: #(e.g. phi, pii, gcslogs, auditlogs, statefiles, and general)
   # Code Block 4.5.f
  network_interface:
    subnetwork:
${google_compute_subnetwork.deep-learning-vm-subnetwork.self_link}
```

```
 service_account:
  email: ${google_service_account.deep-vm-service-account.email}
  scopes:
  - bigquery
  - sql-admin
  - userinfo-email
  - compute-ro
  - storage-ro

 # allow_stopping_for_update: true / false
 # enable_secure_boot: true
# Custom created service account

service_accounts:
- account_id: deep-vm-service-account

terraform_deployments:
 resources:
  config:
   resource:
   - google_compute_snapshot:
     deep-learning-vm-snapshot:
      name: deep-learning-vm-snapshot
      source_disk: ${google_compute_disk.deep-learning-vm-disk.name}
      zone: europe-west3-c
      labels:
       project: aiml-project-name
       connected_disk:
${google_compute_disk.deep-learning-vm-disk.name}
     # Customer Managed Keys must be configured here
      # snapshot_encryption_key:
      #  raw_key:
      #  sha256:
   - google_compute_disk:
     deep-learning-vm-disk:
      name: deep-learning-vm-disk
      type: pd-ssd
      zone: europe-west3-c
      labels:
       project: project-name
       connected_instance: deep-learning-vm-name
      physical_block_size_bytes: 4096
     # Customer Managed Keys must be configured here
      # disk_encryption_key:
      #  raw_key:
      #  sha256:
      #  kms_key_self_link:
```

```
(ex.{google_kms_crypto_key.gcs.self_link})
      - google_compute_subnetwork:
        - deep-learning-vm-subnetwork:
          - name: deep-learning-vm-subnet
            network: ${google_compute_network.private_network.self_link}
            region: europe-west3
            ip_cidr_range: 10.3.0.0/16
```

Identity and Access Management	**User access control**	**Refer to Code Block 3.2.5.a** **iam_members** - Member role for the user. G Suite users/groups and Cloud IAM roles can be used to control access.
Resilience	**Disaster Recovery**	**Refer to Code Block 3.2.5.b** **attached_disk** - An array of objects where each object defines a disk attached to the compute instance.
Data Security	**Information Lifecycle Management**	**Refer to Code Block 3.2.5.c** **deletion_protection** - This feature ensures that the compute instance cannot be deleted and defaults to **false**. Refer to this document for more information. **auto_delete** - Whether the disk attached to the compute instance is automatically deleted after the instance is deleted or not. Defaults to **true**.
	Encryption & Key Management	**Refer to Code Block 3.2.5.d** **disk_encryption_key_raw** - A 256-bit custom managed encryption key to encrypt this disk. If a custom key is not specified, Google Cloud default encryption will be used for the data at rest. See this whitepaper for more information on how Google encrypts data at rest by default and to understand the key management options.

		kms_key_self_link - A key stored on Google Cloud KMS. If a custom KMS key is not specified, Google Cloud default encryption will be used for the data at rest. See this whitepaper for more information on how Google encrypts data at rest by default and to understand the key management options.
	Labelling	**Refer to Code Block 3.2.5.e** Labels are used to identify assets based on their classification. This can be used to identify the appropriate types of data stored within the service.
Infrastructure Security	**Network Security**	**Refer to Code Block 3.2.5.f** **network_interface.network** - The name or self_link of the network to attach this instance to. The network must exist in the same region this instance will be created in. Either network or subnetwork must be provided. **network_interface.subnetwork** - The name or self_link of the subnetwork to attach this instance to. The subnetwork must exist in the same region this instance will be created in. Either network or subnetwork must be provided.

3.2.6 Google Cloud Dataproc

Google Cloud Dataproc is a managed Apache Spark and Apache Hadoop service that lets you take advantage of open source data tools for batch processing, querying, streaming, and machine learning. Dataproc automation helps you create clusters quickly, manage them easily, and save money by turning clusters off when you don't need them. Dataproc uses image templates to bundle operating system, big data components (Hadoop, Spark, Hive, and Pig), and Google Cloud Platform connectors into a package deployed on a cluster.

A Dataproc cluster is deployed as a part of the architecture which can be used to execute various big data jobs such as Hadoop, Spark, Hive and Pig jobs as necessary.

To learn more about Cloud Dataproc and the parameters discussed below, refer to the Cloud Dataproc Documentation and resource configuration respectively.

HIPAA Alignment for Google Cloud Dataproc

Security Rule: Technical Safeguards	*Default Configurations*	*User-Controlled Configurations (ex. via DPT)*
Data Security	Data is encrypted at-rest and in-transit, with Google managed encryption keys.	Custom encryption keys can be used for encryption of data stored on Google Compute Engine disks. See Customer-managed Encryption Keys for more information. See this whitepaper for more information on how Google encrypts data at rest by default and to understand the key management options.
Infrastructure Security	Auto-scaling is available but not enabled by default for Cloud Dataproc.	The auto-scaling policy attached to the cluster can be edited for scaling boundaries, frequency, and aggressiveness to provide fine-grained control over cluster resources throughout cluster lifetime.
Identity and Access Management	Users in the owners' group of the project are allowed to perform any action allowed by the project's organization policies.	Use G Suite to create users and groups. Additional role-member bindings can be added as mentioned in the template to control access to the Dataproc cluster.

DPT Template Configuration for Google Cloud Dataproc

Note: For options for the customizable parameters in the template below, please refer to Terraform documentation for Dataproc Cluster and Dataproc Job. The configurable values in the below template are indicative only. Please modify it to match specific requirements in the context of usage.

Template (Please refer to accompanying *.yaml template for detailed configuration)

```
terraform_deployments:
  resources:
    config:
      resource:
      # Dataproc cluster
      - google_dataproc_cluster:
          mycluster:
            name: dataproc-cluster-name
            region: europe-west3
            labels:
              project: aiml-project-name
            cluster_config:
              master_config:
                num_instances: 1
                machine_type: n1-standard-1
                disk_config:
                  boot_disk_type: pd-ssd
                  boot_disk_size_gb: 15
              worker_config:
                num_instances: 2
                machine_type: n1-standard-1
                min_cpu_platform: Intel Skylake
                disk_config:
                  boot_disk_size_gb: 15
                  num_local_ssds: 1
              preemptible_worker_config:
                num_instances: 0
              software_config:
                image_version: 1.3.7-deb9
                override_properties:
                  dataproc:dataproc.allow.zero.workers: 'true'
      # KMS keys configuration can be enabled. Refer
"https://www.terraform.io/docs/providers/google/r/dataproc_cluster.html"
   # Code Block 3.2.6.a
              #  security_config:
              #    kerberos_config:
              #      kms_key_uri:
(ex.{google_kms_crypto_key.gcs.self_link})
              #      root_principal_password_uri: bucketId/o/objectId
   # Code Block 3.2.6.b
              autoscaling_config:
                policy_uri: ${google_dataproc_autoscaling_policy.asp.name}
              gce_cluster_config:
                service_account:
dataproc-service-account@aiml-project-name.iam.gserviceaccount.com
      - google_dataproc_autoscaling_policy:
          asp:
            policy_id: dataprocs-policies
            location: europe-west3
            worker_config:
              min_instances: 2
              max_instances: 5
            basic_algorithm:
              yarn_config:
                graceful_decommission_timeout: 30s
                scale_up_factor: 0.5
                scale_down_factor: 0.5
    # IAM bindings for dataprocs
   # Code Block 3.2.6.c
      - google_dataproc_cluster_iam_member:
        - editor:
          - cluster: ${google_dataproc_cluster.mycluster.name}
            member: user:user@domain
            role: roles/editor
            region: europe-west3
        - viewer:
          - cluster: ${google_dataproc_cluster.mycluster.name}
            member: user:user@domain
            role: roles/viewer
            region: europe-west3
service_account:
- account_id: dataproc-service-account
```

Data Security	**Encryption & Key Management**	**Refer to Code Block 3.2.6.a** **security_config.kerberos_config.kms_key_uri -** The URI of the KMS key used to encrypt various sensitive files. It is a key stored on Google Cloud KMS. It is an attribute mentioned under the security_config block of the template. If a custom KMS key is not specified, Google Cloud default encryption will be used for the data at rest. See this whitepaper for more information on how Google encrypts data at rest by default and to understand the key management options.
Infrastructure Security	**Capacity Management**	**Refer to Code Block 3.2.6.b**

		cluster_config.autoscaling_config.policy_uri - The autoscaling policy config associated with the cluster.
Identity and Access Management	**User access control**	**Refer to Code Block 3.2.6.c** **google_dataproc_cluster_iam_member** - It is a list of role-member bindings. Each item in a list is populated with a role and a list of members that are granted the role.

3.2.7 Google Cloud Datalab

Google Cloud Datalab is a powerful interactive tool, built on Jupyter, to explore, analyze, transform, and visualize data and build machine learning models on Google Cloud Platform. It runs on Compute Engine and connects to multiple cloud services easily so you can focus on your data science tasks.

To learn more about Cloud Datalab and parameters discussed below, refer to the Cloud Datalab Documentation and module configuration respectively.

HIPAA Alignment for Google Cloud Datalab

Security Rule: Technical Safeguards	***Default Configurations***	***User-Controlled Configurations (ex. via DPT)***
Infrastructure Security	The datalab module is not attached to the network by default and is available via external IP.	The module can either be confined to use a private network or just a subnet of that network. Absence of configuration for the network would fail the deployment.
Data Security	The datalab module has a persistent disk provisioned by default and its contents are automatically backed-up into Cloud Storage.	NA - This is enabled by default when provisioning Datalab. This can be disabled but is not recommended. DPT retains the default configuration as enabled.

DPT Template Configuration for Google Cloud Datalab

Note: For options for the customizable parameters in the template below, please refer to Terraform documentation for Datalab.The configurable values in the template below are indicative only. Please modify it to match specific requirements in the context of usage.

Template (Please refer to accompanying *.yaml template for detailed configuration)

```
terraform_deployments:
  resources:
    config:
      module:
    # Refer
"https://github.com/terraform-google-modules/terraform-google-datalab" for
more information.
      - datalab:
        - datalab_user_email: user@domain
        # adding gpu requires a quota assigned. Refer to link
"https://github.com/terraform-google-modules/terraform-google-datalab/tree/m
aster/examples/basic" for furthur information.
          project_id: aiml-project-name
          source: terraform-google-modules/datalab/google//modules/instance
   # Code Block 3.2.7.a
          network_name: ${google_compute_network.private_network.name}
          subnet_name: ${google_compute_subnetwork.datalab-subnetwork.name}
          version: "~> 1.0"
          zone: europe-west3-c
        # custom service account with the deployer given serviceAccountUser
role
          service_account:
${google_service_account.datalab-service-account.email}
          datalab_enable_swap: true
   # Code Block 3.2.7.b
          create_disk: true
          datalab_enable_backup: true
          datalab_console_log_level: "warn"
          datalab_idle_timeout: "60m"
      resource:
      - google_compute_subnetwork:
        - datalab-subnetwork:
          - name: datalab-subnet
            network: ${google_compute_network.private_network.self_link}
            region: europe-west3
            ip_cidr_range: 10.2.0.0/16
# Custom created service account
service_accounts:
- account_id: datalab-service-account
```

Infrastructure Security	**Network Security**	**Refer to code block 3.2.7.a** **network_name** - The VPC network under which the Cloud Datalab instance is created. This setting must always be configured.

		subnet_name - The subnet under the previously mentioned VPC network under which the Cloud Datalab instance is created. This setting must always be configured.
Data Security	**Information Lifecycle Management**	**Refer to code block 3.2.7.b** **create_disk** - Created a persistent data disk and attaches it to the datalab module. **datalab_enable_backup** - Automatically does backup of the disk's contents to Cloud Storage

3.3 Environment Setup

DPT can be run locally on a computer or by using Google Cloud Shell.

Prior to running DPT locally, the following tools must be installed:

- Bazel - An open-source build and test tool
- Terraform - An infrastructure-as-code provisioning tool
- Cloud SDK - A set of tools for managing resources and applications hosted on Google Cloud.
- Git - A distributed version control system.

This step is not required when using Google Cloud Shell, as the required tools are already installed and ready to use.

3.4 DPT Access Control

The access control for DPT is covered under two sections to enhance the security of the deployments. Deploying DPT requires 'owner' (privileged) rule at the organisation or folder level (https://cloud.google.com/resource-manager/docs/cloud-platform-resource-hierarchy) to deploy resources. Considering privileged access security and separation of access, the access required to deploy resources should be separated from the access required to operate them.

Sections 3.4.1 and 3.4.2 explain in detail the access grants before, during and through the deployment of resources. This approach ensures that only required accesses are granted based on requirement in each stage of the project creation lifecycle. The access control of resources during the project lifecycle is not covered under this solution guide.

3.4.1 Pre-deployment Access Control

Before a template is deployed, DPT requires creation of two groups for each project in the template.

- ***Owner***: {PROJECT_ID}-owners@{DOMAIN}. This group is granted the owner's role for the project, which allows members to do anything permitted by organization policies within the project. Additions to the owner's group should be for a short term and controlled tightly.

- ***Auditor***: {PROJECT_ID}-auditors@{DOMAIN}. The auditor's group is granted the permission to list resources, view IAM configuration, and view contents of audit logs, but not to view any hosted data. If there are multiple data projects, it is advisable to maintain a single auditor's group across all projects.

DPT needs 'owner' permissions for the projects' in the template to provision their resources. So initially, to grant provisioning access, the user identity deploying the template is temporarily added to the owners' group for provisioning projects and resources.

Using a service account: It is ideal to use a service account to deploy a DPT workload rather than a user account, as a service account can be granted the minimum set of permissions required to deploy the DPT template (*Billing User, Project Billing Manager* and *Project Creator*) and its scopes can be restricted for interacting with necessary services only. Similarly, this service account must be added to the project's owners' groups for provisioning.

The *Service Account User* assumes the roles of the service account having owner permissions to deploy the workloads. The time period should be restricted to the period of deployment only.

3.4.2 Post-deployment Access Control

Post deployment of a template, DPT removes the deploying user (or user identity) from the owners' groups of all the projects in the template, effectively restricting further access to the projects. This ensures that only specific pre-approved owners continue to have control post-deployment.

Note: DPT grants roles and permissions to users, groups, and entities (e.g., service accounts). To further customize access after deployment is complete, user groups should be created to control access to the projects and their underlying resources using Google G Suite Admin Console, Cloud Identity, or Google Cloud Directory Sync (GCDS). These user groups can be granted custom roles and permissions using Cloud IAM and conditional access policies. For further information, please refer to the Cloud IAM documentation.

3.5 Deployment Modes

The section demonstrates the deployment of the HIPAA-aligned Analytics and AI/ML Platform DPT configuration templates. Prior to launching any of the following deployment modes, please ensure the following steps are performed:

- Copy the git repository to a folder (locally or on Google Cloud Shell).
- Create *aiml* folder or directory within *./healthcare/deploy*.
- Copy the DPT HIPAA-aligned analytics and AI/ML templates (variables.yaml and config.yaml) into *./healthcare/deploy/aiml.*

The DPT template file can be run in three different modes based on what is being done - validating the template, deploying it for the first time, or updating an existing deployment.

3.5.1 Dry Run Mode

Dry Run mode helps to review everything that the template might do from the beginning through the end of the deployment process without really deploying it.

This helps in validating the configuration file and ensures that it is well-formed. However, one caveat is that it may not detect any runtime errors that might happen during the deployment process.

```
bazel run cmd/apply:apply -- \
      --enable_terraform \
      --config_path=./healthcare/deploy/aiml/variables.yaml \
      --dry_run
```

3.5.2 Project Creation Mode

Project Creation mode helps to deploy the project once the configuration file is validated through the Dry Run mode.

```
bazel run cmd/apply:apply -- \
      --enable_terraform \
      --config_path=./healthcare/deploy/aiml/variables.yaml
```

3.5.3 Project Update Mode

Project Update mode helps to make modifications to an already existing project like adding a resource or updating a setting. The relevant project ID must be specified in the --*projects* option. If there are multiple projects, their project IDs must be separated by commas against the option.

```
bazel run cmd/apply:apply -- \
      --enable_terraform \
      --config_path=./healthcare/deploy/aiml/variables.yaml \
      --projects=[PROJECTS]
```

3.6 Project Types and Deployment Phases

DPT uses Terraform as the primary deployment tool.

- The deployment process kicks-off when the helper script *apply.go* reads the configurations from the configuration file and uses various other scripts from DPT to create Terraform scripts.
- The created Terraform scripts are then deployed in a phased manner.

3.6.1 Project Types

The projects deployed by DPT can be broadly classified into two categories - base projects and data-hosting projects. A DPT template can have multiple data-hosting projects but only one base project per category as discussed below.

- **Data-hosting project:** Any project defined under the *projects* list of the template is a data-hosting project. All the essential workloads and use-cases are developed as data-hosting projects. Eg. Data-warehouse workload, Analytics and AI/ML platforms, etc. Audit logging and resource tracking can either be defined separately for each data-hosting project or defined to be central for all the data-hosting projects as base projects.

- **Audit logs project(optional):** This is a base project to deploy central audit logging for all the data-hosting projects in the template. It is defined under the *audit_logs_project* section of the template. Once configured, it cannot be changed.

- **DevOps project(optional):** This is a base project for centralized state management of resources across all the data hosting projects in the template. It is defined under the *devops* section outside of the data-hosting projects. Once configured, it cannot be changed.

- **Forseti project(optional):** This is a base project to deploy Forseti for configuration monitoring of resources deployed across all the data-hosting projects. It is defined under the *forseti* section of the template.

3.6.2 Deployment Phases

As mentioned above, base projects such as central *DevOps*, *Audit,* and *Forseti* are optional and deployed before the projects which host data (i.e., data-hosting projects). The projects are deployed in phases and dependencies are addressed automatically.

The phases of deploying the template are mentioned as below:

1. **Project Creation:** In this phase, the project is created first, and then a storage bucket, which stores the state of the deployments related to the project.

   NOTE: If a top-level devops block is set in the config, all the state buckets will be created in the devops project.

2. **Resources:** This phase contains multiple deployments grouped as follows:
   a. ***Services*****:** This deployment consists of the default set of services required for the deployment of resources included in the project.
   b. ***Resources*****:** The deployment consists of specific preset default resources such as IAM permissions, logging metrics, and alert policies as defined resources under the project.

3. **Audit:** This phase contains a single deployment, which creates audit log resources defined in the audit block of a project (BigQuery Dataset and Cloud Storage bucket) as well as logging sinks to export audit logs.

   NOTE: If a top-level audit block is set in the config, these resources will be created in the central audit project.

4. **Forseti:** If the Forseti project config is also being applied, a Forseti instance is deployed in the Forseti project at this point and granted the minimum necessary access to each project to monitor security violations across these.

The following table explains the contents of the configuration file (config.yaml) and also details the deployment of data warehouse data-hosting project and its resources in phases.

Note: config.yaml is the base template which is imported into the variables.yaml template during runtime. While the config.yaml defines the deployment process and resources of the workload; parameters and configurations of the workload are declared in the variables.yaml file. Multiple variables.yaml files can be created to re-use config.yaml file deploy same workloads with different configurations.

3.7 Pre-Deployment setup

3.7.1 Initial Setup

● The 'billing_account' parameter sets up the billing account used by base projects and data-hosting projects deployed by this file. This can be used to restrict billing accounts used for projects restricting costs and controlling who can enable billing.
● The domain of the organization under which the projects are deployed must be mentioned against the 'domain' parameter.
	`overall:`
	`  billing_account: {{.BILLING_ACCOUNT}}`
	`  domain: {{.DOMAIN}}`
	`  organization_id: {{.ORGANIZATION_ID}}`
	`# folder_id: {{.FOLDER_ID}}`

3.7.2 Forseti Deployment

Forseti is a collection of open-source tools using rule-based policies to monitor and enforce configuration state on Google Cloud projects and resources. The following section of DPT installs Forseti Security and its core Forseti Security modules. The following modules can then be configured to take a snapshot of GCP resources, monitor and enforce configurations.

1. **Inventory** : Record a snapshot of GCP resources to Cloud SQLto maintain a historical record of resources in Google Cloud.
2. **Scanner**: Use the information collected by Forseti Inventory to regularly compare role-based access policies for your GCP resources.
3. **Enforcer**: Create and Use policies to compare the current state of Compute Engine firewall to the desired state.
4. **Explain**: Add-on module provides visibility into your Cloud Identity and Access Management (Cloud IAM) policies.
5. **Email Notifications**: Send email notifications for Inventory and Scanner using the SendGrid API. SendGrid is currently the only supported email provider.

● The 'forseti' section first creates a project which hosts Forseti services.	`generated_fields_path:`
	`./generated_fields.yaml`
	
	`# Forseti section deploys a forseti image`
	`which does security monitoring of GCP`

- The ‘devops’ section creates a storage bucket mentioned under DevOps sub-section.
- The ’terraform_deployments’ sub-section, first enables the Google Compute Engine API to create compute resources and then deploys a router and configures NAT.
- The Forseti Project can be used to monitor Google Cloud resources provide a single pane of view

Additional configuration options:

- Use Cloud IAM to restrict access to specific users and groups
- Set specific expiration time for data at a service level to control time period of retention
- Delete contents irretrievably if required through the template - to be used in specific use-cases only
- Enable versioning to retain older copies and serve as an audit trail
- Utilize NAT to provide access to internet
- Enable Private IP to restrict SQL access to internal network

```
resources
forseti:
  project:
    project_id: {{.FORSETI_PROJECT_ID}}
    owners_group:
{{.FORSETI_PROJECT_OWNERS_GROUP}}
    auditors_group:
{{.FORSETI_PROJECT_AUDITORS_GROUP}}        #
Auditors group at project level

    # Bigquery dataset stores audit logs from
the forseti project and its resources.
    audit:
      logs_bigquery_dataset:
        dataset_id:
{{.FORSETI_AUDIT_LOGS_BIGQUERY_DATASET_ID}}
        delete_contents_on_destroy:
{{.FORSETI_DELETE_CONTENTS_ON_DESTROY}}
        access:
        - user_by_email:
{{.FORSETIBQ_SERVICE_ACCOUNT_NAME}}@{{.FORSET
I_PROJECT_ID}}.iam.gserviceaccount.com
          role: roles/bigquery.dataEditor
# Can provide roles as per requirement.
        location: {{.LOCATION}}
        # default_encryption_configuration:
        #   kms_key_name:
(ex.{google_kms_crypto_key.gcs.self_link})
        labels:
          data_criticality: medium
          datatype: auditlogs
          project: {{.FORSETI_PROJECT_ID}}

    # Storage bucket stores the terraform
states of the resources in the projects.
    devops:
      state_storage_bucket:
        name:
{{.FORSETI_STATE_STORAGE_BUCKET}}
        # encryption:
        #   default_kms_key_name:
(ex.{google_kms_crypto_key.gcs.self_link})
        location: {{.LOCATION}}

    project_services:
    - service: compute.googleapis.com
    - service:
```

```
servicenetworking.googleapis.com

    # Setup NAT to allow private forseti to
access the internet to fetch the Forseti repo
while
    # having no external IP.
    # See
https://github.com/forseti-security/terraform
-google-forseti/issues/234.
    terraform_deployments:
      resources:
        config:
          resource:
          -
google_service_account_iam_binding:
              admin-account-iam:
                service_account_id:
${google_service_account.forsetibqsa.name}
                role:
roles/iam.serviceAccountUser
                members:
                -
group:{{.FORSETI_PROJECT_OWNERS_GROUP}}
          - google_service_account:
              forsetibqsa:
                account_id:
{{.FORSETIBQ_SERVICE_ACCOUNT_NAME}}
          # Setting up VPC
          - google_compute_network:
              forseti_private_network:
                name:
{{.FORSETI_VPC_NETWORK_NAME}}
                auto_create_subnetworks:
false
          - google_compute_subnetwork:
              forseti_subnetwork:
              - name:
{{.FORSETI_SUBNETWORK_NAME}}
                network:
${google_compute_network.forseti_private_netw
ork.self_link}
                region: {{.REGION}}
                ip_cidr_range:
{{.FORSETI_SUBNET_IP_RANGE}} # (ex.
192.168.0.0/20)
          - google_compute_router:
              forseti-router:
                name:
{{.FORSETI_ROUTER_NAME}}
                project:
{{.FORSETI_PROJECT_ID}}
                network:
${google_compute_network.forseti_private_netw
ork.self_link}
                region: {{.REGION}}
          - google_compute_router_nat:
              forseti-nat:
                name: {{.FORSETI_NAT_NAME}}
                project:
{{.FORSETI_PROJECT_ID}}
                region: {{.REGION}}
                nat_ip_allocate_option:
AUTO_ONLY

source_subnetwork_ip_ranges_to_nat:
ALL_SUBNETWORKS_ALL_IP_RANGES
                router:
${google_compute_router.forseti-router.name}
  properties:
    server_private: true
    client_private: true
    cloudsql_private: true
    network:
${google_compute_network.forseti_private_netw
ork.name}
    subnetwork:
${google_compute_subnetwork.forseti_subnetwor
k.name}
```

3.7.3 Data-Hosting Project Deployments

- This template includes the code for one data-hosting project.
  - It creates a new project if a project with the same name doesn't already exist in the organization.
  - In case of an existing project, it updates the project and its resources as per configuration.

```
  projects:
  - project_id: {{.AIML_PROJECT_ID}}
    owners_group: {{.AIML_OWNERS_GROUP}}
    auditors_group:
{{.AIML_AUDITORS_GROUP}}

    # Storage bucket stores the terraform
states of the resources in the projects.
    devops:
      state_storage_bucket:
        name:
```

- ○ If additional data-hosting projects are required, additional projects with their corresponding resources can be added under the projects list of the existing data-hosting template and the template can be rerun to create them.
- It also adds an ‘owner’ and ‘auditor’ group for access control on the project.

```
{{.AIML_STATE_STORAGE_BUCKET}}
    # encryption:
    #   default_kms_key_name:
(ex.{google_kms_crypto_key.gcs.self_link}
or
projects/my-project/locations/global/keyR
ings/my-ring/cryptoKeys/my-key)
    location: {{.LOCATION}}
    labels:
     data_criticality: low
     datatype: statefiles
     project: {{.AIML_PROJECT_ID}}

  # Bigquery dataset and data storage
bucket to store logs from projects and
their resources.
  audit:
   logs_bigquery_dataset:
    dataset_id:
{{.AIML_AUDIT_LOGS_BIGQUERY_DATASET_ID}}
    # delete_contents_on_destroy:
true
    access:
    - user_by_email:
{{.AUDITBQ_SERVICE_ACCOUNT_NAME}}@{{.AIML
_PROJECT_ID}}.iam.gserviceaccount.com
     role: roles/bigquery.dataEditor
# Can provide roles as per requirement.
    #
default_encryption_configuration:
    #   kms_key_name:
(ex.{google_kms_crypto_key.gcs.self_link}
or
projects/my-project/locations/global/keyR
ings/my-ring/cryptoKeys/my-key)
    location: {{.LOCATION}}
    labels:
     data_criticality: medium
     datatype: auditlogs
     project: {{.AIML_PROJECT_ID}}
   logs_storage_bucket:
    name:
{{.AIML_GCS_LOGS_STORAGE_BUCKET_NAME}}
    location: {{.LOCATION}}
    # encryption:
    #   default_kms_key_name:
(ex.{google_kms_crypto_key.gcs.self_link}
```

	or projects/my-project/locations/global/keyRings/my-ring/cryptoKeys/my-key)
	        lifecycle_rule:
	        - condition:
	            # Number of days from the time of creation, after which, storage objects from GCS logs bucket are moved to secondary storage class
	            age: {{.AIML_GCS_LOGS_AGE_FOR_SECONDARY_STORAGE_CLASS}}
	          action:
	            type: SetStorageClass
	            # The storage class to which, objects from GCS logs bucket will be pushed to, after the specified number of days mentioned above (e.g. STANDARD/REGIONAL/MULTI_REGIONAL/COLDLINE/NEARLINE)
	            storage_class: {{.AIML_GCS_LOGS_SECONDARY_STORAGE_CLASS}}
	        labels:
	          data_criticality: medium
	          datatype: gcslogs
	          project: {{.AIML_PROJECT_ID}}

● This section enables a list of APIs required by the data-hosting project.	# APIs required by AI/ML tasks and also other necessary APIs are enabled here project_services: - service: compute.googleapis.com - service: servicenetworking.googleapis.com - service: dataproc.googleapis.com - service: dlp.googleapis.com - service: speech.googleapis.com - service: lifesciences.googleapis.com - service: translate.googleapis.com - service: videointelligence.googleapis.com - service: vision.googleapis.com - service: language.googleapis.com
● In this section, two (2) BigQuery datasets are deployed under the same project.	# Deploying 2 BigQuery datasets for raw and transformed data bigquery_datasets:

  - One dataset is used for storing unprocessed healthcare data loaded from the Cloud SQL instance, Cloud Storage bucket or healthcare datasets.
  - The other BigQuery dataset stores the transformed/ enriched data that comes from various data transformation tools or tasks.
- The BigQuery table allows the erasure of contents on deletion to protect against unauthorized or accidental retrieval.
- The template also accommodates provisioning of role-based access to individual resources through Google Cloud IAM.

Note: The configurations are customizable and can be changed as required to meet specific use-cases.

```
  - dataset_id:
{{.RAW_DATA_BIGQUERY_DATASET_ID}}
    # delete_contents_on_destroy: true
    depends_on:
      -
google_service_account.{{.RAWDATABQ_SERVICE_ACCOUNT_NAME}}
    access:
    - user_by_email:
${google_service_account.{{.RAWDATABQ_SERVICE_ACCOUNT_NAME}}.email}
      role: roles/bigquery.dataEditor
# Can provide roles as per requirement.
    - special_group:
{{.RAW_DATA_BQ_SPECIAL_GROUP}}
      role:
{{.RAW_DATA_BQ_SPECIAL_GROUP_ROLE}}
    # default_encryption_configuration:
    #   kms_key_name:
(ex.{google_kms_crypto_key.gcs.self_link}
or
projects/my-project/locations/global/keyRings/my-ring/cryptoKeys/my-key)
    location: {{.LOCATION}}
    labels:
      data_criticality:
{{.RAW_DATA_BIGQUERY_DATASET_DATA_CRITICALITY_LABEL}}
      datatype:
{{.RAW_DATA_BIGQUERY_DATASET_DATA_TYPE_LABEL}}
      project: {{.AIML_PROJECT_ID}}

  - dataset_id:
{{.TRANSFORMED_DATA_BIGQUERY_DATASET_ID}}
    delete_contents_on_destroy: true
    depends_on:
      -
google_service_account.{{.TRANSDATABQ_SERVICE_ACCOUNT_NAME}}
    access:
    - user_by_email:
${google_service_account.{{.TRANSDATABQ_SERVICE_ACCOUNT_NAME}}.email}
      role: roles/bigquery.dataEditor
# Can provide roles as per requirement.
    - special_group:
```

	{{.TRANSFORMED_DATA_BQ_SPECIAL_GROUP}}
	      role: {{.TRANSFORMED_DATA_BQ_SPECIAL_GROUP_ROLE}}
	    # default_encryption_configuration:
	    # kms_key_name: (ex.{google_kms_crypto_key.gcs.self_link} or projects/my-project/locations/global/keyRings/my-ring/cryptoKeys/my-key)
	    location: {{.LOCATION}}
	    labels:
	      data_criticality: {{.TRANSFORMED_DATA_BIGQUERY_DATASET_DATA_CRITICALITY_LABEL}}
	      datatype: {{.TRANSFORMED_DATA_BIGQUERY_DATASET_DATA_TYPE_LABEL}}
	      project: {{.AIML_PROJECT_ID}}

● In this section, a Cloud Storage bucket is created under the same project for storing unprocessed data generated by various healthcare tools.
● The retention period of the Cloud Storage bucket can be changed as required. For example, set the trigger for deletion of ePHI data once processed and transferred out of storage.
	# Deploying storage bucket as a data source
	  storage_buckets:
	  - name: {{.RAW_DATA_STORAGE_BUCKET_NAME}}
	    # IAM Role Binding
	    _iam_members:
	    - role: roles/storage.objectCreator
	      member: {{.RAW_DATA_STORAGE_BUCKET_OBJECTCREATOR}}
	    - role: roles/storage.objectViewer
	      member: {{.RAW_DATA_STORAGE_BUCKET_OBJECTVIEWER}}
	    # encryption:
	    # default_kms_key_name: (ex.{google_kms_crypto_key.gcs.self_link})
	    location: {{.LOCATION}}
	    labels:
	      data_criticality: {{.RAW_DATA_STORAGE_BUCKET_DATA_CRITICALITY_LABEL}}
	      datatype: {{.RAW_DATA_STORAGE_BUCKET_DATA_TYPE_LABEL}}
	      project: {{.AIML_PROJECT_ID}}

- In this section, healthcare data stores named DICOM store, FHIR store, and hl7-v2-store are created under a new healthcare dataset.
- While these datasets are provisioned, the Cloud Healthcare API will also be enabled by the template.
- Along with Google Cloud IAM, access to the API can be restricted to a limited set of users based on roles, responsibilities, and dataset.

```
# Datastores for personal data which aid
in features of Healthcare API
  healthcare_datasets:
  - name: {{.HEALTHCARE_DATASET_NAME}}
    location: {{.REGION}}
    # IAM Role Binding
    _iam_members:
    - role:
roles/healthcare.datasetViewer
      member:
{{.HEALTHCARE_DATASET_VIEWER}}
    _dicom_stores:
    - name:
{{.HEALTHCARE_DICOM_STORE_NAME}}
      _iam_members:
      - role:
roles/healthcare.dicomEditor
        member:
{{.HEALTHCARE_DICOM_EDITOR}}
      - role:
roles/healthcare.dicomStoreAdmin
        member:
{{.HEALTHCARE_DICOM_STOREADMIN}}
    _fhir_stores:
    - name:
{{.HEALTHCARE_FHIR_STORE_NAME}}
      _iam_members:
      - role:
roles/healthcare.fhirResourceReader
        member:
{{.HEALTHCARE_FHIR_STORE_READER}}
      - role:
roles/healthcare.fhirResourceEditor
        member:
{{.HEALTHCARE_FHIR_STORE_EDITOR}}
    _hl7_v2_stores:
    - name:
{{.HEALTHCARE_HL7V2_STORE_NAME}}
      _iam_members:
      - role:
roles/healthcare.hl7V2StoreAdmin
        member:
{{.HEALTHCARE_HL7V2_STOREADMIN}}
      - role:
roles/healthcare.hl7V2Ingest
        member:
{{.HEALTHCARE_HL7V2_INGEST}}
```

```
      - role:
roles/healthcare.hl7V2Editor
        member:
{{.HEALTHCARE_HL7V2_STORE_EDITOR}}
```

- This section creates a Private IP SQL instance and its replica which stores the unprocessed health data.
- These instances are attached to a private VPC network which effectively restricts access to them and also protects them against exposure to public networks. This can be further customized to build micro-segments within the network.

```
terraform_deployments:
  resources:
    config:
      resource:
      - google_compute_global_address:
          private_ip_addresses:
            name:
{{.CLOUD_SQL_PRIVATE_IP_NAME}}
            purpose: VPC_PEERING
            address_type: INTERNAL
            prefix_length: 16
            network:
${google_compute_network.private_network.
self_link}
      -
google_service_networking_connection:
          private_vpc_connection:
            network:
${google_compute_network.private_network.
self_link}
            service:
servicenetworking.googleapis.com
            reserved_peering_ranges:
            -
${google_compute_global_address.private_i
p_addresses.name}
        # Setting up master and replica
SQL instances
      - google_sql_database_instance:
          instance:
            name:
{{.MASTER_CLOUD_SQL_NAME}}
            region: {{.REGION}}
            depends_on:
            -
google_service_networking_connection.priv
ate_vpc_connection
            settings:
              availability_type: ZONAL
              tier: db-f1-micro
              ip_configuration:
                ipv4_enabled: false
                private_network:
${google_compute_network.private_network.
self_link}
              backup_configuration:
                binary_log_enabled: true
                enabled: true
          replica-instance:
            name:
{{.REPLICA_CLOUD_SQL_NAME}}
            region: {{.REGION}}
            master_instance_name:
${google_sql_database_instance.instance.n
ame}
            depends_on:
            -
google_service_networking_connection.priv
ate_vpc_connection
            settings:
              availability_type: ZONAL
              tier: db-f1-micro
              ip_configuration:
                ipv4_enabled: false
                private_network:
${google_compute_network.private_network.
self_link}
            replica_configuration:
              failover_target: true
              master_heartbeat_period:
60000
```

- This section deploys a Datalab module, built on Jupyter, which enables analysis of unprocessed data on BigQuery datasets, storage buckets, and cloud SQL instances.
- This Datalab module is attached to a private VPC network through a dedicated subnet which effectively restricts access to the module and protects it against exposure to public networks.
- This module is also assigned a service account to either restrict or allow its access for other resources.

```
terraform_deployments:
  resources:
    config:
      module:
      - datalab:
        - datalab_user_email:
{{.DATALAB_EMAIL}}
        # adding gpu requires a quota
assigned. Refer to link
"https://github.com/terraform-google-modu
les/terraform-google-datalab/tree/master/
examples/basic" for furthur information.
          network_name:
${google_compute_network.private_network.
name}
          project_id:
{{.AIML_PROJECT_ID}}
```

```
          source:
terraform-google-modules/datalab/google//
modules/instance
          subnet_name:
${google_compute_subnetwork.datalab-subne
twork.name}
          version: "~> 1.0"
          zone: {{.ZONE}}
        # custom service account with the
deployer given serviceAccountUser role
          service_account:
${google_service_account.{{.DATALAB_SERVI
CE_ACCOUNT_NAME}}.email}
          datalab_enable_swap: true
          datalab_enable_backup: true
          datalab_console_log_level:
"warn"
          datalab_idle_timeout: "60m"
          create_disk: true
```

- This section deploys a Deep Learning VM instance, with pre-installed Jupyter notebooks and machine learning libraries, which enables analysis of unprocessed data on BigQuery datasets, storage buckets, and cloud SQL instances.
- This Deep Learning VM instance is attached to a private VPC network through a dedicated subnet which effectively restricts access to the instance and protects it against exposure to public networks.
- This instance is also provided with a disk attached to it whose back-up is taken by a snapshot.
- This instance is also assigned a service account to either restrict or allow its access for other resources.

```
terraform_deployments:
  resources:
    config:
      resource:
      - google_compute_snapshot:
          deep-learning-vm-snapshot:
            name:
deep-learning-vm-snapshot
            source_disk:
${google_compute_disk.deep-learning-vm-di
sk.name}
            zone: {{.ZONE}}
            labels:
              project:
{{.AIML_PROJECT_ID}}
              connected_disk:
${google_compute_disk.deep-learning-vm-di
sk.name}
          # Customer Managed Keys must be
configured here
            # snapshot_encryption_key:
            #   raw_key:
            #   sha256:
      - google_compute_disk:
          deep-learning-vm-disk:
            name: deep-learning-vm-disk
            type: pd-ssd
            zone: {{.ZONE}}
            labels:
              project:
{{.AIML_PROJECT_ID}}
              connected_instance:
{{.DEEP_LEARNING_VM_NAME}}
            physical_block_size_bytes:
4096
          # Customer Managed Keys must be
configured here
            # disk_encryption_key:
            #   raw_key:
            #   sha256:
            #   kms_key_self_link:
(ex.{google_kms_crypto_key.gcs.self_link}
)
# Deep Learning VMs
compute_instances:
- name: {{.DEEP_LEARNING_VM_NAME}}
  labels:
    data_criticality:
{{.DEEP_LEARNING_VM_DATA_CRITICALITY_LABE
L}}
    datatype:
{{.DEEP_LEARNING_VM_DATA_TYPE_LABEL}}
    project: {{.AIML_PROJECT_ID}}
  zone: {{.ZONE}}
  machine_type: n1-standard-1
  attached_disk:
  - source:
${google_compute_disk.deep-learning-vm-di
sk.self_link}
    mode: READ_ONLY
  # A 256-bit customer supported key
stored with them.
    # disk_encryption_key_raw:
  # A key stored on Google Cloud KMS.
    # kms_key_self_link:
(ex.{google_kms_crypto_key.gcs.self_link}
or
projects/my-project/locations/global/keyR
ings/my-ring/cryptoKeys/my-key)
  boot_disk:
    auto_delete: false
    mode: READ_WRITE           # can be
changed to READ_WRITE or READ_ONLY
    initialize_params:
```

```
    # deep learning vm containing
tensorflow pre-installed as example. Use
"gcloud compute images list --project
deeplearning-platform-release
--no-standard-images" to find a list of
more VMs users can use.
    # using VMs with GPU requires
specific configurations. Check the deep
learning VM Google documentation for the
same.
      image:
https://www.googleapis.com/compute/v1/pro
jects/deeplearning-platform-release/globa
l/images/tf2-latest-cpu-20200227
  # A 256-bit customer supported key
stored with them.
    # disk_encryption_key_raw:
  # A key stored on Google Cloud KMS.
    # kms_key_self_link:
(ex.{google_kms_crypto_key.gcs.self_link}
or
projects/my-project/locations/global/keyR
ings/my-ring/cryptoKeys/my-key)
  network_interface:
    subnetwork:
${google_compute_subnetwork.deep-learning
-vm-subnetwork.self_link}
  service_account:
    email:
${google_service_account.{{.DEEPVM_SERVIC
E_ACCOUNT_NAME}}.email}
    scopes:
    - bigquery
    - sql-admin
    - userinfo-email
    - compute-ro
    - storage-ro
  _iam_members:
  - role: roles/editor
    member:
{{.DEEP_LEARNING_VM_EDITOR_ROLE_MEMBER}}
  - role: roles/viewer
    member:
{{.DEEP_LEARNING_VM_VIEWER_ROLE_MEMBER}}
  # allow_stopping_for_update: true /
false          #  Enables an instance to
be stopped for an updation purpose)
```

```
  deletion_protection: false
  # shielded_instance_config:
#  Shielded VM Config can only be set
when using a UEFI-compatible disk
  # enable_secure_boot: true
```

- This section deploys a Dataproc cluster, which processes Hadoop jobs like Spark, Hive and Pig on unprocessed data from BigQuery datasets, Storage buckets, and Cloud SQL instances.
- This cluster is assigned a service account to either restrict or allow its access for other resources.
- This section also highlights the syntax to define a Dataproc job.

```
terraform_deployments:
  resources:
    config:
      resource:
      # Dataproc cluster
      - google_dataproc_cluster:
          mycluster:
            name:
{{.DATAPROC_CLUSTER_NAME}}
            region: {{.REGION}}
            labels:
              project:
{{.AIML_PROJECT_ID}}
            cluster_config:
              master_config:
                num_instances: 1
                machine_type:
n1-standard-1
                disk_config:
                  boot_disk_type: pd-ssd
                  boot_disk_size_gb: 15
              worker_config:
                num_instances: 2
                machine_type:
n1-standard-1
                min_cpu_platform: Intel
Skylake
                disk_config:
                  boot_disk_size_gb: 15
                  num_local_ssds: 1
              preemptible_worker_config:
                num_instances: 0
              software_config:
                image_version: 1.3.7-deb9
                override_properties:

dataproc:dataproc.allow.zero.workers:
'true'
      # KMS keys configuration can be
enabled. Refer
"https://www.terraform.io/docs/providers/
```

```
google/r/dataproc_cluster.html"
              #  security_config:
              #    kerberos_config:
              #      kms_key_uri:
(ex.{google_kms_crypto_key.gcs.self_link}
or
projects/my-project/locations/global/keyR
ings/my-ring/cryptoKeys/my-key)
              #
root_principal_password_uri:
bucketId/o/objectId
              autoscaling_config:
                policy_uri:
${google_dataproc_autoscaling_policy.asp.
name}
              gce_cluster_config:
                service_account:
{{.DATAPROC_SERVICE_ACCOUNT_NAME}}@{{.AIM
L_PROJECT_ID}}.iam.gserviceaccount.com
      -
google_dataproc_autoscaling_policy:
          asp:
            policy_id: dataprocs-policys
            location: {{.REGION}}
            worker_config:
              min_instances: 2
              max_instances: 5
            basic_algorithm:
              yarn_config:

graceful_decommission_timeout: 30s
                scale_up_factor: 0.5
                scale_down_factor: 0.5
    # IAM  for dataprocs
      -
google_dataproc_cluster_iam_member:
        - editor:
          - cluster:
${google_dataproc_cluster.mycluster.name}
            member:
{{.DATAPROC_CLUSTER_EDITOR_ROLE_MEMBER}}
            role: roles/editor
            region: {{.REGION}}
        - viewer:
          - cluster:
${google_dataproc_cluster.mycluster.name}
            member:
```

```
{{.DATAPROC_CLUSTER_VIEWER_ROLE_MEMBER}}
      role: roles/viewer
      region: {{.REGION}}
   - google_dataproc_job_iam_member:
    - editor:
     - job_id:
${google_dataproc_job.pyspark.reference[0
].job_id}
      member:
{{.DATAPROC_JOB_EDITOR_ROLE_MEMBER}}
      role: roles/editor
      region: {{.REGION}}
    - viewer:
     - job_id:
${google_dataproc_job.pyspark.reference[0
].job_id}
      member:
{{.DATAPROC_JOB_VIEWER_ROLE_MEMBER}}
      role: roles/viewer
      region: {{.REGION}}
   # Example pyspark dataproc job.
Refer
"https://www.terraform.io/docs/providers/
google/r/dataproc_job.html" for more
examples.
   - google_dataproc_job:
     pyspark:
      region:
${google_dataproc_cluster.mycluster.regio
n}
      placement:
       cluster_name:
${google_dataproc_cluster.mycluster.name}
      pyspark_config:
       main_python_file_uri:
gs://dataproc-examples-2f10d78d114f6aaec7
6462e3c310f31f/src/pyspark/hello-world/he
llo-world.py
       properties:
        spark.logConf: 'true'
  # Setting up VPC network for Datalab,
Cloud SQL and Deep Learning VM
   - google_compute_network:
     private_network:
      name:
{{.PRIVATE_VPC_NETWORK_NAME}}
      auto_create_subnetworks:
```

```
false
   - google_compute_global_address:
     private_ip_addresses:
      name:
{{.CLOUD_SQL_PRIVATE_IP_NAME}}
      purpose: VPC_PEERING
      address_type: INTERNAL
      prefix_length: 16
      network:
${google_compute_network.private_network.
self_link}
   -
google_service_networking_connection:
     private_vpc_connection:
      network:
${google_compute_network.private_network.
self_link}
      service:
servicenetworking.googleapis.com
      reserved_peering_ranges:
      -
${google_compute_global_address.private_i
p_addresses.name}
```

- This section creates services for many resources deployed by the template such as Dataproc cluster, Deep Learning VM instance and Datalab module.
- These service accounts are referred to in their respective resource sections of the template.
- These services accounts effectively restrict or allow the resources access for other resources and thus, define tight scopes.

```
# Custom created service account
service_accounts:
- account_id:
{{.DATAPROC_SERVICE_ACCOUNT_NAME}}
- account_id:
{{.DATALAB_SERVICE_ACCOUNT_NAME}}
- account_id:
{{.DEEPVM_SERVICE_ACCOUNT_NAME}}
- account_id:
{{.AUDITBQ_SERVICE_ACCOUNT_NAME}}
- account_id:
{{.RAWDATABQ_SERVICE_ACCOUNT_NAME}}
- account_id:
{{.TRANSDATABQ_SERVICE_ACCOUNT_NAME}}
```

- This section deploys a private VPC network and its subnets which are used by resources like Cloud SQL instances, Datalab modules and Deep Learning VM instances.

```
terraform_deployments:
 resources:
  config:
   resource:
   - google_compute_network:
     private_network:
      name:
```

- Two separate subnets are created under that private network, one for Datalab and the other for Deep Learning VM.	{{.PRIVATE_VPC_NETWORK_NAME}}
	            auto_create_subnetworks:
	false
	      - google_compute_subnetwork:
	        - datalab-subnetwork:
	          - name: datalab-subnet
	            network:
	${google_compute_network.private_network.
	self_link}
	            region: {{.REGION}}
	            ip_cidr_range: 10.2.0.0/16
	        - deep-learning-vm-subnetwork:
	          - name: deep-learning-vm-subnet
	            network:
	${google_compute_network.private_network.
	self_link}
	            region: {{.REGION}}
	            ip_cidr_range: 10.3.0.0/16

Note: The deployment template can have the deployment variables statically declared or passed during execution. The template above implements the latter and integrates with another template which contains all the variable declarations required during the runtime. This helps in separating the code from the variables, enabling re-usability, modularity, and security. To learn more about modularization and reusability of DPT templates, refer here.

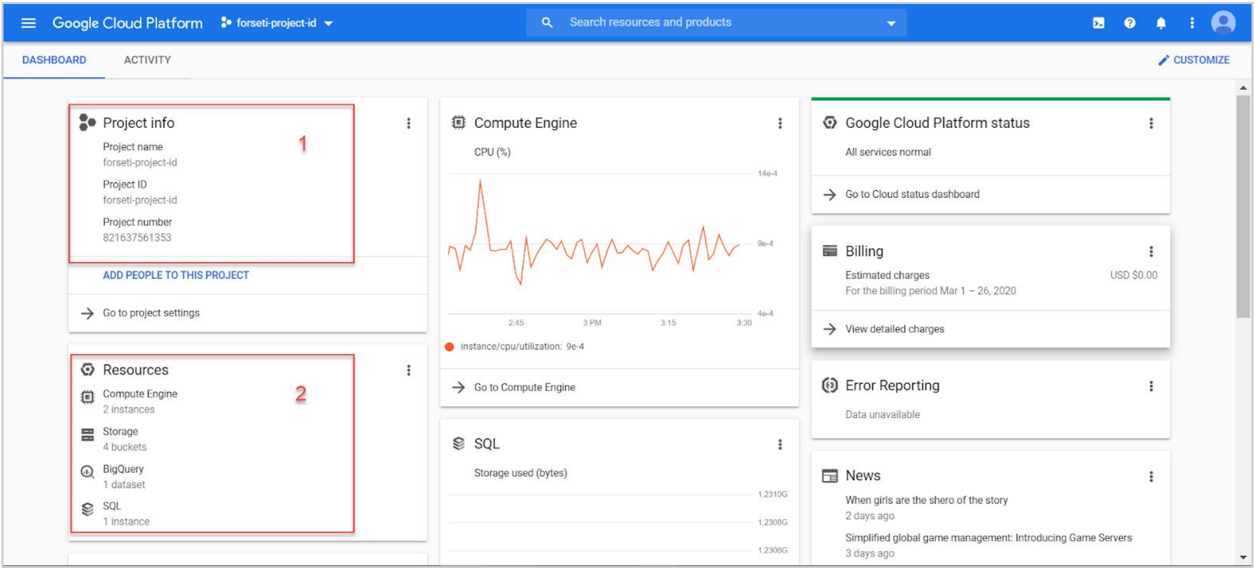
3.8 Post-Deployment Verification

Post-deployment of the DPT template, two projects are created:

- Primary data project (“dpth-ai-ml” in this example)
- A separate Forseti project for monitoring (“forseti-project-id” in this example)

3.8.1 Forseti Project

Cloud Console Dashboard: Using the Cloud Console, a list of projects that are deployed can be viewed and selected. Under the Forseti project (forseti-project-id), the project information and the resources that are deployed using DPT are listed as highlighted in Figure 2.



Project info

Project name
forseti-project-id

Project ID
forseti-project-id

Project number
821637561353

ADD PEOPLE TO THIS PROJECT

→ Go to project settings

Resources

Compute Engine
2 instances

Storage
4 buckets

BigQuery
1 dataset

SQL
1 instance

Compute Engine

CPU (%)

● instance/cpu/utilization: 9e-4

→ Go to Compute Engine

SQL

Storage used (bytes)

Google Cloud Platform status

All services normal

→ Go to Cloud status dashboard

Billing

Estimated charges USD $0.00
For the billing period Mar 1 – 26, 2020

→ View detailed charges

Error Reporting

Data unavailable

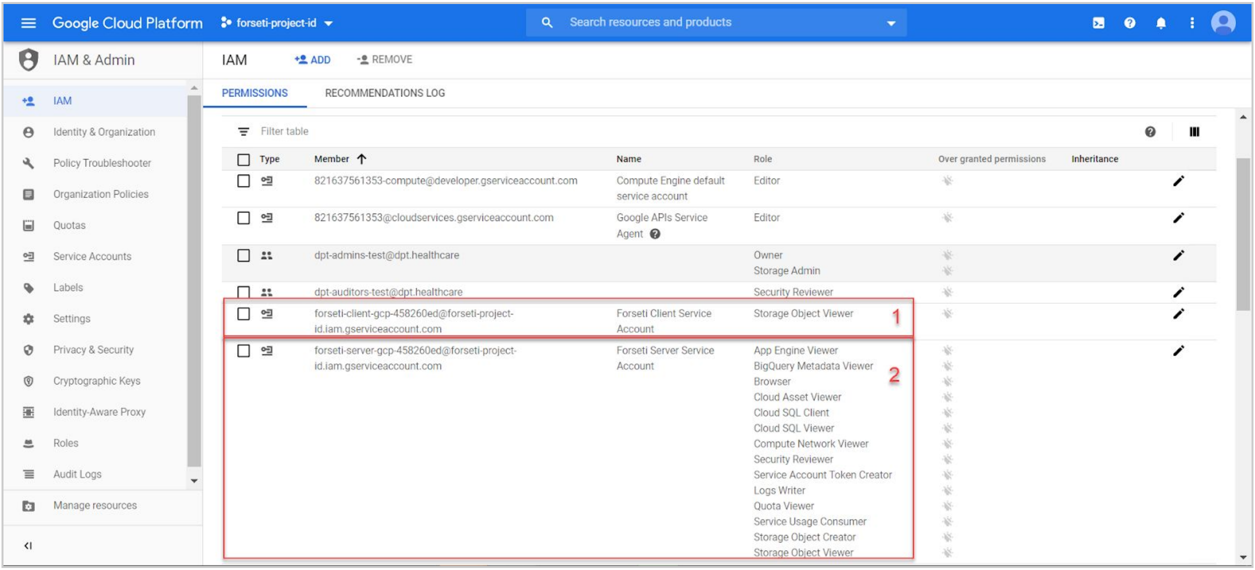
News

When girls are the shero of the story
2 days ago

Simplified global game management: Introducing Game Servers
3 days ago

Figure 2 - 1. Forseti Project Info   2. Forseti Project Resources

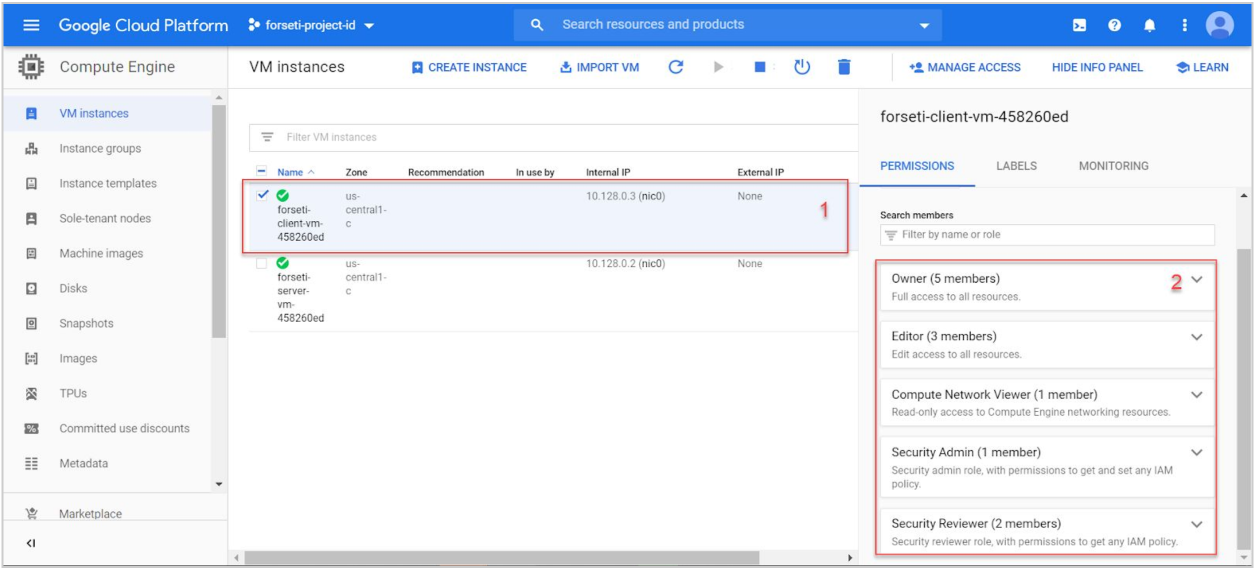
IAM Console: The IAM permissions can be viewed through the IAM console (Figure 3). Here, the Forseti service accounts are created, and the appropriate roles are assigned as defined in the template. The Storage Object Viewer role is provided to one of the service accounts, and the AppEngine Viewer role is provided to the other account.



☐ Type	Member ↑	Name	Role	Over granted permissions	Inheritance
☐	821637561353-compute@developer.gserviceaccount.com	Compute Engine default service account	Editor		
☐	821637561353@cloudservices.gserviceaccount.com	Google APIs Service Agent	Editor		
☐	dpt-admins-test@dpt.healthcare		Owner Storage Admin		
☐	dpt-auditors-test@dpt.healthcare		Security Reviewer		
☐	forseti-client-gcp-458260ed@forseti-project-id.iam.gserviceaccount.com	Forseti Client Service Account	Storage Object Viewer		
☐	forseti-server-gcp-458260ed@forseti-project-id.iam.gserviceaccount.com	Forseti Server Service Account	App Engine Viewer BigQuery Metadata Viewer Browser Cloud Asset Viewer Cloud SQL Client Cloud SQL Viewer Compute Network Viewer Security Reviewer Service Account Token Creator Logs Writer Quota Viewer Service Usage Consumer Storage Object Creator Storage Object Viewer		

Figure 3 - 1. Forseti Client Service Account   2. Forseti Server Service Account

Compute Engine Console: View the list of servers created on the 'VM instances' console. Two Forseti Virtual Machine (VM) instances are created as a part of the Forseti project. The list of permissions on the VM's, as highlighted in Figure 4 and 5, can be viewed by selecting the VM.



Name	Zone	Recommendation	In use by	Internal IP	External IP
✓ forseti-client-vm-458260ed	us-central1-c			10.128.0.3 (nic0)	None
forseti-server-vm-458260ed	us-central1-c			10.128.0.2 (nic0)	None

forseti-client-vm-458260ed

Owner (5 members)
Full access to all resources.

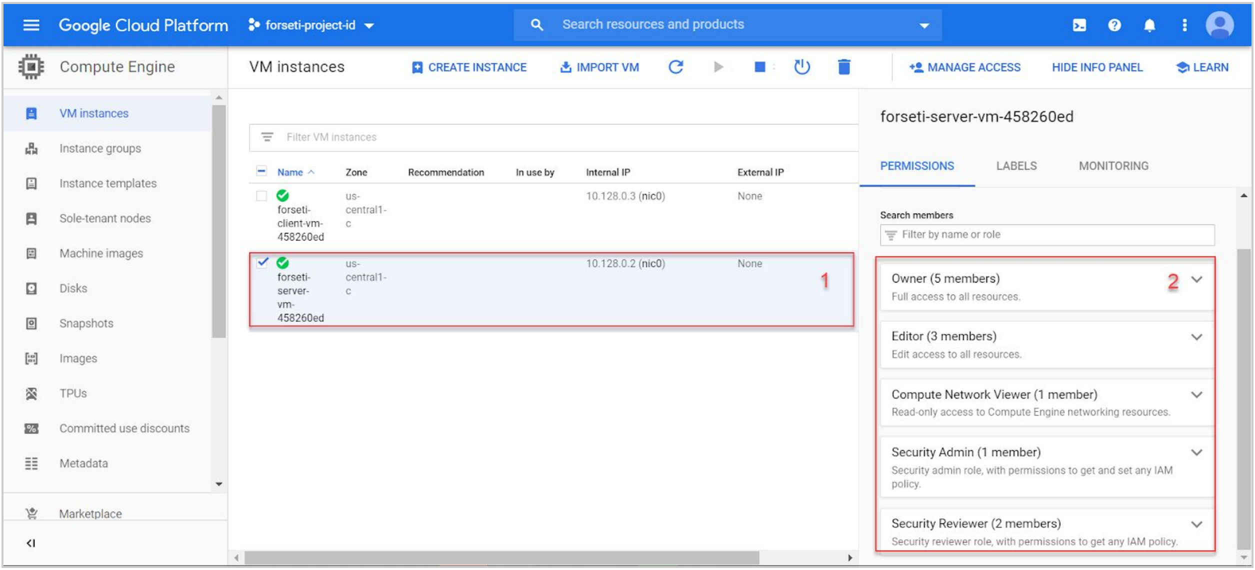
Editor (3 members)
Edit access to all resources.

Compute Network Viewer (1 member)
Read-only access to Compute Engine networking resources.

Security Admin (1 member)
Security admin role, with permissions to get and set any IAM policy.

Security Reviewer (2 members)
Security reviewer role, with permissions to get any IAM policy.

Figure 4 - 1. Forseti Client Instance 2. Instance Permissions



Name	Zone	Recommendation	In use by	Internal IP	External IP
forseti-client-vm-458260ed	us-central1-c			10.128.0.3 (nic0)	None
✓ forseti-server-vm-458260ed	us-central1-c			10.128.0.2 (nic0)	None

forseti-server-vm-458260ed

Owner (5 members)
Full access to all resources.

Editor (3 members)
Edit access to all resources.

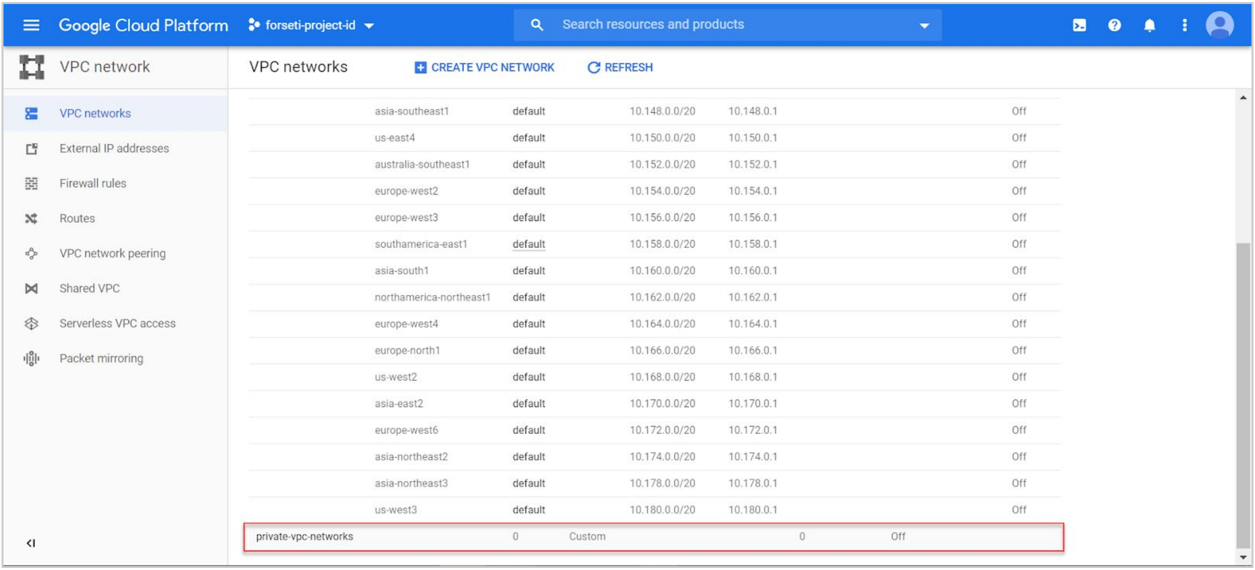
Compute Network Viewer (1 member)
Read-only access to Compute Engine networking resources.

Security Admin (1 member)
Security admin role, with permissions to get and set any IAM policy.

Security Reviewer (2 members)
Security reviewer role, with permissions to get any IAM policy.

Figure 5 - 1. Forseti Server Instance 2. Instance Permissions

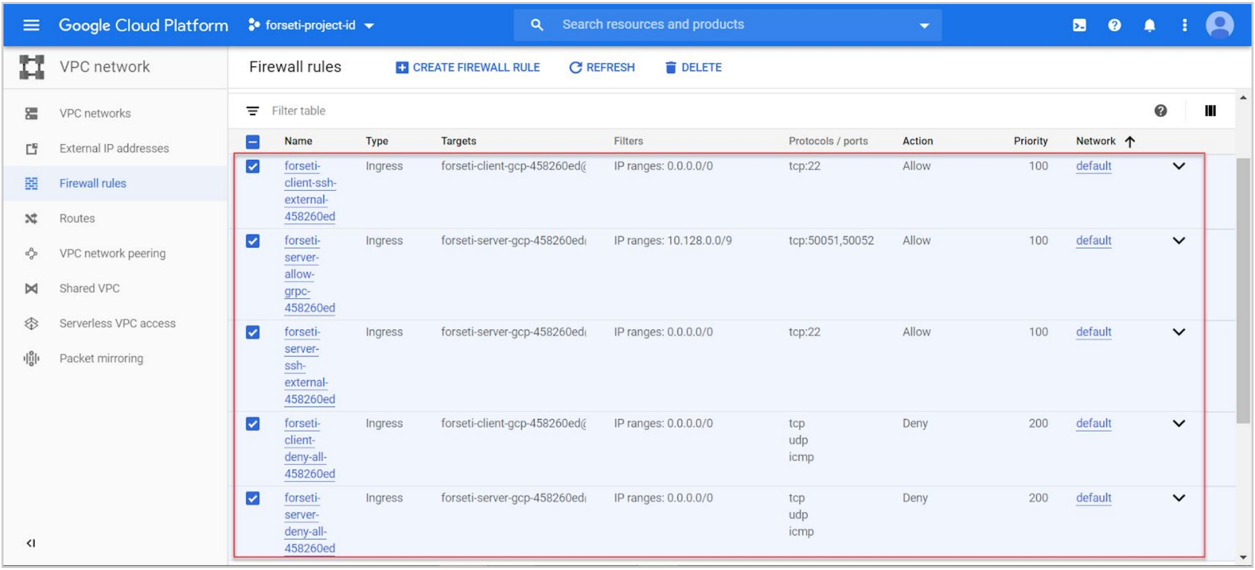
VPC Network Console: Figure 6 shows the Virtual Private Network created for Forseti on the [VPC network Console](#).



	asia-southeast1	default		10.148.0.0/20	10.148.0.1			Off
	us-east4	default		10.150.0.0/20	10.150.0.1			Off
	australia-southeast1	default		10.152.0.0/20	10.152.0.1			Off
	europe-west2	default		10.154.0.0/20	10.154.0.1			Off
	europe-west3	default		10.156.0.0/20	10.156.0.1			Off
	southamerica-east1	default		10.158.0.0/20	10.158.0.1			Off
	asia-south1	default		10.160.0.0/20	10.160.0.1			Off
	northamerica-northeast1	default		10.162.0.0/20	10.162.0.1			Off
	europe-west4	default		10.164.0.0/20	10.164.0.1			Off
	europe-north1	default		10.166.0.0/20	10.166.0.1			Off
	us-west2	default		10.168.0.0/20	10.168.0.1			Off
	asia-east2	default		10.170.0.0/20	10.170.0.1			Off
	europe-west6	default		10.172.0.0/20	10.172.0.1			Off
	asia-northeast2	default		10.174.0.0/20	10.174.0.1			Off
	asia-northeast3	default		10.178.0.0/20	10.178.0.1			Off
	us-west3	default		10.180.0.0/20	10.180.0.1			Off
private-vpc-networks		0	Custom			0	Off	

Figure 6 - Private VPC

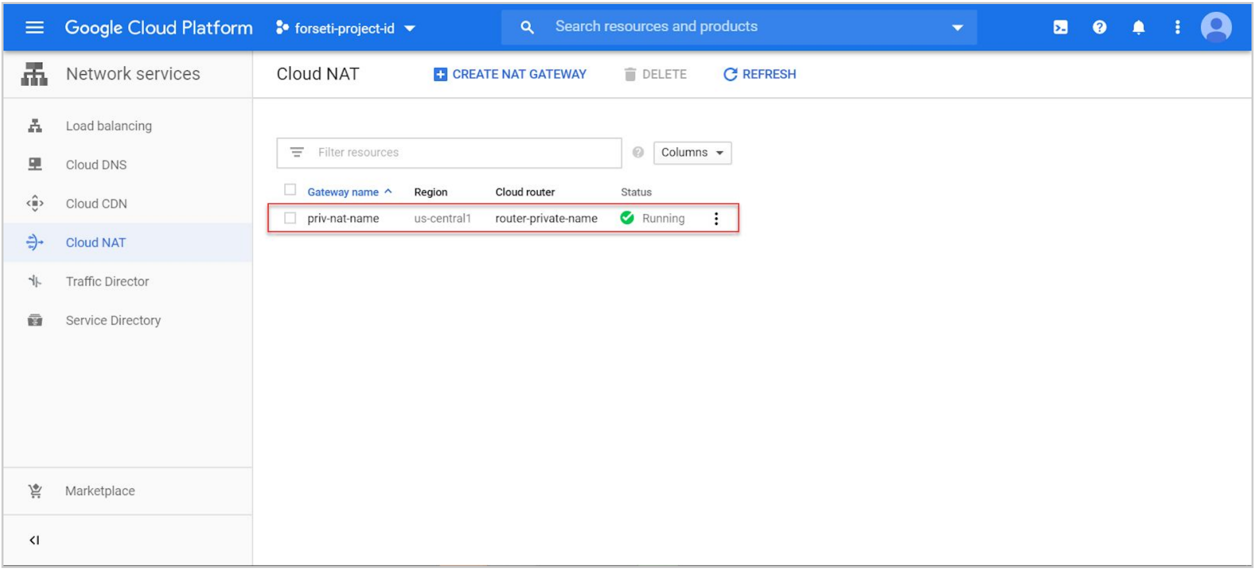
The firewall rules section on the [VPC network Console](#) displays firewall rules created as a part of this project (Figure 7). The ‘Firewall rules’ console displays the Rule name, rule type selected (Ingress/Egress), IP ranges, and the Instance to which the firewall rules are applied.



☐	Name	Type	Targets	Filters	Protocols / ports	Action	Priority	Network ↑
☑	forseti-client-ssh-external-458260ed	Ingress	forseti-client-gcp-458260ed(	IP ranges: 0.0.0.0/0	tcp:22	Allow	100	default
☑	forseti-server-allow-grpc-458260ed	Ingress	forseti-server-gcp-458260ed(	IP ranges: 10.128.0.0/9	tcp:50051,50052	Allow	100	default
☑	forseti-server-ssh-external-458260ed	Ingress	forseti-server-gcp-458260ed(	IP ranges: 0.0.0.0/0	tcp:22	Allow	100	default
☑	forseti-client-deny-all-458260ed	Ingress	forseti-client-gcp-458260ed(	IP ranges: 0.0.0.0/0	tcp udp icmp	Deny	200	default
☑	forseti-server-deny-all-458260ed	Ingress	forseti-server-gcp-458260ed(	IP ranges: 0.0.0.0/0	tcp udp icmp	Deny	200	default

Figure 7 - Forseti Firewall Rules

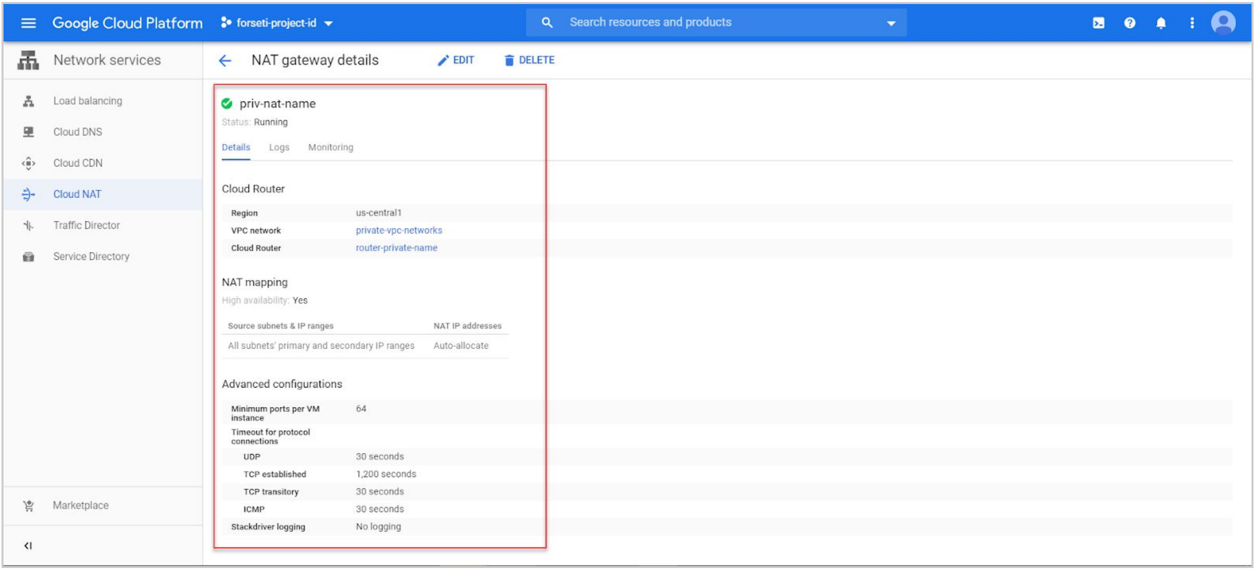
Network Services Console: The Cloud NAT created under the Forseti project is deployed on the [Network Services Console](#) (Figure 8).



Gateway name	Region	Cloud router	Status
priv-nat-name	us-central1	router-private-name	Running

Figure 8 - Network Services Console

Hybrid Connectivity Console: The details of the NAT gateway are available in the [Network Services Console](#) (Figure 9). It provides the details about the subnets selected and the routes configured on the NAT gateway.



priv-nat-name

Status: Running

Cloud Router

Region	us-central1
VPC network	private-vpc-networks
Cloud Router	router-private-name

NAT mapping

High availability: Yes

Source subnets & IP ranges	NAT IP addresses
All subnets' primary and secondary IP ranges	Auto-allocate

Advanced configurations

Minimum ports per VM instance	64
Timeout for protocol connections	
UDP	30 seconds
TCP established	1,200 seconds
TCP transitory	30 seconds
ICMP	30 seconds
Stackdriver logging	No logging

Figure 9 - Hybrid Connectivity Console

BigQuery Console: A BigQuery dataset has been created for analyzing the logs generated by resources, as can be seen in the [BigQuery Console](#) under the Forseti project (Figure 10).

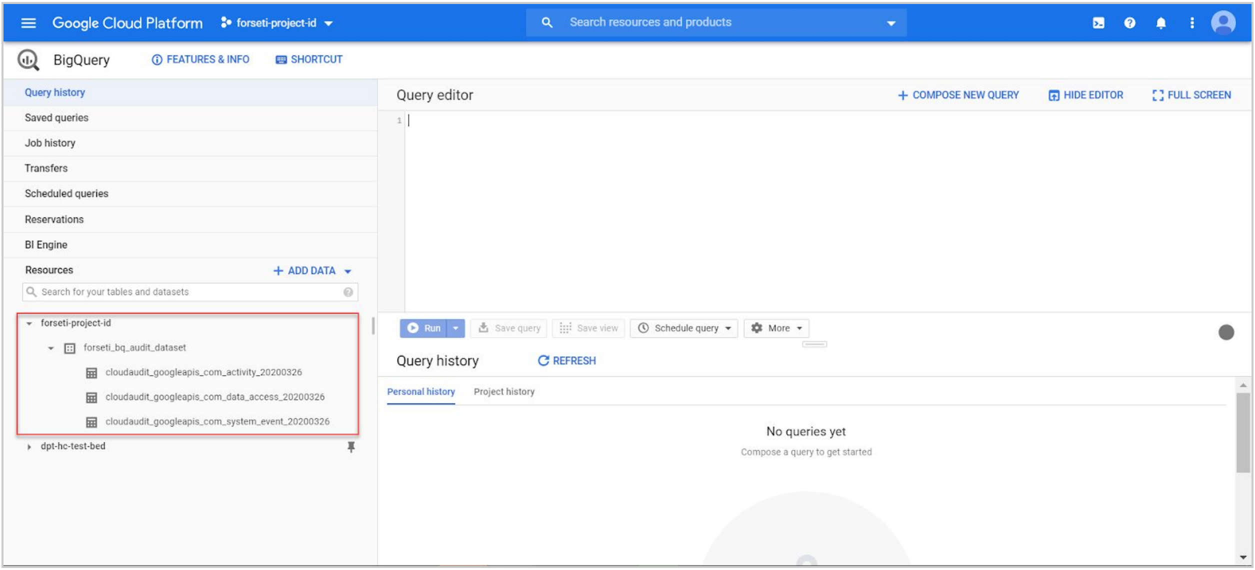
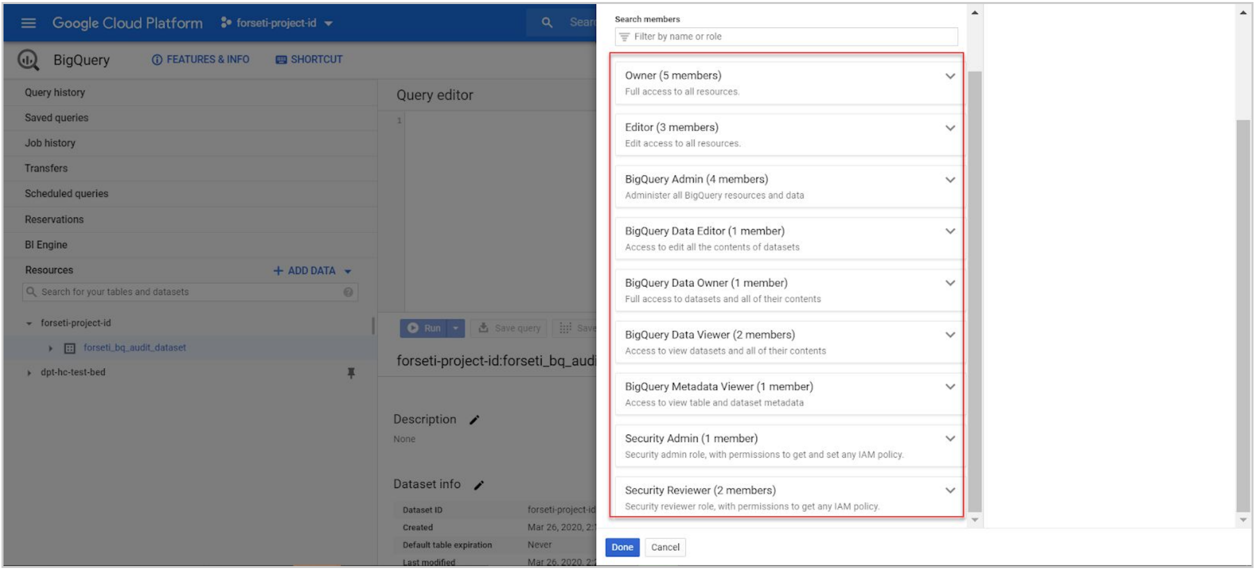


Figure 10 - BigQuery Console

Please note that access to the datasets can be further regulated using Cloud IAM. The IAM permissions provided for the BigQuery dataset are viewable under the *Share Dataset* option (Figure 11). These can be further customized to control access to the BigQuery datasets.



Owner (5 members)
Full access to all resources.

Editor (3 members)
Edit access to all resources.

BigQuery Admin (4 members)
Administer all BigQuery resources and data

BigQuery Data Editor (1 member)
Access to edit all the contents of datasets

BigQuery Data Owner (1 member)
Full access to datasets and all of their contents

BigQuery Data Viewer (2 members)
Access to view datasets and all of their contents

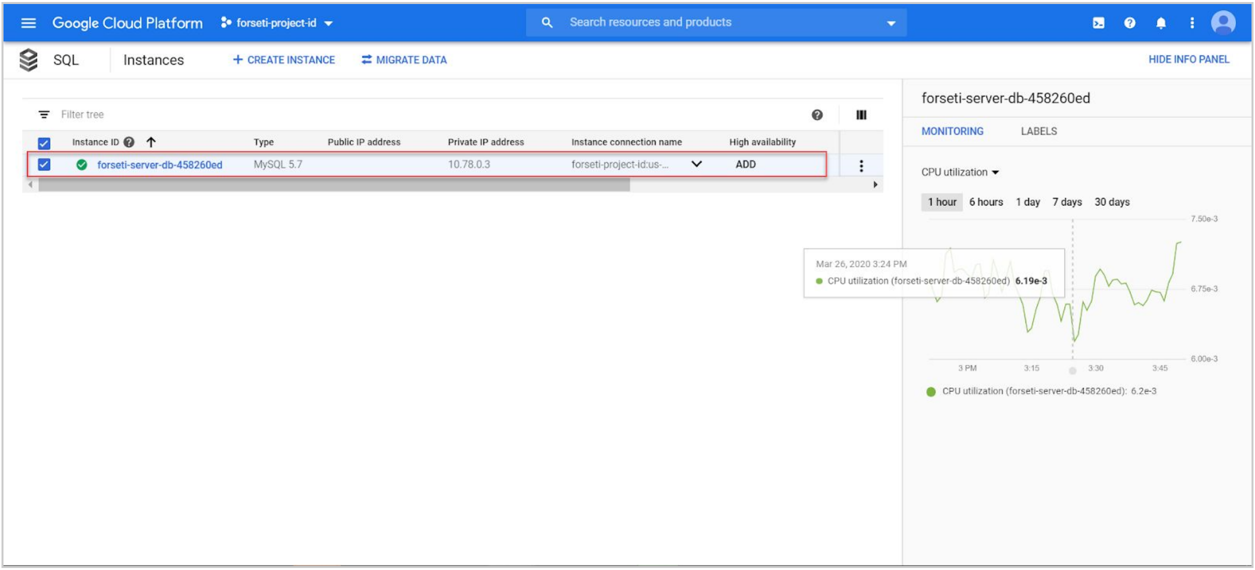
BigQuery Metadata Viewer (1 member)
Access to view table and dataset metadata

Security Admin (1 member)
Security admin role, with permissions to get and set any IAM policy.

Security Reviewer (2 members)
Security reviewer role, with permissions to get any IAM policy.

Figure 11 - IAM Permissions for BigQuery Data

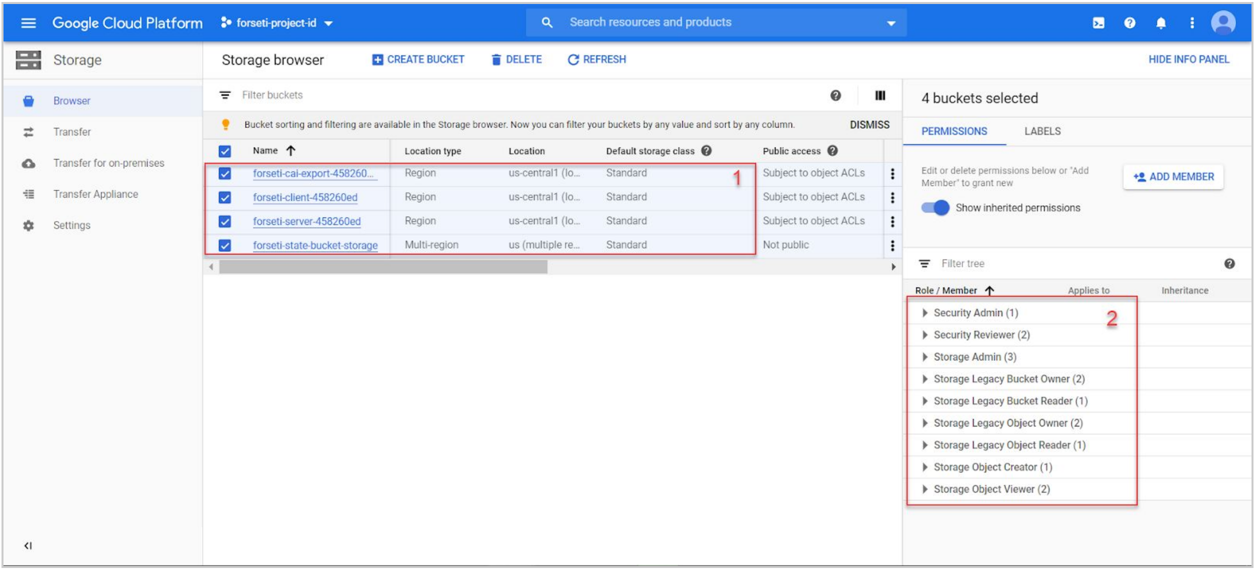
Cloud SQL Console: A Forseti SQL Instance has been created, as can be seen in the Cloud SQL Console screen (Figure 12).



☑	Instance ID	Type	Public IP address	Private IP address	Instance connection name	High availability
☑	forseti-server-db-458260ed	MySQL 5.7		10.78.0.3	forseti-project-id:us-...	ADD

Figure 12 - Cloud SQL Console

Storage Browser Console: In the Storage Browser Console, the storage buckets that are deployed through the template as part of the Forseti project can be seen (Figure 13). The storage buckets and their corresponding permissions can be viewed by selecting the respective buckets in the list.



Bucket sorting and filtering are available in the Storage browser. Now you can filter your buckets by any value and sort by any column.

☑	Name ↑	Location type	Location	Default storage class	Public access
☑	forseti-cai-export-458260...	Region	us-central1 (lo...	Standard	Subject to object ACLs
☑	forseti-client-458260ed	Region	us-central1 (lo...	Standard	Subject to object ACLs
☑	forseti-server-458260ed	Region	us-central1 (lo...	Standard	Subject to object ACLs
☑	forseti-state-bucket-storage	Multi-region	us (multiple re...	Standard	Not public

4 buckets selected

Edit or delete permissions below or "Add Member" to grant new

● Show inherited permissions

Role / Member ↑	Applies to	Inheritance
Security Admin (1)		
Security Reviewer (2)		
Storage Admin (3)		
Storage Legacy Bucket Owner (2)		
Storage Legacy Bucket Reader (1)		
Storage Legacy Object Owner (2)		
Storage Legacy Object Reader (1)		
Storage Object Creator (1)		
Storage Object Viewer (2)		

Figure 13 - 1.Forseti Storage Buckets   2. Bucket Permissions

3.8.2 Data-hosting Project (dpth-ai-ml)

BigQuery Cloud Console: As a part of *dpth-ai-ml* project, three datasets are mapped - one dataset for storing *audit logs*, the other for *raw input data,* and the last one for *enriched data.* The datasets can be seen in the BigQuery Console (Figure 14**)**.

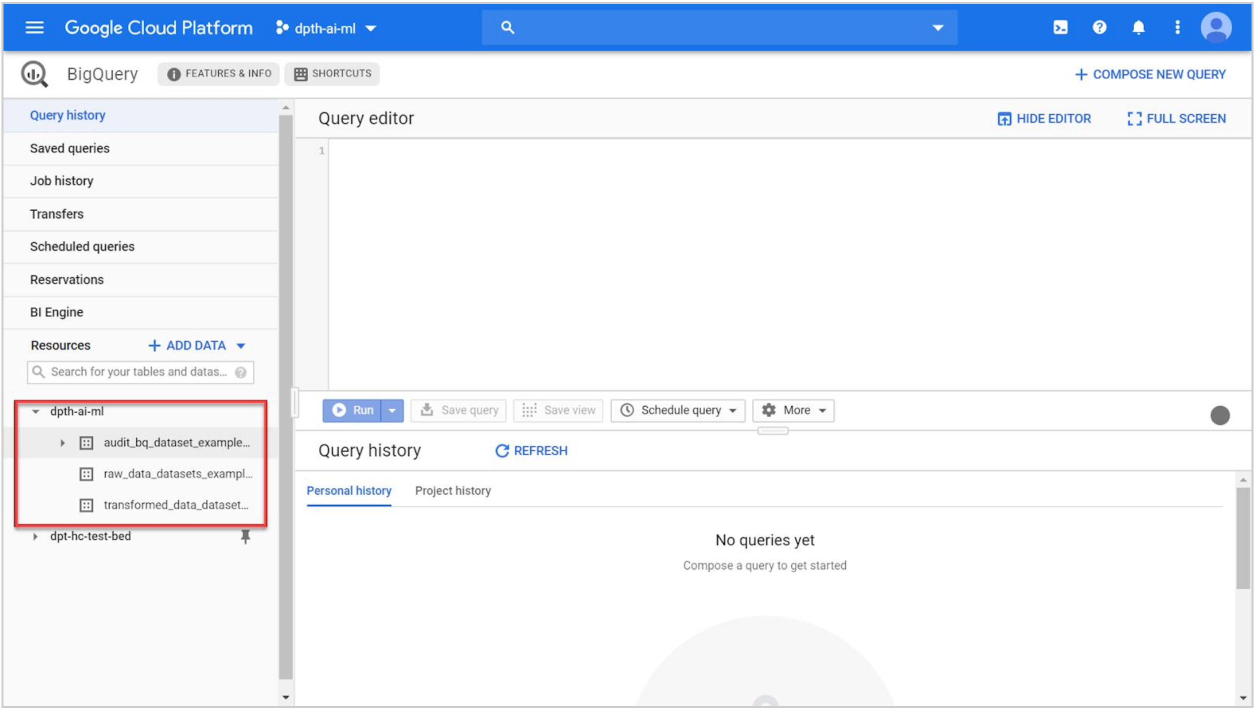
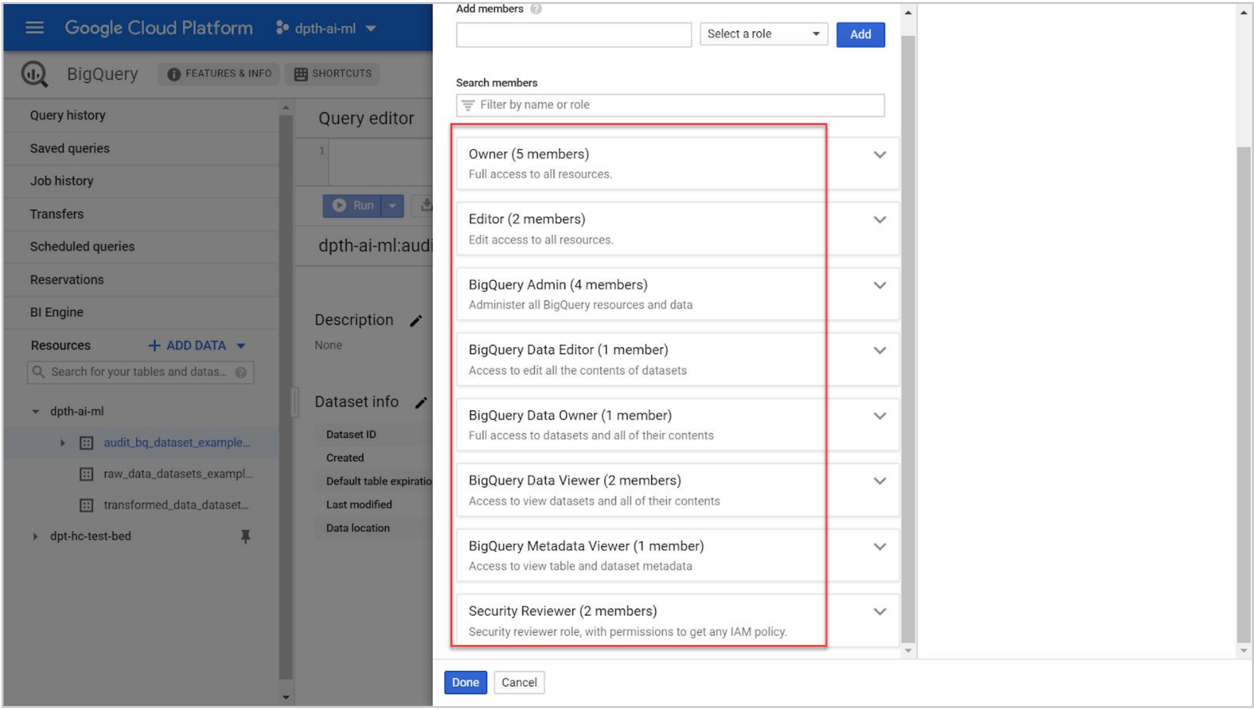


Figure 14 - BigQuery Datasets

Figure 15 and Figure 16 below indicate the permissions granted on the *Audit BigQuery* dataset and the *Transformed Data* dataset respectively. Since the project *dpth-ai-ml* is created inside a folder, some permissions are also inherited from the *Folder level permissions* by the datasets. Under the project name *dpth-ai-ml,* select a dataset and click on *SHARE DATASET* to view the respective permissions.



Owner (5 members)
Full access to all resources.

Editor (2 members)
Edit access to all resources.

BigQuery Admin (4 members)
Administer all BigQuery resources and data

BigQuery Data Editor (1 member)
Access to edit all the contents of datasets

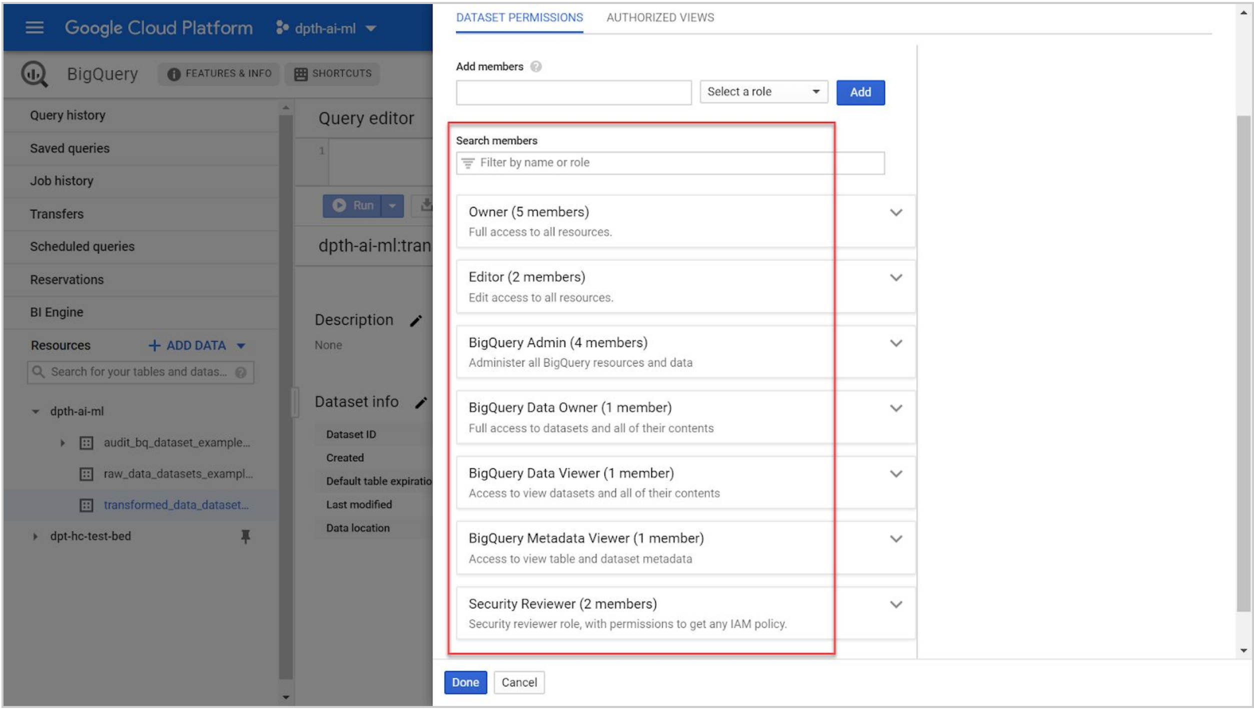
BigQuery Data Owner (1 member)
Full access to datasets and all of their contents

BigQuery Data Viewer (2 members)
Access to view datasets and all of their contents

BigQuery Metadata Viewer (1 member)
Access to view table and dataset metadata

Security Reviewer (2 members)
Security reviewer role, with permissions to get any IAM policy.

Figure 15 - Audit BigQuery dataset permissions



Owner (5 members)
Full access to all resources.

Editor (2 members)
Edit access to all resources.

BigQuery Admin (4 members)
Administer all BigQuery resources and data

BigQuery Data Owner (1 member)
Full access to datasets and all of their contents

BigQuery Data Viewer (1 member)
Access to view datasets and all of their contents

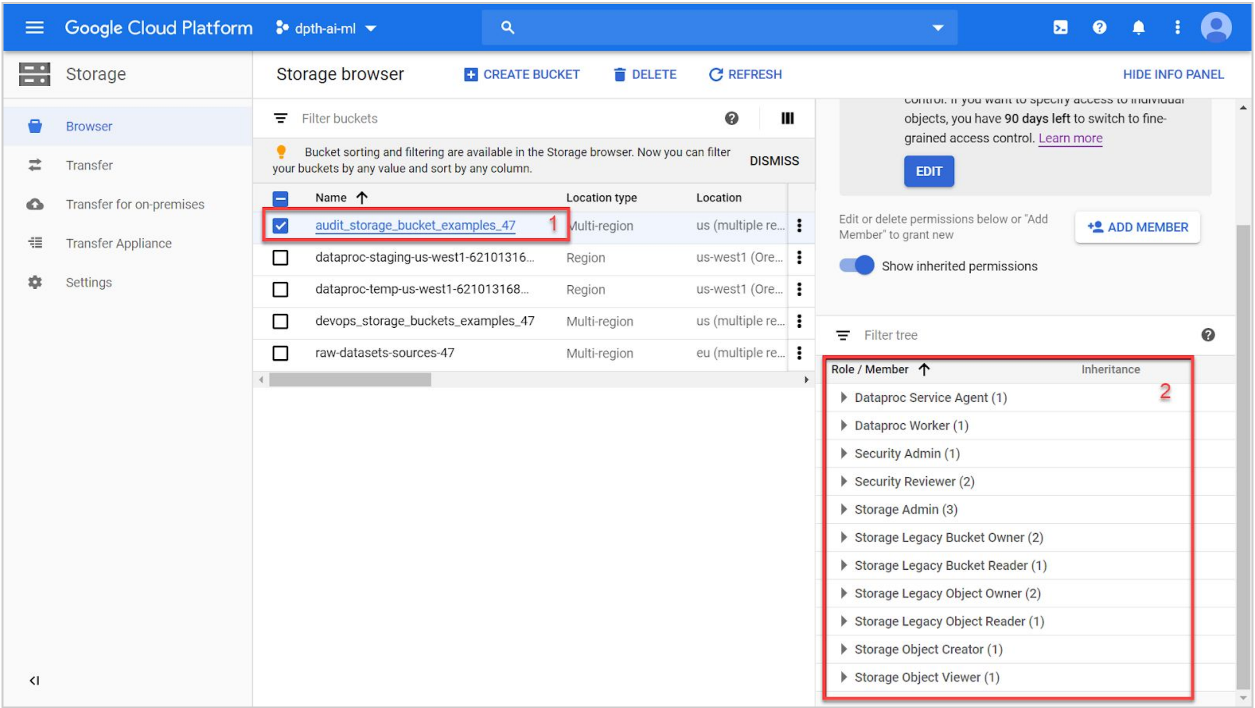
BigQuery Metadata Viewer (1 member)
Access to view table and dataset metadata

Security Reviewer (2 members)
Security reviewer role, with permissions to get any IAM policy.

Figure 16 -Transformed Data dataset permissions

Storage Console: Storage Browser Console shows five Storage buckets created - One bucket for raw data source, two buckets for audit logs and devops state storage and two buckets for

Cloud Dataproc. Select a bucket to show the permissions on it as highlighted in Figure 17, Figure 18 and Figure 19 below.



Bucket sorting and filtering are available in the Storage browser. Now you can filter your buckets by any value and sort by any column.

	Name	Location type	Location
☑	audit_storage_bucket_examples_47	Multi-region	us (multiple re...
□	dataproc-staging-us-west1-62101316...	Region	us-west1 (Ore...
□	dataproc-temp-us-west1-621013168...	Region	us-west1 (Ore...
□	devops_storage_buckets_examples_47	Multi-region	us (multiple re...
□	raw-datasets-sources-47	Multi-region	eu (multiple re...

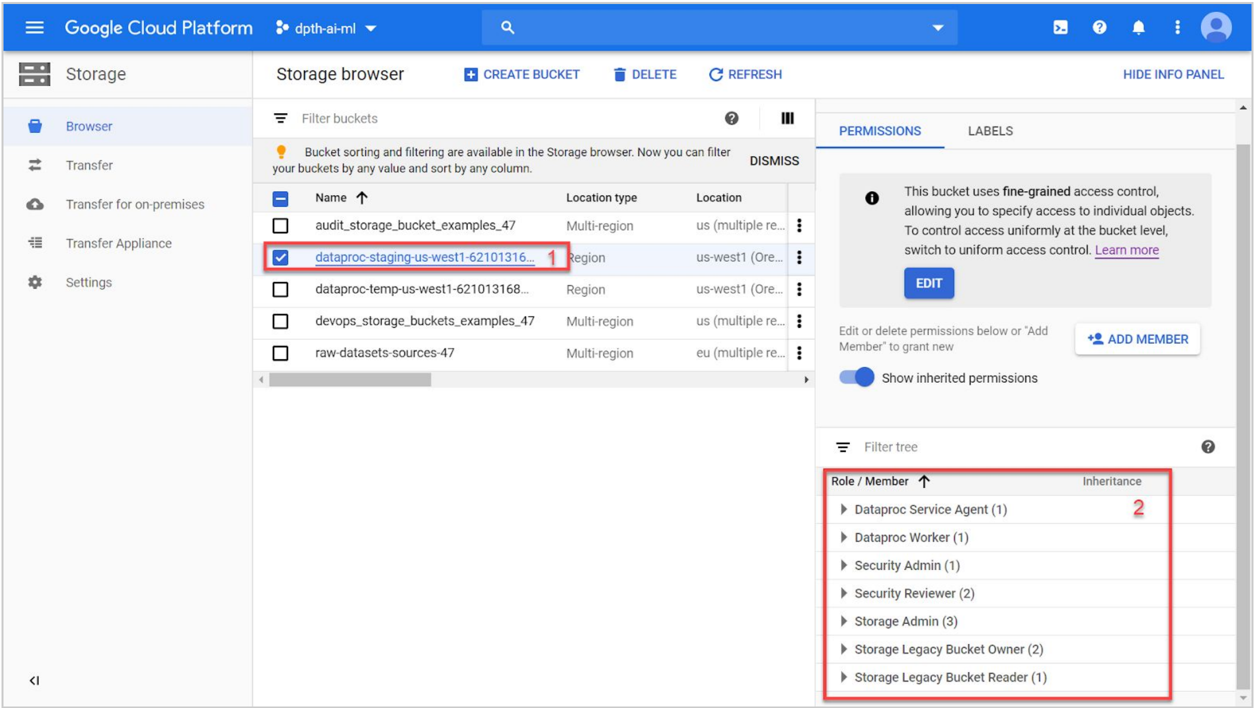
control. If you want to specify access to individual objects, you have **90 days left** to switch to fine-grained access control. Learn more

Edit or delete permissions below or "Add Member" to grant new

Show inherited permissions

Role / Member	Inheritance
Dataproc Service Agent (1)	
Dataproc Worker (1)	
Security Admin (1)	
Security Reviewer (2)	
Storage Admin (3)	
Storage Legacy Bucket Owner (2)	
Storage Legacy Bucket Reader (1)	
Storage Legacy Object Owner (2)	
Storage Legacy Object Reader (1)	
Storage Object Creator (1)	
Storage Object Viewer (1)	

Figure 17 - 1. Audit Storage Bucket   2. Bucket Permissions



Bucket sorting and filtering are available in the Storage browser. Now you can filter your buckets by any value and sort by any column.

	Name	Location type	Location
□	audit_storage_bucket_examples_47	Multi-region	us (multiple re...
☑	dataproc-staging-us-west1-62101316...	Region	us-west1 (Ore...
□	dataproc-temp-us-west1-621013168...	Region	us-west1 (Ore...
□	devops_storage_buckets_examples_47	Multi-region	us (multiple re...
□	raw-datasets-sources-47	Multi-region	eu (multiple re...

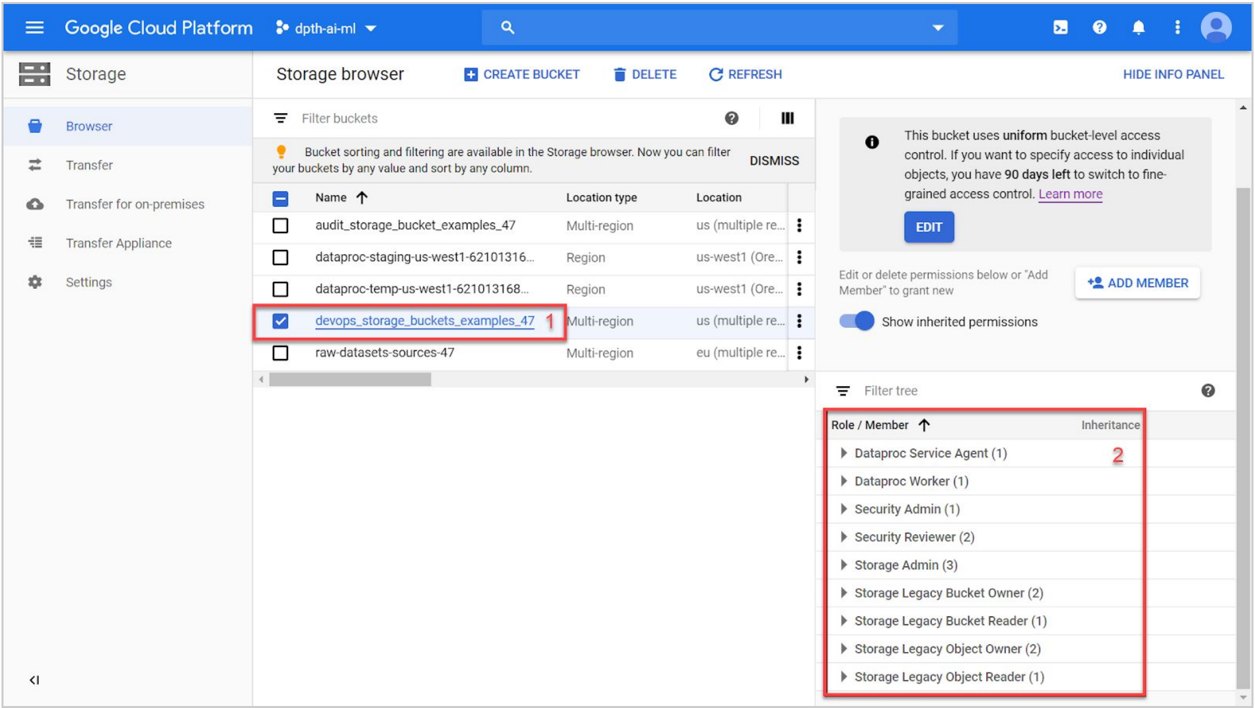
This bucket uses **fine-grained** access control, allowing you to specify access to individual objects. To control access uniformly at the bucket level, switch to uniform access control. Learn more

Edit or delete permissions below or "Add Member" to grant new

Show inherited permissions

Role / Member	Inheritance
Dataproc Service Agent (1)	
Dataproc Worker (1)	
Security Admin (1)	
Security Reviewer (2)	
Storage Admin (3)	
Storage Legacy Bucket Owner (2)	
Storage Legacy Bucket Reader (1)	

Figure 18 - 1.Dataproc Storage Bucket   2. Bucket Permissions



Bucket sorting and filtering are available in the Storage browser. Now you can filter your buckets by any value and sort by any column.

☐	Name ↑	Location type	Location
☐	audit_storage_bucket_examples_47	Multi-region	us (multiple re…
☐	dataproc-staging-us-west1-62101316…	Region	us-west1 (Ore…
☐	dataproc-temp-us-west1-621013168…	Region	us-west1 (Ore…
☑	devops_storage_buckets_examples_47	Multi-region	us (multiple re…
☐	raw-datasets-sources-47	Multi-region	eu (multiple re…

This bucket uses **uniform** bucket-level access control. If you want to specify access to individual objects, you have **90 days left** to switch to fine-grained access control. Learn more

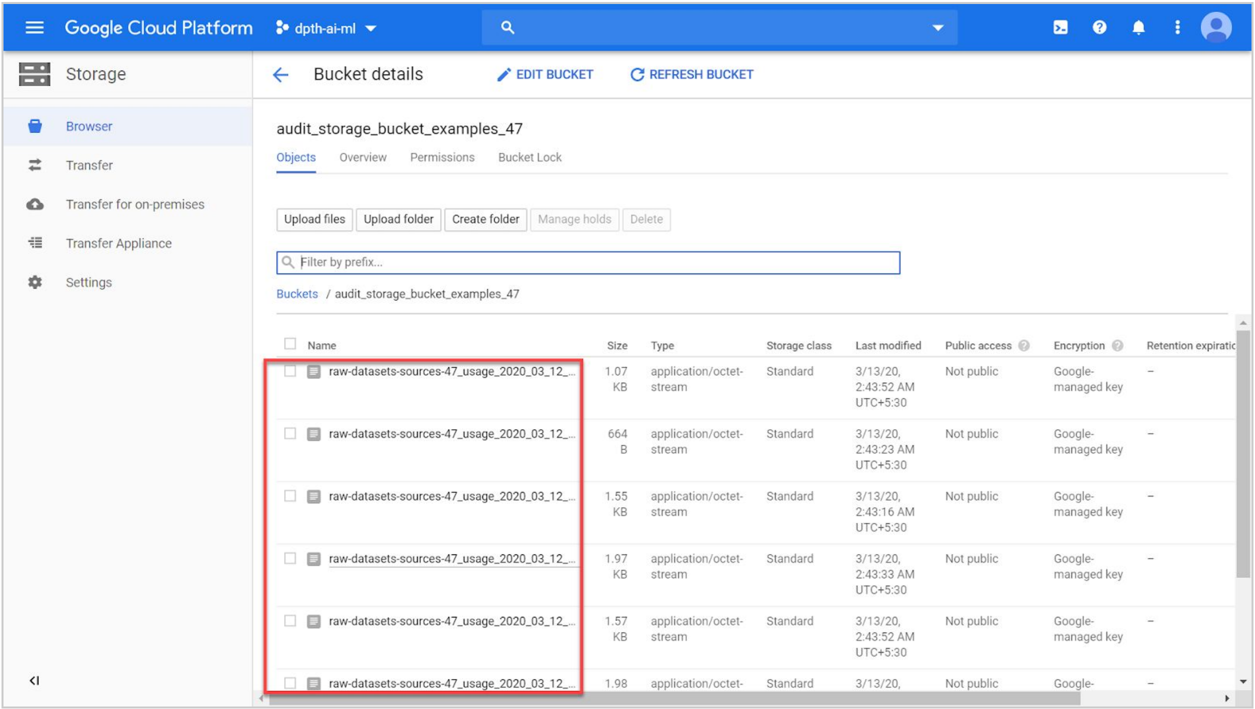
Edit or delete permissions below or "Add Member" to grant new

Show inherited permissions

Role / Member ↑	Inheritance
Dataproc Service Agent (1)	
Dataproc Worker (1)	
Security Admin (1)	
Security Reviewer (2)	
Storage Admin (3)	
Storage Legacy Bucket Owner (2)	
Storage Legacy Bucket Reader (1)	
Storage Legacy Object Owner (2)	
Storage Legacy Object Reader (1)	

Figure 19 - 1.DevOps Storage Bucket   2. Bucket Permissions

Datasets created also have objects stored inside. The Audit Storage bucket has usage and log objects corresponding to different resources as shown in Figure 20. The Dataproc Staging bucket has objects relating to master and worker instances, jobs and properties as highlighted in Figure 21. The Devops Storage bucket has terraform states corresponding to different DPT phases of deployment stored in them as highlighted in Figure 22. Click on the buckets to view the object.

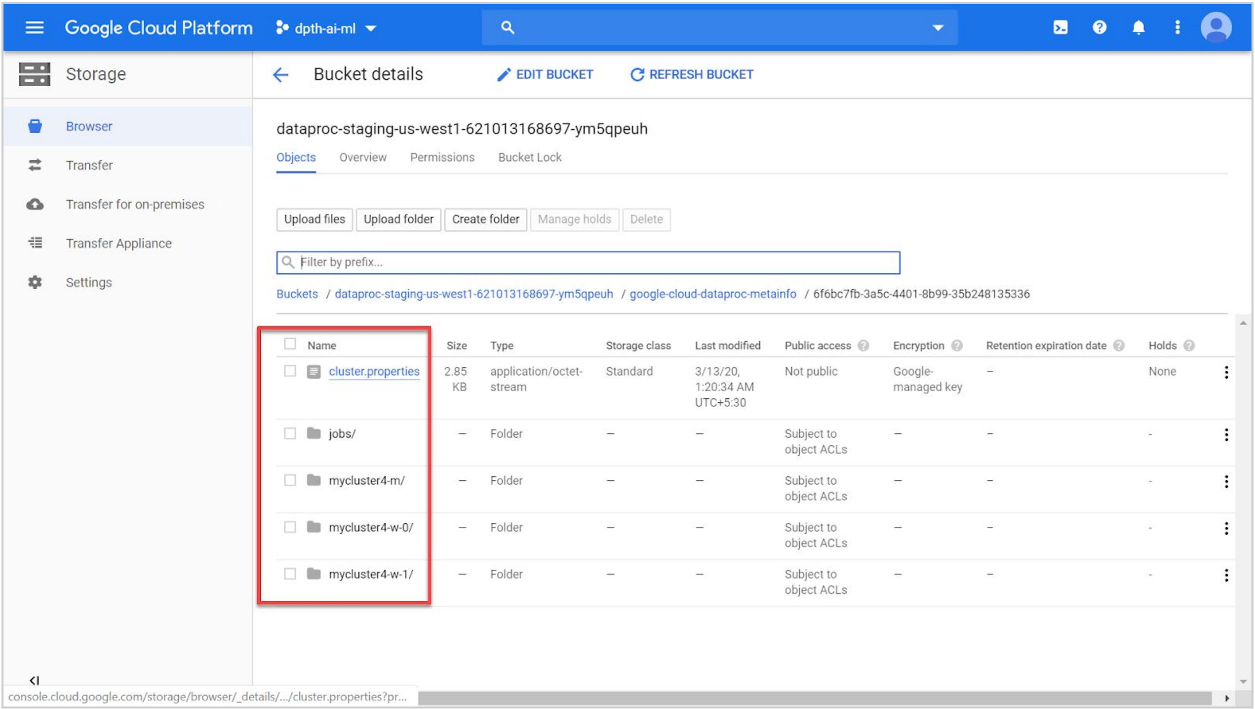


audit_storage_bucket_examples_47

Buckets / audit_storage_bucket_examples_47

Name	Size	Type	Storage class	Last modified	Public access	Encryption	Retention expiratic
raw-datasets-sources-47_usage_2020_03_12_...	1.07 KB	application/octet-stream	Standard	3/13/20, 2:43:52 AM UTC+5:30	Not public	Google-managed key	–
raw-datasets-sources-47_usage_2020_03_12_...	664 B	application/octet-stream	Standard	3/13/20, 2:43:23 AM UTC+5:30	Not public	Google-managed key	–
raw-datasets-sources-47_usage_2020_03_12_...	1.55 KB	application/octet-stream	Standard	3/13/20, 2:43:16 AM UTC+5:30	Not public	Google-managed key	–
raw-datasets-sources-47_usage_2020_03_12_...	1.97 KB	application/octet-stream	Standard	3/13/20, 2:43:33 AM UTC+5:30	Not public	Google-managed key	–
raw-datasets-sources-47_usage_2020_03_12_...	1.57 KB	application/octet-stream	Standard	3/13/20, 2:43:52 AM UTC+5:30	Not public	Google-managed key	–
raw-datasets-sources-47_usage_2020_03_12_...	1.98	application/octet-	Standard	3/13/20,	Not public	Google-	–

Figure 20 - Audit objects



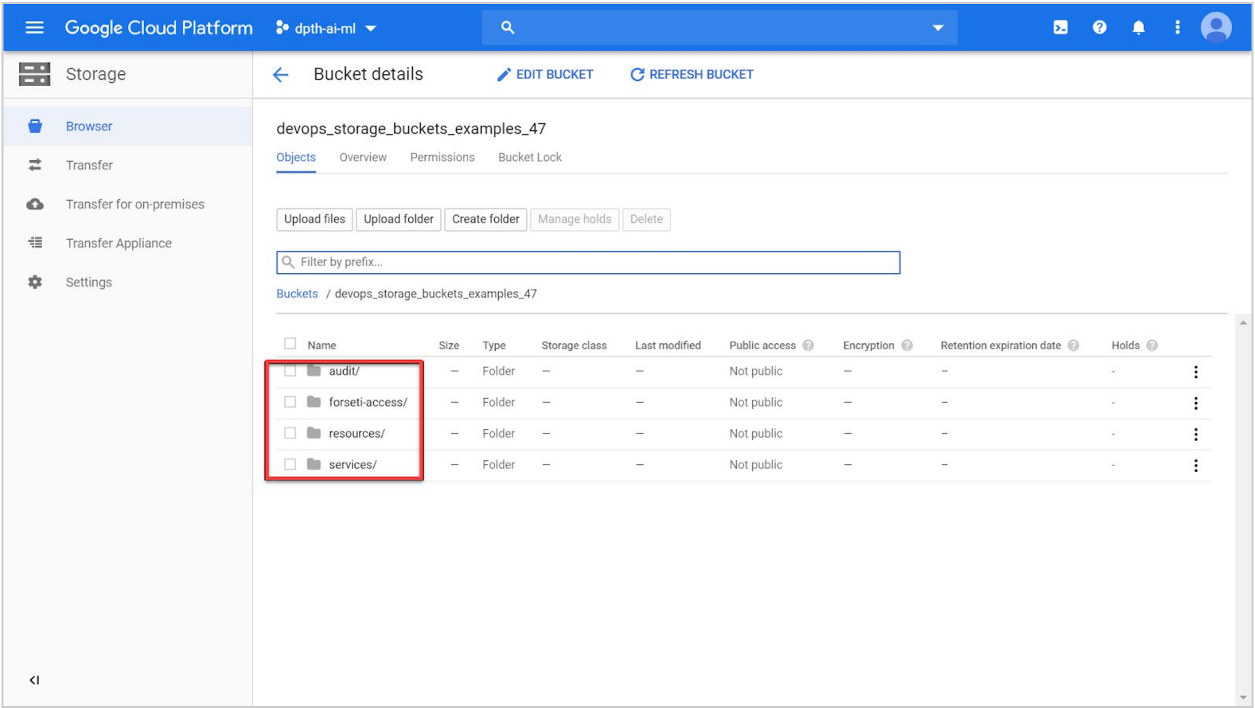
dataproc-staging-us-west1-621013168697-ym5qpeuh

Buckets / dataproc-staging-us-west1-621013168697-ym5qpeuh / google-cloud-dataproc-metainfo / 6f6bc7fb-3a5c-4401-8b99-35b248135336

Name	Size	Type	Storage class	Last modified	Public access	Encryption	Retention expiration date	Holds
cluster.properties	2.85 KB	application/octet-stream	Standard	3/13/20, 1:20:34 AM UTC+5:30	Not public	Google-managed key	–	None
jobs/	–	Folder	–	–	Subject to object ACLs	–	–	-
mycluster4-m/	–	Folder	–	–	Subject to object ACLs	–	–	-
mycluster4-w-0/	–	Folder	–	–	Subject to object ACLs	–	–	-
mycluster4-w-1/	–	Folder	–	–	Subject to object ACLs	–	–	-

console.cloud.google.com/storage/browser/_details/.../cluster.properties?pr...

Figure 21 - Dataproc objects

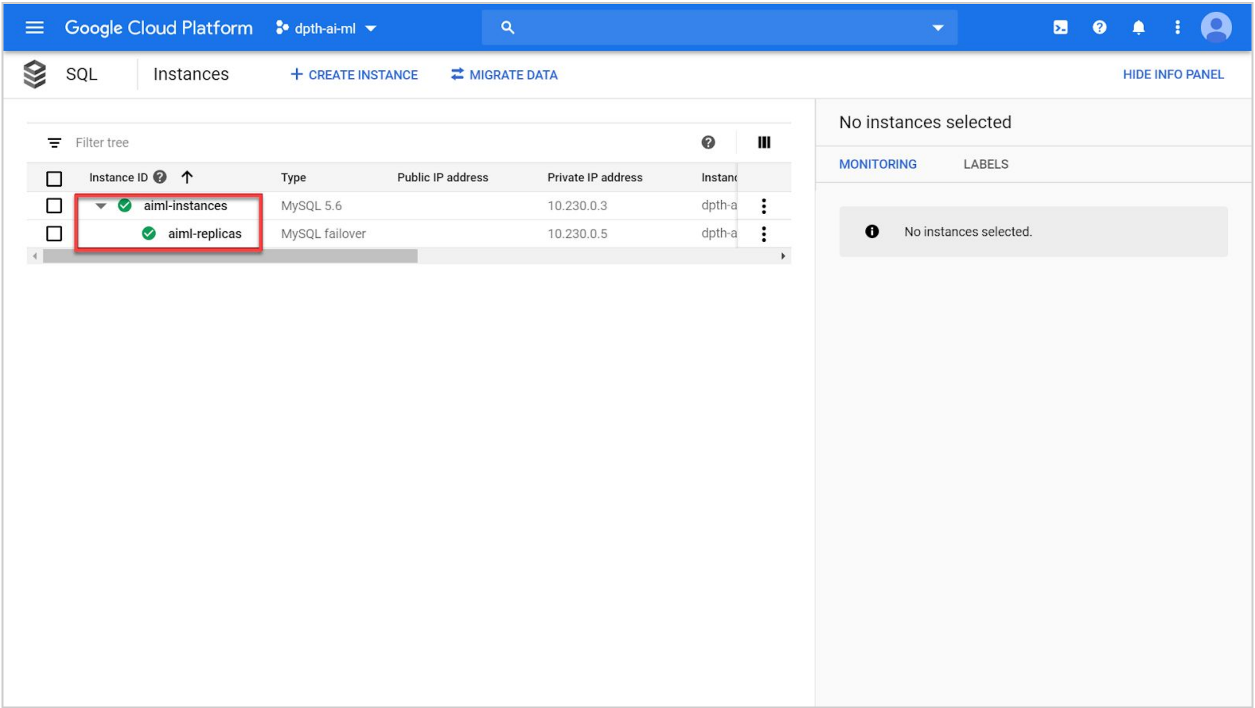


devops_storage_buckets_examples_47

Name	Size	Type	Storage class	Last modified	Public access	Encryption	Retention expiration date	Holds
audit/	–	Folder	–	–	Not public	–	–	-
forseti-access/	–	Folder	–	–	Not public	–	–	-
resources/	–	Folder	–	–	Not public	–	–	-
services/	–	Folder	–	–	Not public	–	–	-

Figure 22 - DevOps objects

Cloud SQL: Browse the Cloud SQL console to review the private SQL instance, and its replica, which were created through the template below (Figure 23).



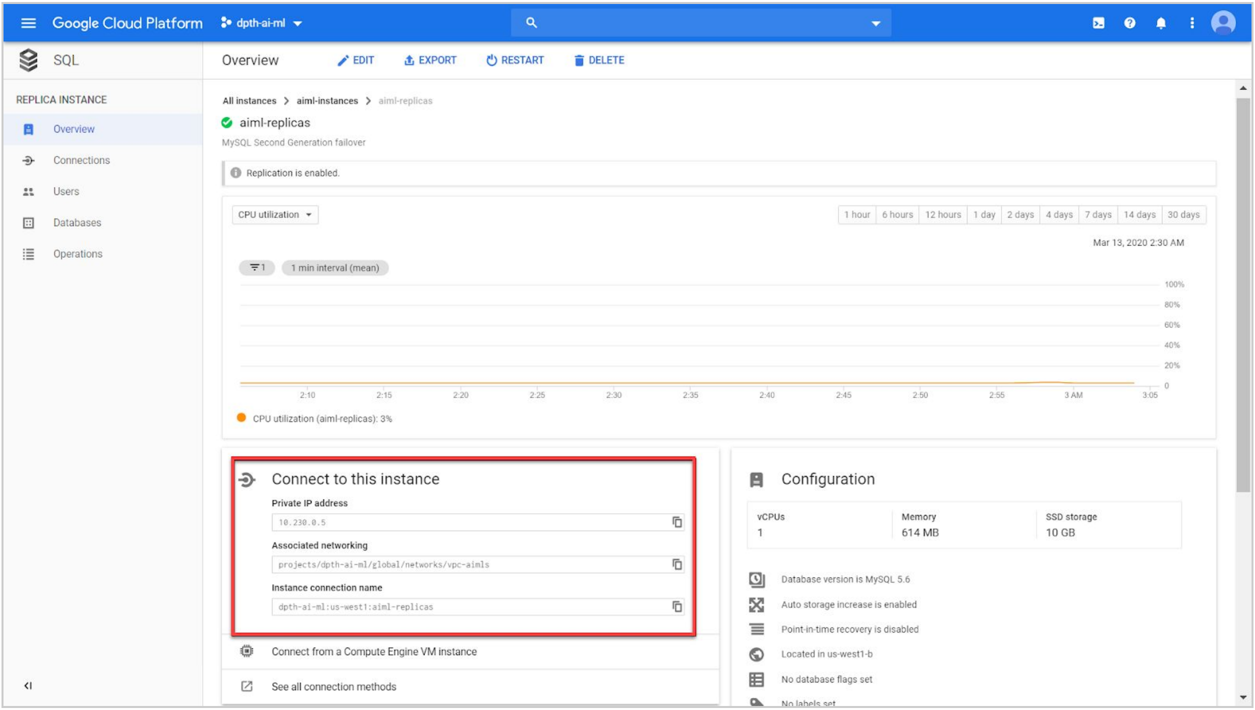
□	Instance ID	Type	Public IP address	Private IP address	Instanc
□	aiml-instances	MySQL 5.6		10.230.0.3	dpth-a
□	aiml-replicas	MySQL failover		10.230.0.5	dpth-a

No instances selected

No instances selected.

Figure 23 - SQL Instances

More details about the private SQL instances are available by clicking the instances. Private IP address and associated networking details is highlighted in Figure 24. This instance can only be accessed privately by the users in Virtual Private Cloud as per the rules provided.



aiml-replicas

MySQL Second Generation failover

Replication is enabled.

Connect to this instance

Private IP address

```
10.230.0.5
```

Associated networking

```
projects/dpth-ai-ml/global/networks/vpc-aimls
```

Instance connection name

```
dpth-ai-ml:us-west1:aiml-replicas
```

Connect from a Compute Engine VM instance

See all connection methods

Configuration

vCPUs	Memory	SSD storage
1	614 MB	10 GB

Database version is MySQL 5.6

Auto storage increase is enabled

Point-in-time recovery is disabled

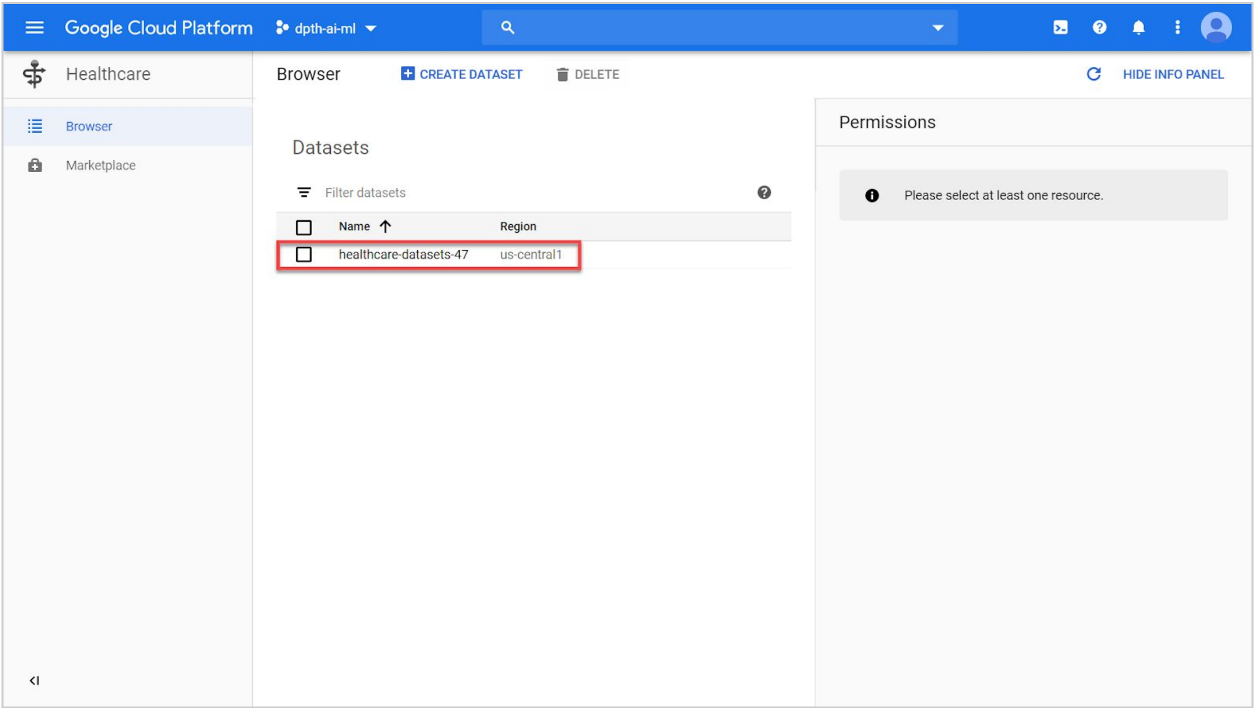
Located in us-west1-b

No database flags set

No labels set

Figure 24 - SQL Instance networking details

Healthcare Console: In the Healthcare Console, the Healthcare Dataset created by the DPT template is listed. Permissions on the dataset can be seen by selecting the dataset (Figure 25).

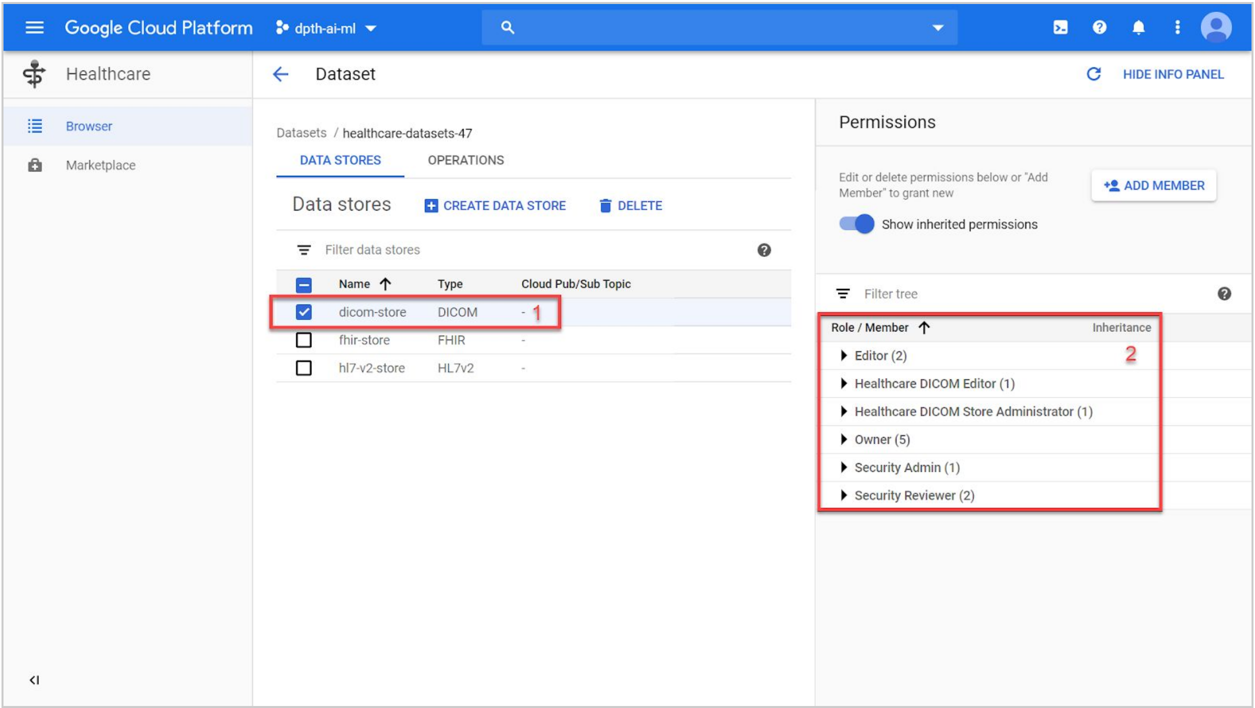


Name	Region
healthcare-datasets-47	us-central1

Please select at least one resource.

Figure 25 - Healthcare Dataset

Select the dataset to see the *DICOM, FHIR* and *HL7-V2* datastores. Figure 26, Figure 27 and Figure 28 show the permissions for *DICOM*, *FHIR* and *HL7-V2* respectively.



Datasets / healthcare-datasets-47

Data stores

	Name ↑	Type	Cloud Pub/Sub Topic
☑	dicom-store	DICOM	-
□	fhir-store	FHIR	-
□	hl7-v2-store	HL7v2	-

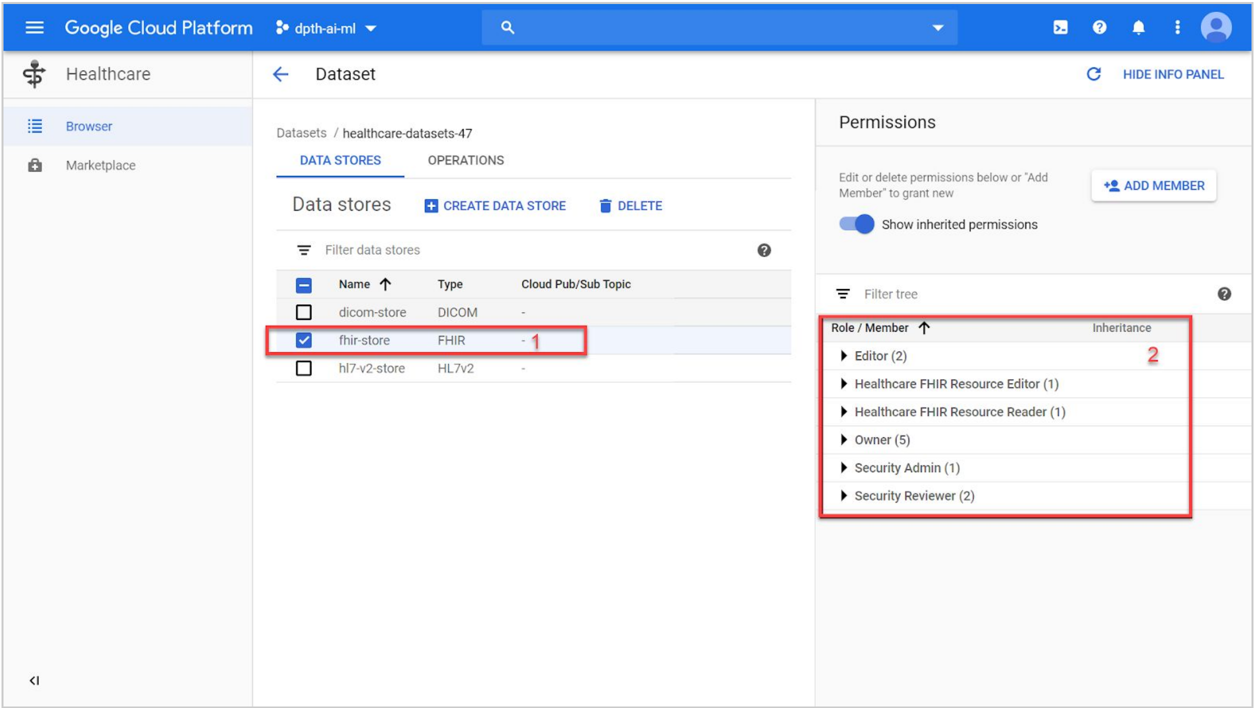
Permissions

Edit or delete permissions below or "Add Member" to grant new

Show inherited permissions

Role / Member ↑	Inheritance
Editor (2)	
Healthcare DICOM Editor (1)	
Healthcare DICOM Store Administrator (1)	
Owner (5)	
Security Admin (1)	
Security Reviewer (2)	

Figure 26 - 1. DICOM Datastore   2. Datastore Permissions



Datasets / healthcare-datasets-47

Data stores

	Name ↑	Type	Cloud Pub/Sub Topic
□	dicom-store	DICOM	-
☑	fhir-store	FHIR	-
□	hl7-v2-store	HL7v2	-

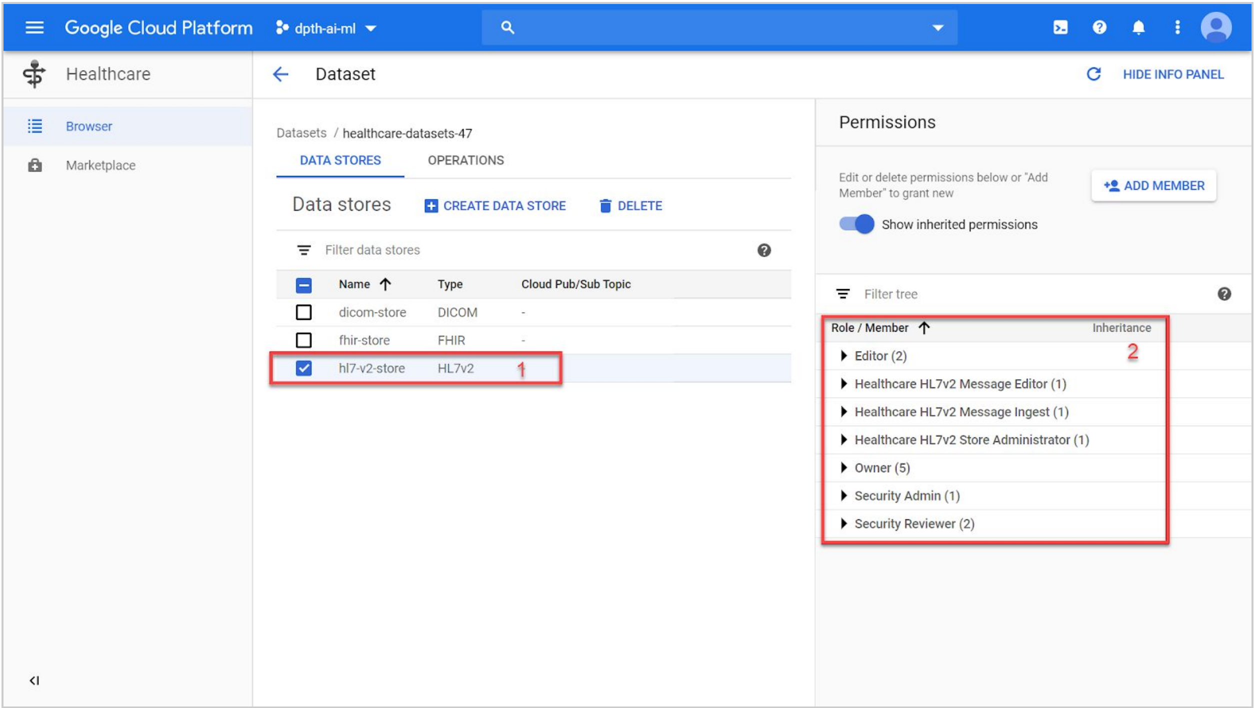
Permissions

Edit or delete permissions below or "Add Member" to grant new

Show inherited permissions

Role / Member ↑	Inheritance
Editor (2)	
Healthcare FHIR Resource Editor (1)	
Healthcare FHIR Resource Reader (1)	
Owner (5)	
Security Admin (1)	
Security Reviewer (2)	

Figure 27 - 1. FHIR Datastore   2. Datastore Permissions



Google Cloud Platform · dpth-ai-ml

Healthcare · Dataset · HIDE INFO PANEL

Browser · Marketplace

Datasets / healthcare-datasets-47

DATA STORES · OPERATIONS

Data stores · CREATE DATA STORE · DELETE

Filter data stores

	Name ↑	Type	Cloud Pub/Sub Topic
□	dicom-store	DICOM	-
□	fhir-store	FHIR	-
☑	hl7-v2-store	HL7v2	

1

Permissions

Edit or delete permissions below or "Add Member" to grant new

ADD MEMBER

Show inherited permissions

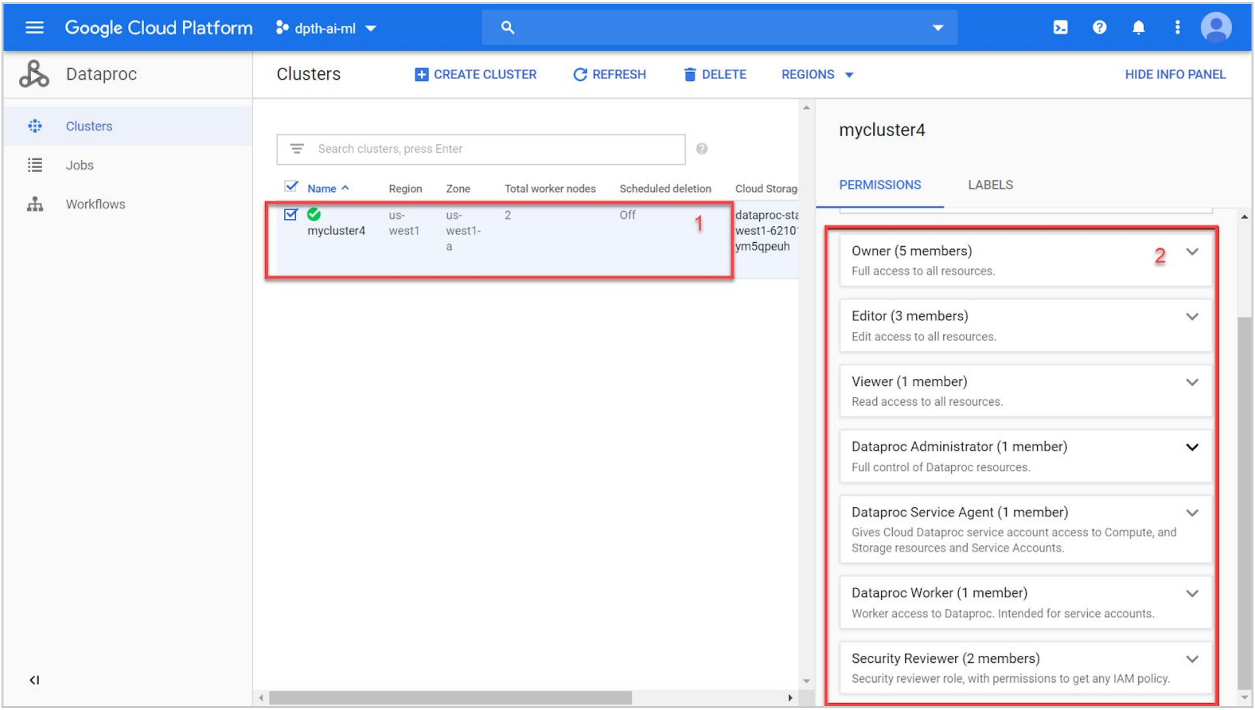
Filter tree

Role / Member ↑	Inheritance
Editor (2)	
Healthcare HL7v2 Message Editor (1)	
Healthcare HL7v2 Message Ingest (1)	
Healthcare HL7v2 Store Administrator (1)	
Owner (5)	
Security Admin (1)	
Security Reviewer (2)	

2

Figure 28 - 1. HL7-v2 Datastore 2. Datastore Permissions

Cloud Dataproc: Go to the Dataproc Clusters page. The cluster created is seen in the cluster list as shown below. Selecting the cluster shows the permission on it as highlighted in Figure 29.



Google Cloud Platform · dpth-ai-ml

Dataproc · Clusters · CREATE CLUSTER · REFRESH · DELETE · REGIONS · HIDE INFO PANEL

Clusters · Jobs · Workflows

Search clusters, press Enter

☑	Name ^	Region	Zone	Total worker nodes	Scheduled deletion	Cloud Storag
☑	✓ mycluster4	us-west1	us-west1-a	2	Off	dataproc-sta west1-6210 ym5qpeuh

1

mycluster4

PERMISSIONS · LABELS

Owner (5 members)
Full access to all resources.

2

Editor (3 members)
Edit access to all resources.

Viewer (1 member)
Read access to all resources.

Dataproc Administrator (1 member)
Full control of Dataproc resources.

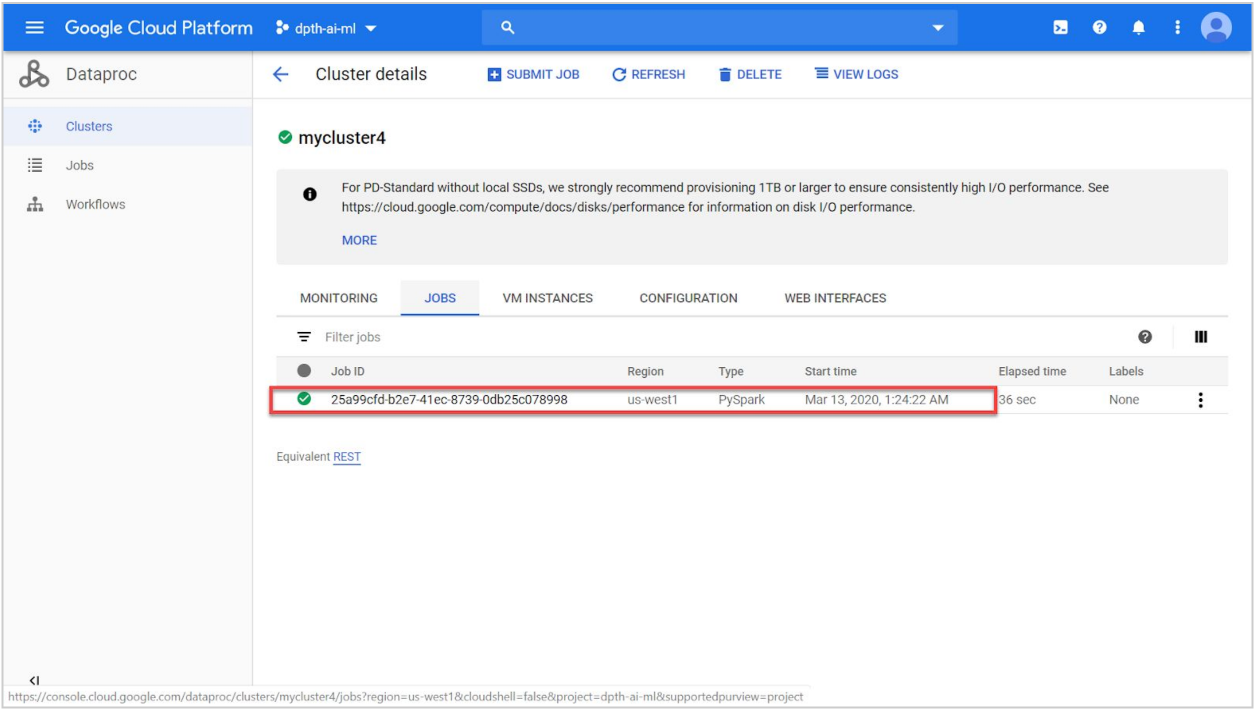
Dataproc Service Agent (1 member)
Gives Cloud Dataproc service account access to Compute, and Storage resources and Service Accounts.

Dataproc Worker (1 member)
Worker access to Dataproc. Intended for service accounts.

Security Reviewer (2 members)
Security reviewer role, with permissions to get any IAM policy.

Figure 29 - 1.Dataproc Cluster 2. Cluster permissions

In the cluster under *JOBS* tab, view jobs related to the cluster. Figure 30 shows the example *PySpark* created.



mycluster4

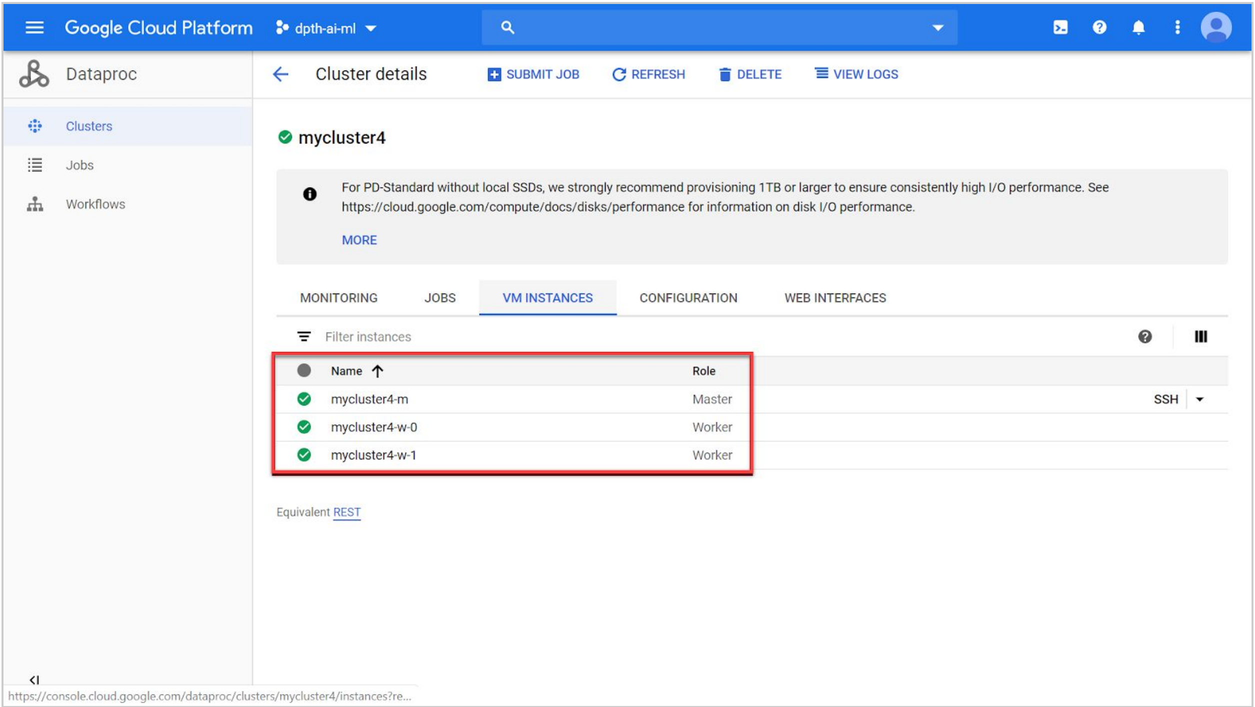
For PD-Standard without local SSDs, we strongly recommend provisioning 1TB or larger to ensure consistently high I/O performance. See https://cloud.google.com/compute/docs/disks/performance for information on disk I/O performance.

Job ID	Region	Type	Start time	Elapsed time	Labels
25a99cfd-b2e7-41ec-8739-0db25c078998	us-west1	PySpark	Mar 13, 2020, 1:24:22 AM	36 sec	None

https://console.cloud.google.com/dataproc/clusters/mycluster4/jobs?region=us-west1&cloudshell=false&project=dpth-ai-ml&supportedpurview=project

Figure 30 - Dataproc Cluster

Under *VM INSTANCES* tab, view master and worker instances of the cluster. Figure 31 shows one master cluster (mycluster4-m) and two worker clusters (mycluster4-w0 and mycluster4-w1).

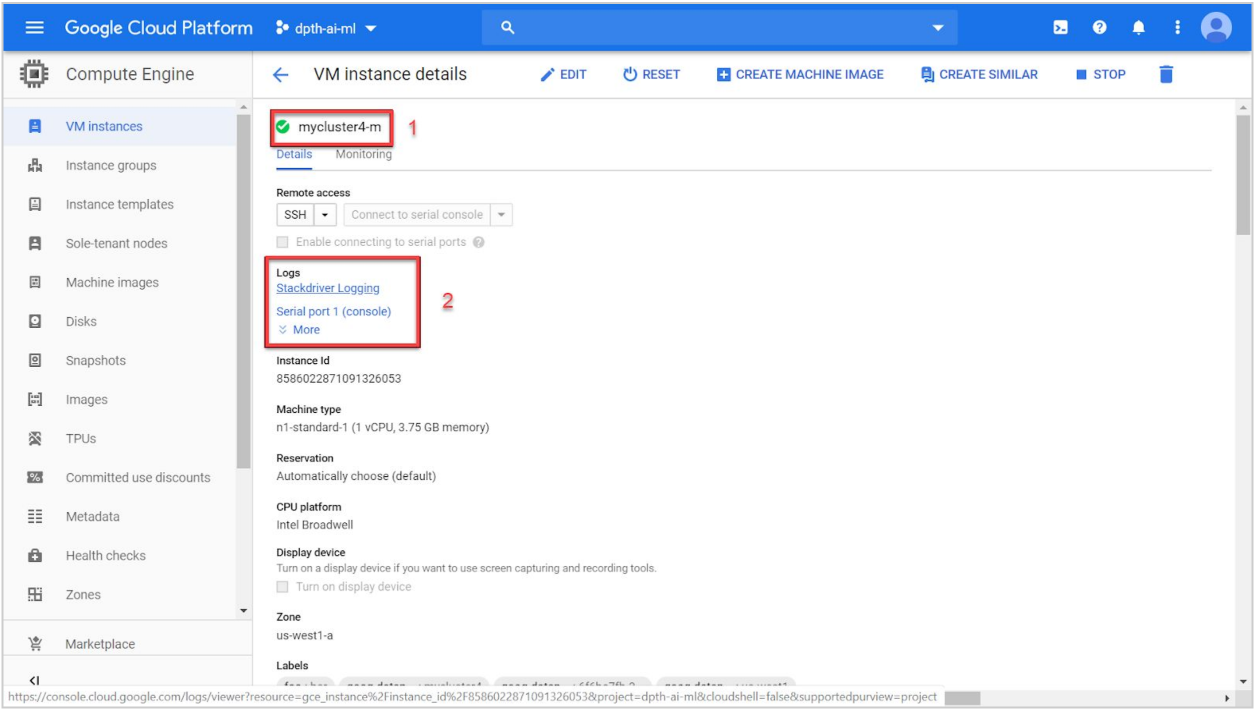


For PD-Standard without local SSDs, we strongly recommend provisioning 1TB or larger to ensure consistently high I/O performance. See https://cloud.google.com/compute/docs/disks/performance for information on disk I/O performance.

Name	Role
mycluster4-m	Master
mycluster4-w-0	Worker
mycluster4-w-1	Worker

Figure 31 - Dataproc VM Instances

Selecting the instances opens the corresponding details page. Figure 32 shows the details page for *mycluster4-m*. Note that Stackdriver Logging is enabled as highlighted in the figure.



mycluster4-m

Remote access

□ Enable connecting to serial ports

Logs
Stackdriver Logging
Serial port 1 (console)
More

Instance Id
8586022871091326053

Machine type
n1-standard-1 (1 vCPU, 3.75 GB memory)

Reservation
Automatically choose (default)

CPU platform
Intel Broadwell

Display device
Turn on a display device if you want to use screen capturing and recording tools.
□ Turn on display device

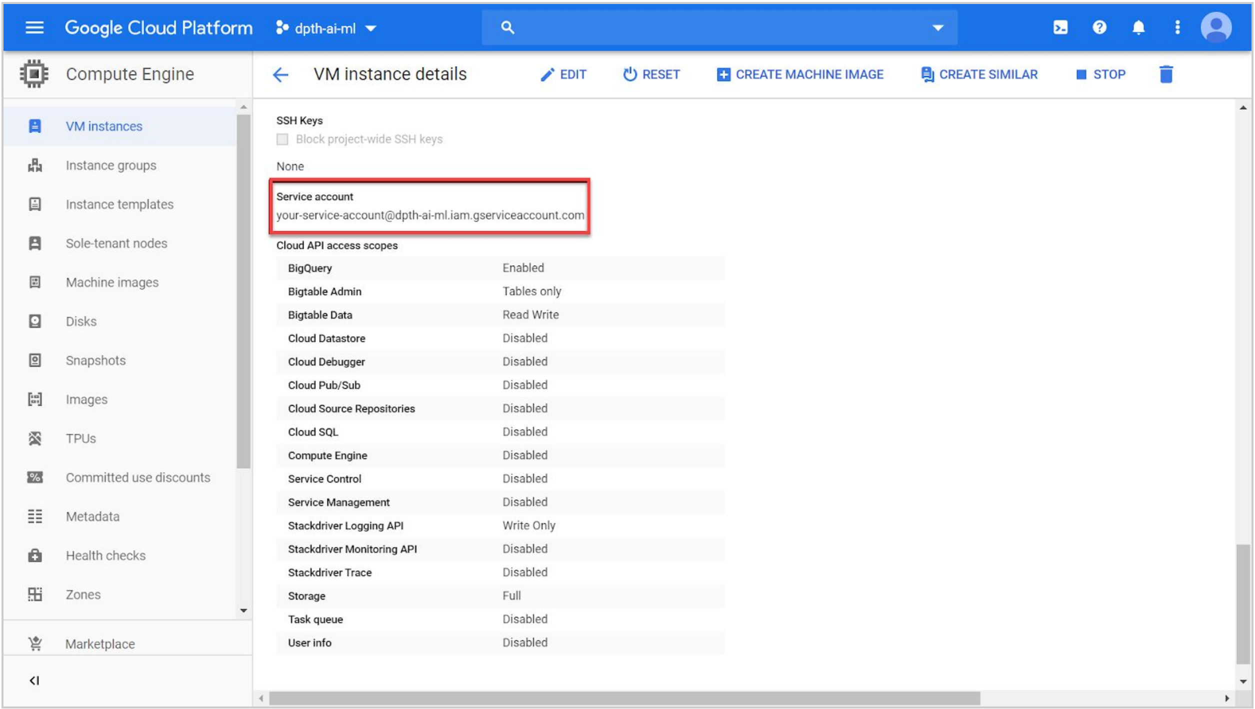
Zone
us-west1-a

Labels

https://console.cloud.google.com/logs/viewer?resource=gce_instance%2Finstance_id%2F8586022871091326053&project=dpth-ai-ml&cloudshell=false&supportedpurview=project

Figure 32 - 1. Dataproc Master Instance 2. Stackdriver Logging

Scroll down to see the custom service account associated with the Dataproc VMs as highlighted in Figure 33.



SSH Keys

□ Block project-wide SSH keys

None

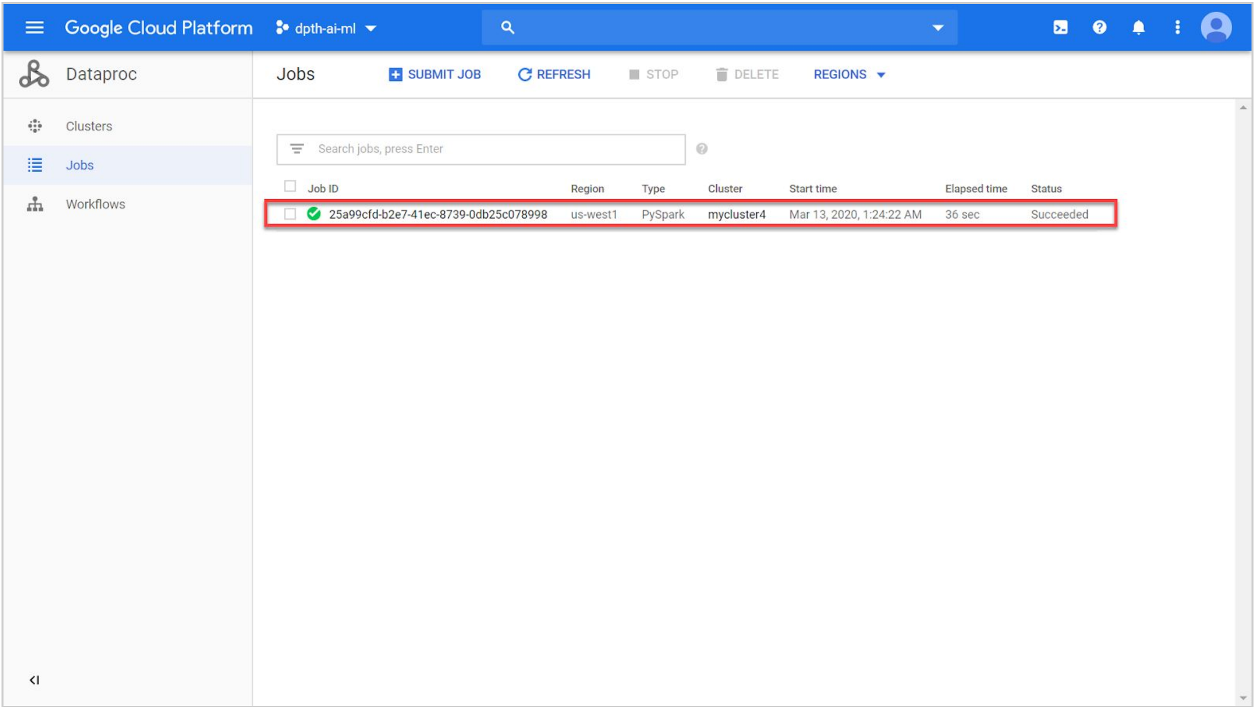
Service account
your-service-account@dpth-ai-ml.iam.gserviceaccount.com

Cloud API access scopes

BigQuery	Enabled
Bigtable Admin	Tables only
Bigtable Data	Read Write
Cloud Datastore	Disabled
Cloud Debugger	Disabled
Cloud Pub/Sub	Disabled
Cloud Source Repositories	Disabled
Cloud SQL	Disabled
Compute Engine	Disabled
Service Control	Disabled
Service Management	Disabled
Stackdriver Logging API	Write Only
Stackdriver Monitoring API	Disabled
Stackdriver Trace	Disabled
Storage	Full
Task queue	Disabled
User info	Disabled

Figure 33. - Dataproc Service Account

All jobs are listed in the Jobs pane. Figure 34. shows the PySpark job listed.

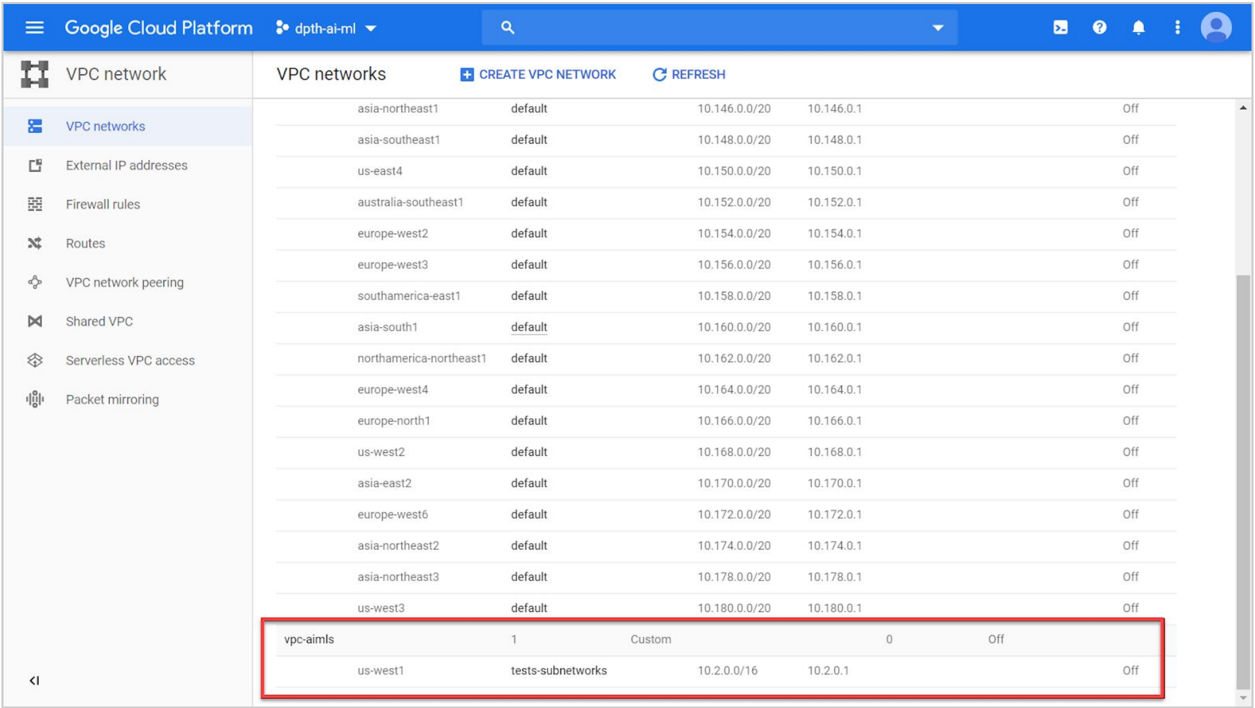


Jobs

□	Job ID	Region	Type	Cluster	Start time	Elapsed time	Status
□	✓ 25a99cfd-b2e7-41ec-8739-0db25c078998	us-west1	PySpark	mycluster4	Mar 13, 2020, 1:24:22 AM	36 sec	Succeeded

Figure 34 - Example PySpark Job

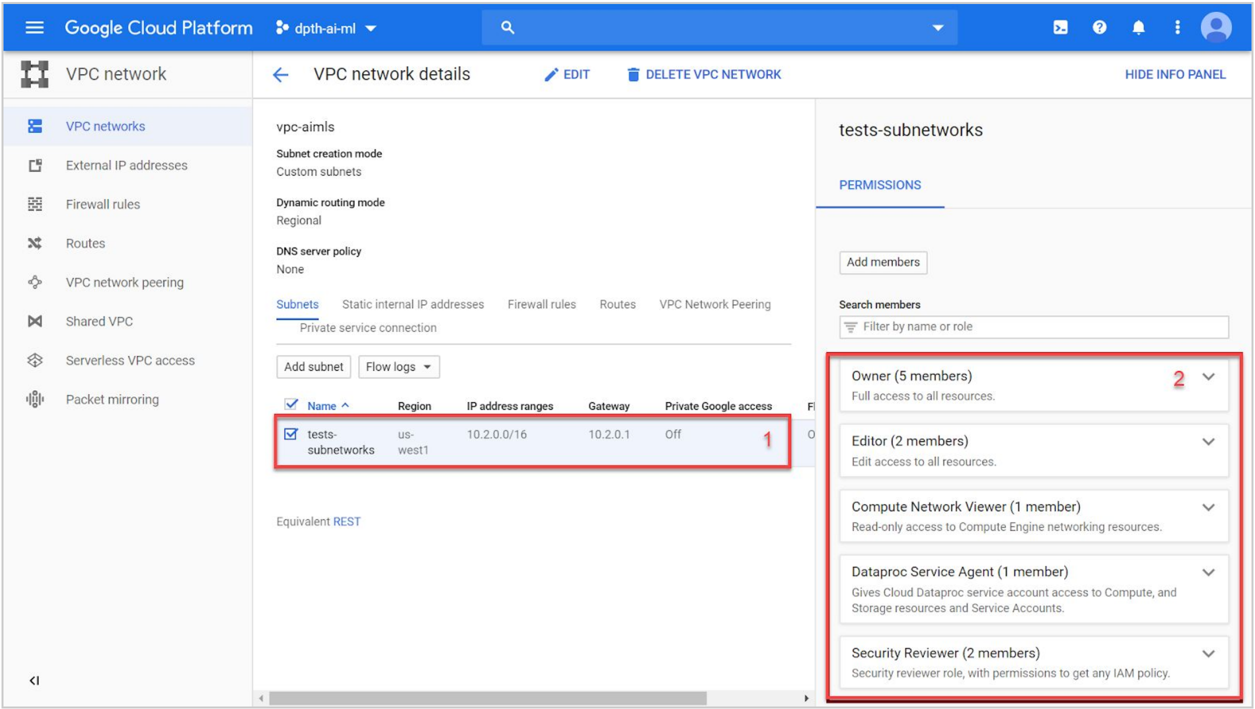
Virtual Private Cloud Networks: On the VPC network Console, the custom created VPC is listed with the corresponding subnetwork. Figure 35 shows the custom created *vpc-aimls* network and its subnetwork *tests-subnetworks* highlighted.



	asia-northeast1	default		10.146.0.0/20	10.146.0.1			Off
	asia-southeast1	default		10.148.0.0/20	10.148.0.1			Off
	us-east4	default		10.150.0.0/20	10.150.0.1			Off
	australia-southeast1	default		10.152.0.0/20	10.152.0.1			Off
	europe-west2	default		10.154.0.0/20	10.154.0.1			Off
	europe-west3	default		10.156.0.0/20	10.156.0.1			Off
	southamerica-east1	default		10.158.0.0/20	10.158.0.1			Off
	asia-south1	default		10.160.0.0/20	10.160.0.1			Off
	northamerica-northeast1	default		10.162.0.0/20	10.162.0.1			Off
	europe-west4	default		10.164.0.0/20	10.164.0.1			Off
	europe-north1	default		10.166.0.0/20	10.166.0.1			Off
	us-west2	default		10.168.0.0/20	10.168.0.1			Off
	asia-east2	default		10.170.0.0/20	10.170.0.1			Off
	europe-west6	default		10.172.0.0/20	10.172.0.1			Off
	asia-northeast2	default		10.174.0.0/20	10.174.0.1			Off
	asia-northeast3	default		10.178.0.0/20	10.178.0.1			Off
	us-west3	default		10.180.0.0/20	10.180.0.1			Off
vpc-aimls		1	Custom			0	Off	
	us-west1	tests-subnetworks		10.2.0.0/16	10.2.0.1			Off

Figure 35 - VPC Network

For more details, Select the custom created VPC. Figure 36 shows the details page with permissions highlighted.



vpc-aimls

Subnet creation mode
Custom subnets

Dynamic routing mode
Regional

DNS server policy
None

Name	Region	IP address ranges	Gateway	Private Google access
☑ tests-subnetworks	us-west1	10.2.0.0/16	10.2.0.1	Off

tests-subnetworks

Search members

Owner (5 members)
Full access to all resources.

Editor (2 members)
Edit access to all resources.

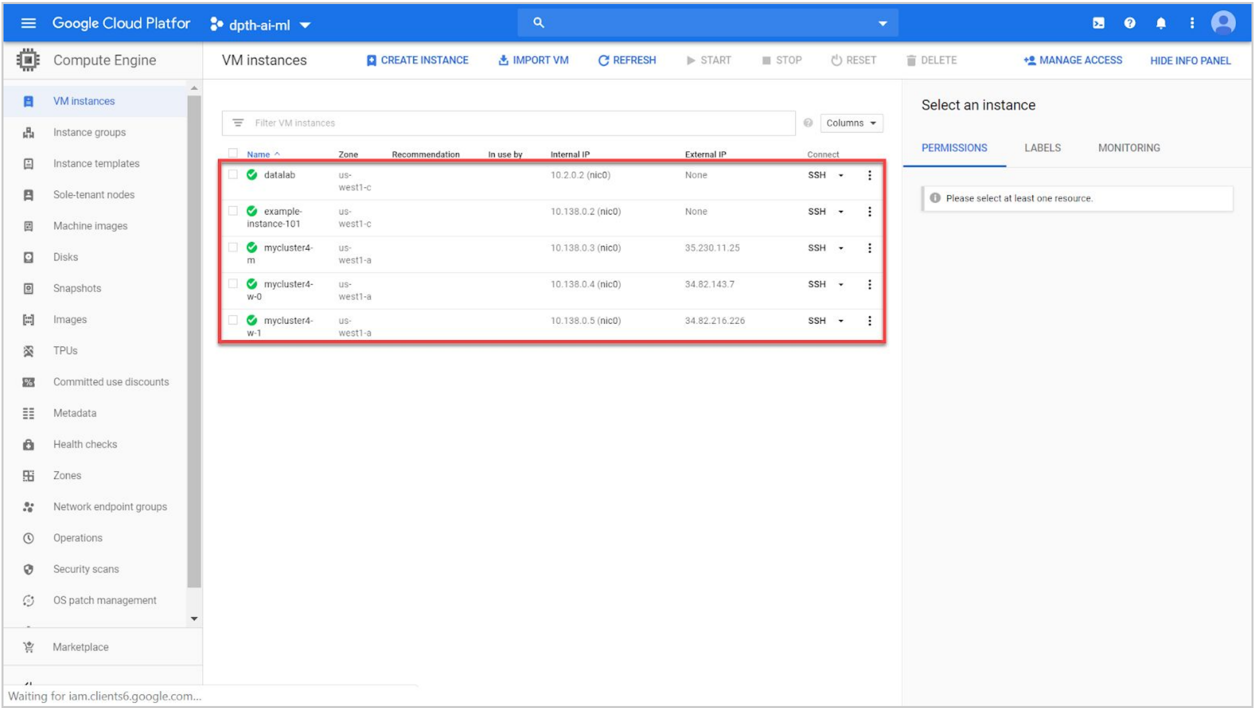
Compute Network Viewer (1 member)
Read-only access to Compute Engine networking resources.

Dataproc Service Agent (1 member)
Gives Cloud Dataproc service account access to Compute, and Storage resources and Service Accounts.

Security Reviewer (2 members)
Security reviewer role, with permissions to get any IAM policy.

Figure 36 - 1. Subnetwork   2. Permissions

Virtual Machines: On VM instances console, view the list of VM instances. Figure 37 shows five instances created - one for Datalab, one for Deep Learning Virtual Machines and three for Cloud Dataproc.



VM instances

	Name ^	Zone	Recommendation	In use by	Internal IP	External IP	Connect
□	✓ datalab	us-west1-c			10.2.0.2 (nic0)	None	SSH
□	✓ example-instance-101	us-west1-c			10.138.0.2 (nic0)	None	SSH
□	✓ mycluster4-m	us-west1-a			10.138.0.3 (nic0)	35.230.11.25	SSH
□	✓ mycluster4-w-0	us-west1-a			10.138.0.4 (nic0)	34.82.143.7	SSH
□	✓ mycluster4-w-1	us-west1-a			10.138.0.5 (nic0)	34.82.216.226	SSH

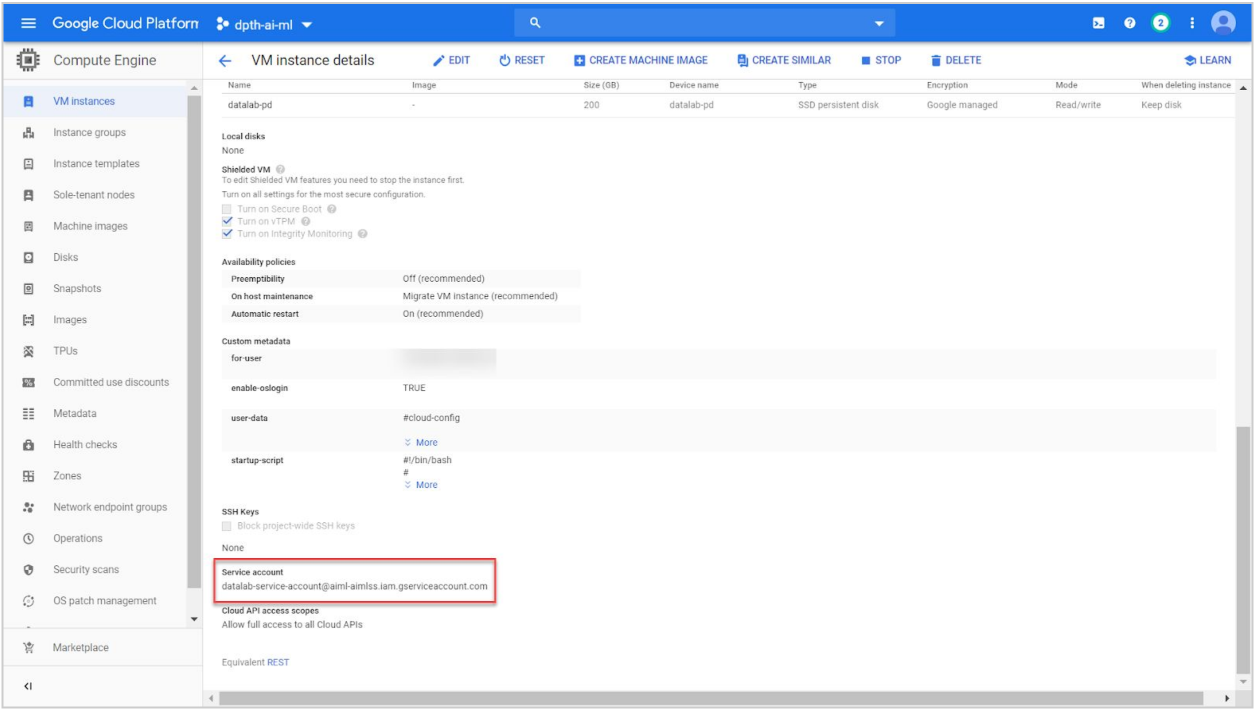
Select an instance

Please select at least one resource.

Waiting for iam.clients6.google.com...

Figure 37 - Virtual Machines

For more details on the Datalab instance as shown below in Figure 38, select the Datalab instance in the *VM instances* page. Note that a custom service account is used as highlighted in the figure below.



VM instance details

Name	Image	Size (GB)	Device name	Type	Encryption	Mode	When deleting instance
datalab-pd	-	200	datalab-pd	SSD persistent disk	Google managed	Read/write	Keep disk

Local disks
None

Shielded VM
To edit Shielded VM features you need to stop the instance first.
Turn on all settings for the most secure configuration.
□ Turn on Secure Boot
☑ Turn on vTPM
☑ Turn on Integrity Monitoring

Availability policies

Preemptibility	Off (recommended)
On host maintenance	Migrate VM instance (recommended)
Automatic restart	On (recommended)

Custom metadata

for-user	[redacted]
enable-oslogin	TRUE
user-data	#cloud-config More
startup-script	#!/bin/bash # More

SSH Keys
□ Block project-wide SSH keys

None

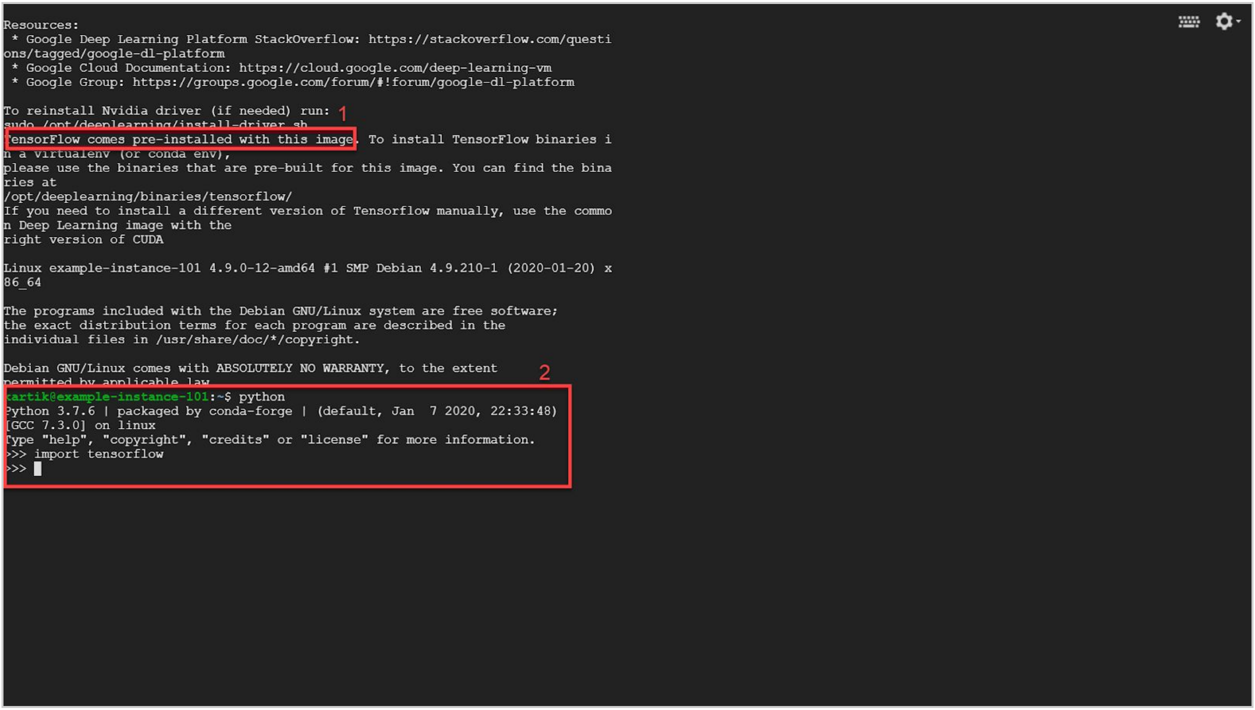
Service account
datalab-service-account@aiml-aimlss.iam.gserviceaccount.com

Cloud API access scopes
Allow full access to all Cloud APIs

Equivalent REST

Figure 38 - Datalab Details

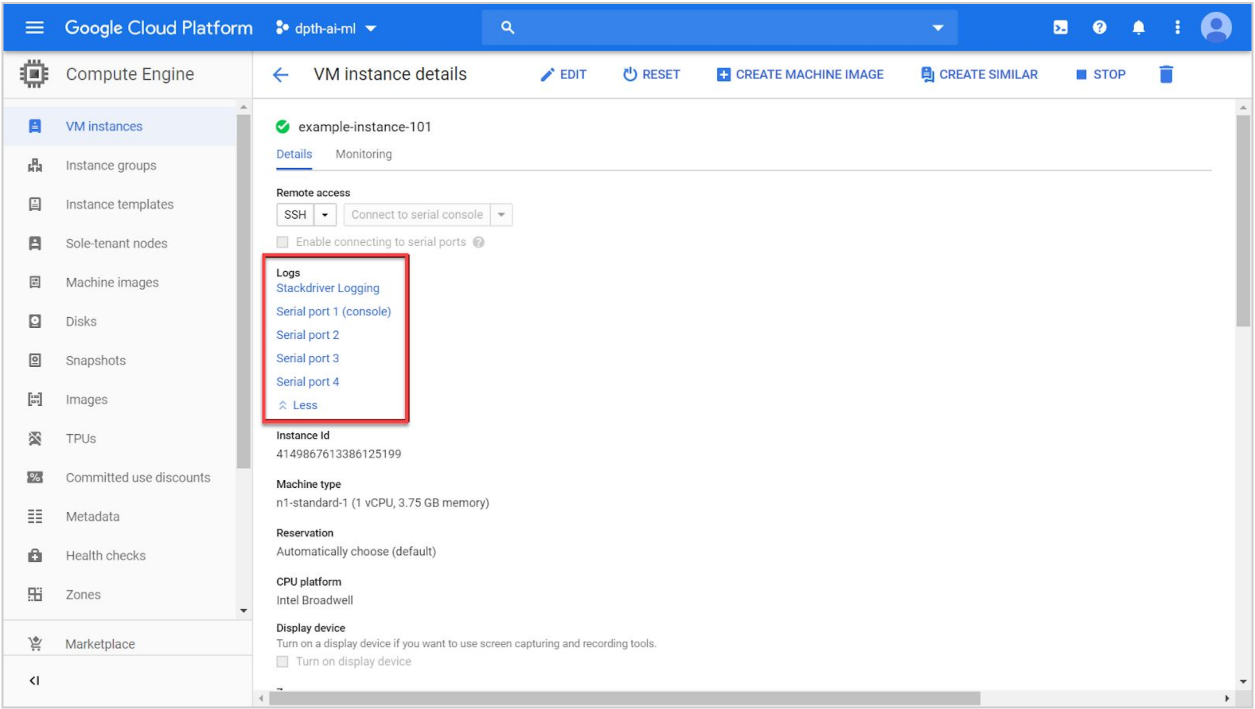
From the *VM instances* page, click on the SSH button in front of the Deep Learning Virtual Machine (*example-instance-101*) to connect to the VM through SSH. Figure 39 shows the window that pops up after successful connection. Note that this VM has TensorFlow pre-installed as seen in the highlighted section by successfully importing Tensorflow in Python.



```
Resources:
 * Google Deep Learning Platform StackOverflow: https://stackoverflow.com/questi
ons/tagged/google-dl-platform
 * Google Cloud Documentation: https://cloud.google.com/deep-learning-vm
 * Google Group: https://groups.google.com/forum/#!forum/google-dl-platform

To reinstall Nvidia driver (if needed) run:
sudo /opt/deeplearning/install-driver.sh
TensorFlow comes pre-installed with this image. To install TensorFlow binaries i
n a virtualenv (or conda env),
please use the binaries that are pre-built for this image. You can find the bina
ries at
/opt/deeplearning/binaries/tensorflow/
If you need to install a different version of Tensorflow manually, use the commo
n Deep Learning image with the
right version of CUDA

Linux example-instance-101 4.9.0-12-amd64 #1 SMP Debian 4.9.210-1 (2020-01-20) x
86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
kartik@example-instance-101:~$ python
Python 3.7.6 | packaged by conda-forge | (default, Jan  7 2020, 22:33:48)
[GCC 7.3.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import tensorflow
>>> 
```

Figure 39 - 1. Tensorflow Installed   2. Python importing Tensorflow

Stackdriver Logging is enabled for all VM instances. Figure 40 shows Stackdriver Logging enabled for *example-instance-101*.



example-instance-101

Remote access

□ Enable connecting to serial ports

Logs
Stackdriver Logging
Serial port 1 (console)
Serial port 2
Serial port 3
Serial port 4

Instance Id
4149867613386125199

Machine type
n1-standard-1 (1 vCPU, 3.75 GB memory)

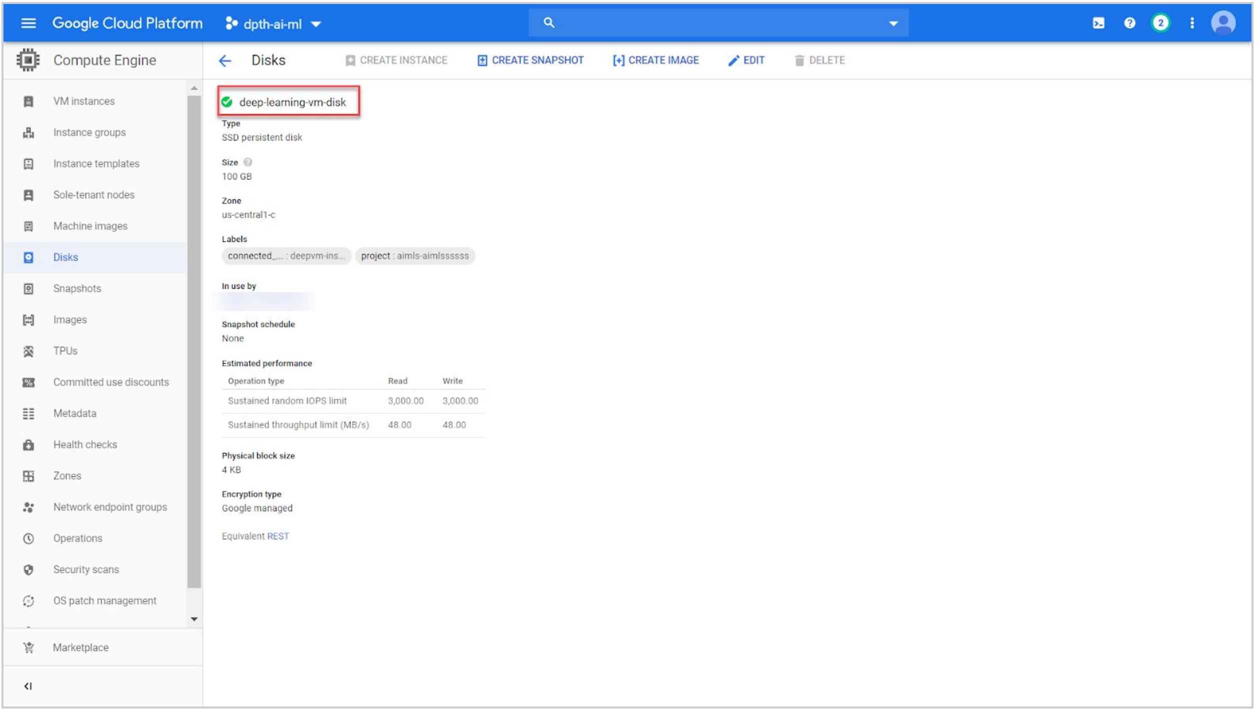
Reservation
Automatically choose (default)

CPU platform
Intel Broadwell

Display device
Turn on a display device if you want to use screen capturing and recording tools.
□ Turn on display device

Figure 40 - VM instance details

Compute Disk: Figure 41 below shows the details of the disk attached to the Deep Learning VM instance.



deep-learning-vm-disk

Type
SSD persistent disk

Size
100 GB

Zone
us-central1-c

Labels
connected_... : deepvm-ins...
project : aimls-aimlssssss

In use by
[redacted]

Snapshot schedule
None

Estimated performance

Operation type	Read	Write
Sustained random IOPS limit	3,000.00	3,000.00
Sustained throughput limit (MB/s)	48.00	48.00

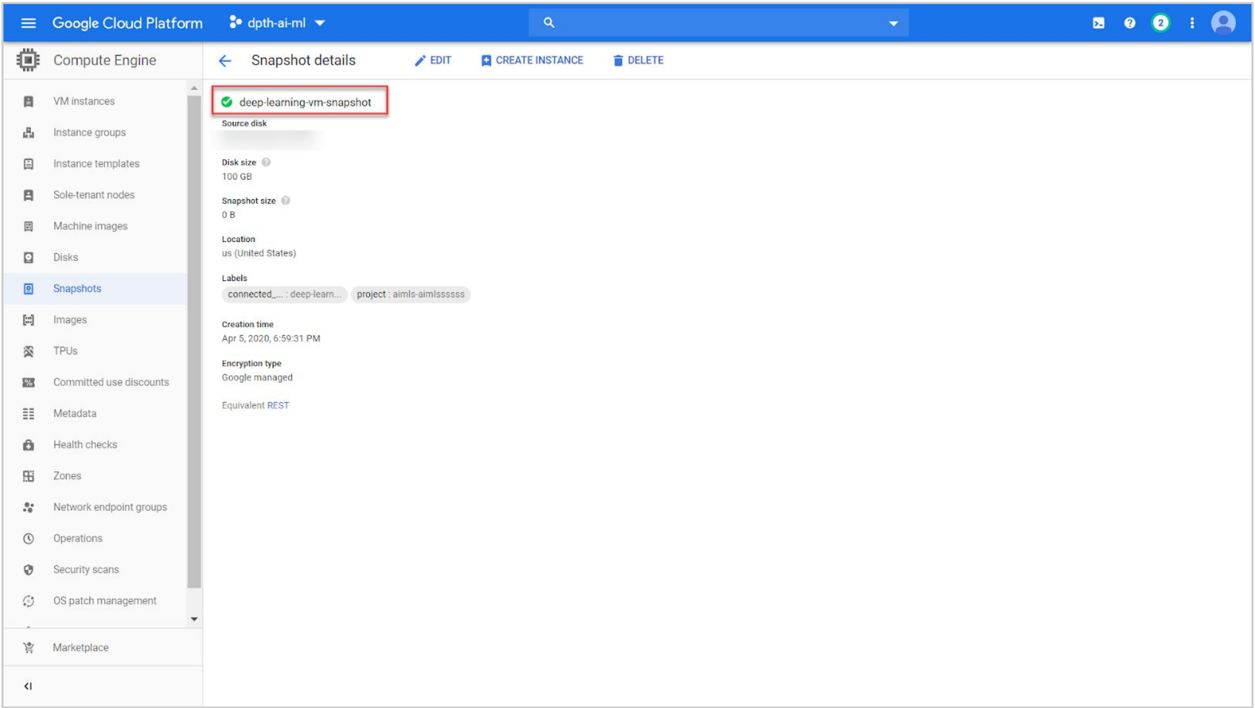
Physical block size
4 KB

Encryption type
Google managed

Equivalent REST

Figure 41 - Compute Disk details

Compute Snapshot: Figure 42 below shows the snapshot details of the compute disk attached to the Deep Learning VM.



deep-learning-vm-snapshot

Source disk
[redacted]

Disk size
100 GB

Snapshot size
0 B

Location
us (United States)

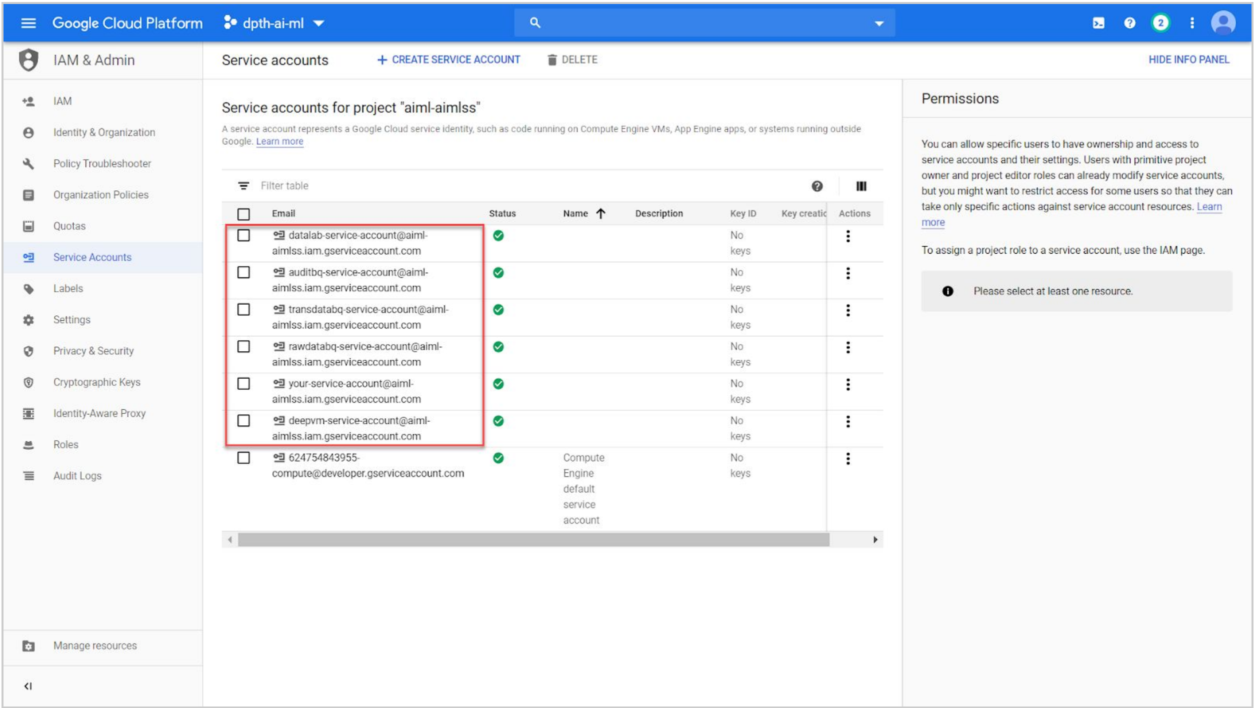
Labels
connected_... : deep-learn... project : aimls-aimlssssss

Creation time
Apr 5, 2020, 6:59:31 PM

Encryption type
Google managed

Figure 42 - Compute Snapshot details

Service Accounts: Figure 43 below shows six custom service accounts for SQL, Datalab, Dataproc and BigQuery datasets.



Service accounts for project "aiml-aimlss"

A service account represents a Google Cloud service identity, such as code running on Compute Engine VMs, App Engine apps, or systems running outside Google. Learn more

□	Email	Status	Name ↑	Description	Key ID	Key creatic	Actions
□	datalab-service-account@aiml-aimlss.iam.gserviceaccount.com	✓			No keys		⋮
□	auditbq-service-account@aiml-aimlss.iam.gserviceaccount.com	✓			No keys		⋮
□	transdatabq-service-account@aiml-aimlss.iam.gserviceaccount.com	✓			No keys		⋮
□	rawdatabq-service-account@aiml-aimlss.iam.gserviceaccount.com	✓			No keys		⋮
□	your-service-account@aiml-aimlss.iam.gserviceaccount.com	✓			No keys		⋮
□	deepvm-service-account@aiml-aimlss.iam.gserviceaccount.com	✓			No keys		⋮
□	624754843955-compute@developer.gserviceaccount.com	✓	Compute Engine default service account		No keys		⋮

Permissions

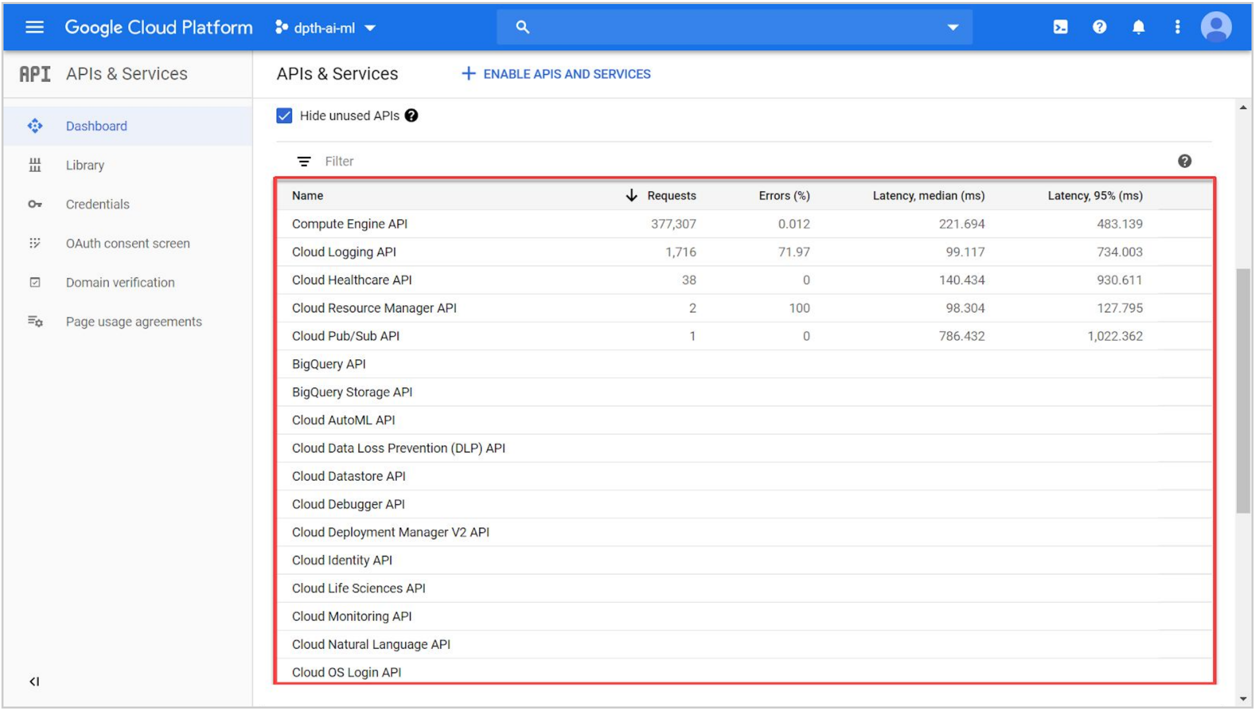
You can allow specific users to have ownership and access to service accounts and their settings. Users with primitive project owner and project editor roles can already modify service accounts, but you might want to restrict access for some users so that they can take only specific actions against service account resources. Learn more

To assign a project role to a service account, use the IAM page.

Please select at least one resource.

Figure 43 - Service Accounts

API & Services Console: Figure 44 below shows all the APIs enabled as a part of the data hosting project.



☑ Hide unused APIs

Name	Requests	Errors (%)	Latency, median (ms)	Latency, 95% (ms)
Compute Engine API	377,307	0.012	221.694	483.139
Cloud Logging API	1,716	71.97	99.117	734.003
Cloud Healthcare API	38	0	140.434	930.611
Cloud Resource Manager API	2	100	98.304	127.795
Cloud Pub/Sub API	1	0	786.432	1,022.362
BigQuery API				
BigQuery Storage API				
Cloud AutoML API				
Cloud Data Loss Prevention (DLP) API				
Cloud Datastore API				
Cloud Debugger API				
Cloud Deployment Manager V2 API				
Cloud Identity API				
Cloud Life Sciences API				
Cloud Monitoring API				
Cloud Natural Language API				
Cloud OS Login API				

Figure 44 - API & Services Console

4. Extended Product Guidance

The section outlines the HIPAA-aligned security configurations for other products and services that are not part of the base Analytics and AI/ML platform deployment, but which can be included based on specific requirements. The instructions for deploying an example project using DPT is explained in Section 3.4.

For the following products and services, Section 4 outlines the default out-of-the-box configurations, including user-configurable settings, that are available.

- Google Cloud Spanner
- Google Cloud Bigtable
- Google Cloud Firewall

Note:

1. *Google Cloud products and services not included in the list above, but which are supported by Terraform can be integrated by adding the respective script(s) within the terraform_deployments section in the Analytics and AI/ML Platform template shown in Section 3. The list of Google Cloud products and services supported by Terraform can be found here.*
2. *Requirements for the creation of customized groups with specific access-level permissions using Cloud IAM and conditional access varies across organizations and can be customized in DPT.*

4.1 Google Cloud Spanner

Google Cloud Spanner is a fully managed, horizontally scalable, and highly available relational database solution offering consistency, scalability, and availability at the Cloud level.

Cloud Spanner can handle high volume real-time data. This further enhances its usability in real-time decision making, along with a very low latency. Cloud Spanner offers global consistency alongside SQL in applications requiring seamless integration across multiple regions.

To learn more about Cloud Spanner and the parameters discussed below, refer to the Cloud Spanner documentation and resource configuration respectively.

HIPAA Alignment for Google Cloud Spanner

Security Rule: Technical Safeguards	*Default Configurations*	*User-Controlled Configurations (ex. via DPT)*
Infrastructure Security	Cloud Spanner is replicated at byte level at the underlying	A node provides up to 2TB of storage and the template here is configured to use two

	distributed file system that it's built on. Besides that, Cloud Spanner uses a concept called splits which replicates the data in the instance and provides additional compute power for each copy of replication. This ensures high availability of data and also the service in the event of a physical or technical incident.	nodes per instance which can be scaled The number of nodes can be configured in the template based on the requirement.
Identity and Access Management	Users in the owners' group of the project are allowed to perform any action allowed by the project's organization policies.	Use G Suite to create users and groups. Additional role-member bindings can be added in the data section of the template as suggested to control access to the spanner instance.

DPT Template Configuration for Google Cloud Spanner

Note: For options for the customizable parameters in the template below, please refer to Terraform documentation for Cloud Spanner Database and Cloud Spanner Instance.The configurable values in the template below are indicative only. Please modify it to match specific requirements in the context of usage.

Template (Please refer to accompanying *.yaml template for detailed configuration)

```
terraform_deployments:
 resources:
 # the bindings and the corresponding roles which the IAM policy refers to are in the data section, and instance, database, and policy are in the resource section.
 # add roles in the data section under binding
  config:
   resource:
   - google_spanner_instance:
    - main:
   # Code Block 4.1.a
     - config: regional-europe-west1
      display_name: main-instance
      num_nodes: '2'
   - google_spanner_database:
    - database:
     - ddl:
      - CREATE TABLE t1 (t1 INT64 NOT NULL,) PRIMARY KEY(t1)
      - CREATE TABLE t2 (t2 INT64 NOT NULL,) PRIMARY KEY(t2)
      instance: ${google_spanner_instance.main.name}
      name: my-database
   # Code Block 4.1.b
   - google_spanner_instance_iam_member:
    - owner:
     - instance: ${google_spanner_instance.main.name}
      role: roles/owner
      member: user:user@domain
     reviewer:
     - instance: ${google_spanner_instance.main.name}
      role: roles/iam.securityReviewer
      member: user:user@domain
```

Infrastructure Security	**Asset Management**	**Refer to code block 4.1.a** **config** - The name of the instance's configuration (similar but not quite the same as a region), which defines the geographic placement and replication of databases in this instance. It determines where the data is stored.
Identity and Access Management	**User access control**	**Refer to code block 4.1.b** **google_spanner_instance_iam_.binding** - Member role for the user. G Suite users/groups and Cloud IAM roles can be used to control access.

4.2 Google Cloud Bigtable

Google Cloud Bigtable is Google Cloud's NoSQL database service. Bigtable can quickly scale to support billions of rows and columns and can store petabytes of data. It supports high read and write throughput at low latency, and it is an ideal data source for MapReduce operations.

Google Cloud encrypts all data stored in the Cloud Bigtable by default, but row-level or cell-level IAM restrictions are not offered by Cloud Bigtable.

To learn more about Cloud Bigtable and the parameters discussed below, refer to the Cloud Bigtable documentation and resource configuration respectively.

HIPAA Alignment for Google Bigtable

Security Rule: Technical Safeguards	*Default Configurations*	*User-Controlled Configurations (ex. via DPT)*

Infrastructure Security	Replication in different zones of the same region typically reduces the replication latency between the clusters.	The template here can be configured to have up to four clusters. Considering the higher replication latency, these clusters can be created in zones spanning across different regions. Also, the template is configured to use 3 nodes per cluster which increases the performance of each cluster.

DPT Template Configuration for Google Cloud Bigtable

Note: For options for the customizable parameters in the template below, please refer to Terraform documentation for [Bigtable Instance]() and [Bigtable Table](). The configurable values in the template below are indicative only. Please modify it to match specific requirements in the context of usage.

Template (Please refer to accompanying *.yaml template for detailed configuration)

```
terraform_deployments:
  resources:
    config:
      resource:
      - google_bigtable_instance:
        - instance:
          - cluster:
              # appropriate id must be chosen based on its purpose
            - cluster_id: tf-instance-cluster
          # Code Block 4.2.a
              num_nodes: 3
              zone: europe-west1-b
              storage_type: HDD
            name: tf-instance
      # Deploying table under bigtable instance
      # This section must be changed as per the data
      - google_bigtable_table:
        - table:
          - instance_name: ${google_bigtable_instance.instance.name}
            name: tf-table
            split_keys:
            - a
            - b
            - c
```

Infrastructure Security	**Disaster Recovery**	**Refer to code block 4.2.a** **zone** - The zone mentioned here is **europe-west1-b,** an European Union region that complies with all the HIPAA policies. **num_nodes** - The number of nodes in the Cloud Bigtable cluster.

4.3 Google Cloud Firewall

Google Cloud Firewall is used to enforce network rules and to control which IP's are allowed to access resources on Google Cloud. Cloud Firewall rules can be used to limit access to SQL and Google Cloud Storage from external networks and to also enforce network segmentation and control internal access within the VPC's and between VPC's.

To learn more about Google Cloud Firewall and the parameters discussed below, refer to the Cloud Firewall documentation and resource configuration respectively.

HIPAA Alignment for Google Cloud Firewall

Security Rule: Technical Safeguards	*Default Configurations*	*User-Controlled Configurations (ex. via DPT)*
Infrastructure Security	A VPC needs to be created for the firewall to be used allowing access through some ports using protocols like TCP and ICMP.	The firewall rules can be customized to block or allow traffic on ports or by protocol. This template here is configured to use a private VPC network and a firewall allowing access through some ports using protocols like TCP and ICMP. The firewall rules can also be applied based on an IP address that belongs to either a source or a destination range, or an IP Address which has a particular tag or even an IP address with a particular.

DPT Template Configuration for Google Cloud Firewall

Note: For options for the customizable parameters in the template below, please refer to Terraform documentation for [Compute Firewall](https://www.terraform.io/docs/providers/google/r/compute_firewall.html). The configurable values in the template below are indicative only. Please modify it to match specific requirements in the context of usage.

Template (Please refer to accompanying *.yaml template for detailed configuration)

```yaml
compute_firewalls:
- name: test-firewall
  network: example-network
  # Code Block 4.3.a
  allow:
  - protocol: icmp
  - protocol: tcp
    ports:
    - '80'
    - '8080'
    - 1000-2000
  source_tags:
  - web
  # For EGRESS traffic, it is NOT supported to specify source_ranges OR source_tags.
  # Code Block 4.3.b
  direction: INGRESS
  # Enable logging for a particular firewall rule. If logging is enabled, logs will be exported to Stackdriver.
  # Code Block 4.3.c
  enable_logging: true
  # Uncomment the following attributes as per the requirements. For more information on the attributes, refer to "https://www.terraform.io/docs/providers/google/r/compute_firewall.html".
  # Code Block 4.3.d
  # destination_ranges:
  # Code Block 4.3.e
  # source_ranges:
  # Code Block 4.3.f
  # source_service_accounts:
  # Code Block 4.3.g
  # source_tags:
```

Infrastructure Security	**Network Security**	**Refer to code block 4.3.a** **allow** - The list of ALLOW rules specified by this firewall. Each tuple of the allow block represents a **PERMITTED** connection. It has 2 fields:

		● **protocol -** The IP Protocol to which this rule will apply. Example - TCP, UDP, ICMP, etc. ● **ports -** List of ports to which this rule applies.
		Refer to code block 4.3.b **direction** - The direction in which traffic flows. Accepted values are INGRESS (incoming) or EGRESS (outgoing). The default value is **INGRESS**.
		Refer to code block 4.3.c **enable_logging** - This field denotes whether to enable logging for a particular firewall rule.
		Refer to code block 4.3.d **destination_ranges** - Ensures that the firewall rule(s) will apply only to traffic that has a destination IP address in these ranges. These ranges must be expressed in CIDR format. Only **IPv4** is supported.
		Refer to code block 4.3.e **source_ranges** - Ensures that the firewall rule(s) will apply only to traffic that has source IP addresses in these ranges. These ranges must be expressed in CIDR format. Only **IPv4** is supported.
		Refer to code block 4.3.f **source_service_accounts** - The firewall rule(s) will apply only to traffic originating from an instance with a service account in this list.
		Refer to code block 4.3.g **source_tags** - The firewall rule(s) will apply only to traffic with source IP that belongs to a tag listed in source tags.

A. Appendix

A.1 HIPAA and Google Cloud

Google Cloud Platform (GCP) enables healthcare organizations to realize the benefits of cloud computing and supports their HIPAA compliance efforts as part of the shared responsibility between a customer and Google. There is no certification recognized by the U.S. Department of Health and Human Services for HIPAA compliance. Google will enter into a Business Associate Agreement (BAA) with customers, while customers continue to be responsible for evaluating their own HIPAA compliance including their use of Google Cloud Services.

Google Cloud customers own their data and control how it is used. It is crucial to remember that enterprises and individuals utilizing Google Cloud are responsible for understanding HIPAA and its implications concerning use of Google Cloud products and services hosting applications and services. Some of these aspects are listed below.

- Applicability of the provisions and requirements of HIPAA across applications, platforms, and Google Cloud Infrastructure
- Classification and inventory of data, particularly protected health information (PHI), along with the business and information systems that process this data
- Alignment of current controls, policies, and processes for managing and protecting healthcare data with HIPAA requirements
- Understanding of the existing data protection features on Google Cloud for meeting HIPAA requirements
- If applicable, review and accept Google Cloud's data processing terms for the G Suite and Cloud Identity Data Processing Amendment (link) and for the Google Cloud Data Processing and Security Terms (link).

Please refer to the following documents carefully before proceeding:

- HIPAA Compliance on GCP: https://cloud.google.com/security/compliance/hipaa
- Google Cloud services in scope for the HIPAA BAA: https://cloud.google.com/security/compliance/hipaa-compliance/
- HIPAA Overview Guide: https://cloud.google.com/files/gcp-hipaa-overview-guide.pdf
- GCP Terms of Service: https://cloud.google.com/terms
- HIPAA Guide for professionals: https://www.hhs.gov/hipaa/for-professionals/index.html
- HSS Guidance on HIPAA and Cloud Computing: https://hhs.gov/hipaa/for-professionals/special-topics/cloud-computing/index.html

A.2 Compliance

Security and privacy on the Cloud is a shared responsibility. Google Cloud is responsible for the security of the cloud infrastructure and services. Google Cloud's customers are responsible for the security, privacy, and compliance of their workloads in the cloud. Google Cloud's focus on data security, privacy, and transparency has provided a foundation towards achieving HIPAA compliance for Google Cloud. In addition, Google Cloud offers data security, data privacy, data portability, and threat protection products and features that can support HIPAA compliance efforts, some of which have been described in this solution guide. These capabilities can be leveraged not only to prevent abuse or unauthorized access to personal data but also to maintain security of data and meet HIPAA requirements.

G Suite and Google Cloud Platform are regularly tested, assessed, and evaluated for the effectiveness of technical and organizational security and privacy measures via third-party audits and certifications, such as those listed below.

1. ISO 27001 for information security management systems
2. ISO 27017 for cloud security controls
3. ISO 27018 for protection of personally identifiable information (PII) in public clouds acting as PII processors
4. SOC 2 and SOC 3 for evaluating systems' and controls' security, availability, processing integrity, and confidentiality or privacy

Google Cloud Platform is also certified under the HITRUST CSF security framework and has attained a CSA Star SOC2+ report.

To learn more about security and compliance for the Google Cloud Platform, and view our comprehensive compliance documentation, refer to Cloud Compliance & Regulations Resources.

A.3 Google Cloud Shared Responsibility Model

The shared responsibility model depends on the particular service model. This starts from the bottom of the stack and moves upwards, from the infrastructure as a service (IaaS) layer where only the hardware, storage, and network are Google's responsibility, up to software as a service (SaaS) where most components of the stack except the content (i.e., data) and access policies are up to the provider.

To learn more about Google Cloud's Shared Responsibility Model, refer to the Google Infrastructure Security Design Overview.

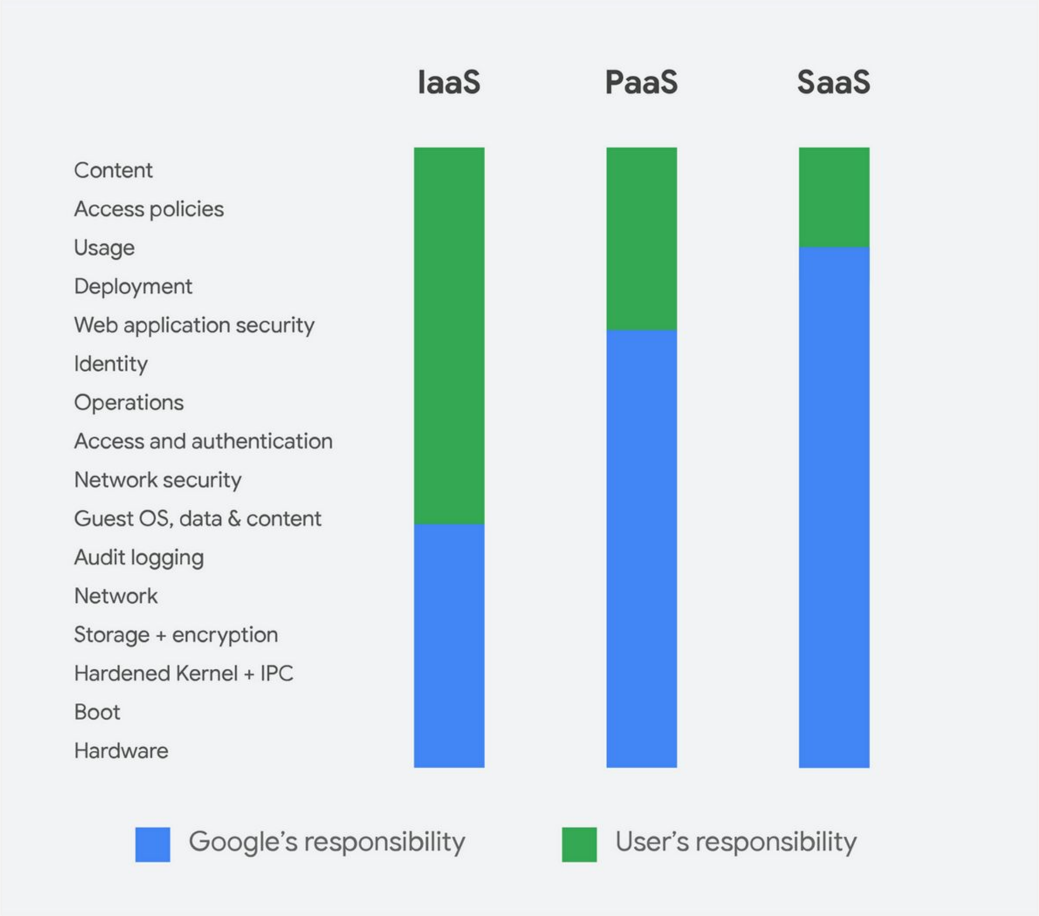


Figure 45 - Google Cloud Shared Responsibility Model

In general, Google is responsible for the security of the underlying infrastructure, including hardware, firmware, kernel, OS, storage, network, and more. This includes encrypting data at rest by default, encrypting data in transit, using custom-designed hardware, laying private network cables, protecting data centers from physical access, and following secure software development practices.

Conversely, customer responsibility for security and compliance in the cloud is listed in Appendix A.4.

A.4 Customer Responsibilities

The following are typical examples of security and compliance capabilities for which the customer is responsible. This is not an exhaustive list.

A.4.1 Identity and Access Management

1. User access provisioning
2. Custom roles to control access
3. Monitoring to detect unusual activity by users and administrators
4. Role-based access controls and separation of duties
5. Multi-factor authentication and 2-step verification for access critical environments and sensitive data
6. Periodic cadence for reviewing access lists

A.4.2 Governance, Risk and Compliance

1. Governance and implementation of organization-specific security policies and standards
2. Definition of security-specific key performance indicators (KPIs) and key risk indicators (KRIs)
3. Security awareness training and secure coding practices training and reporting
4. Background verification checks by authorized parties prior to granting access

A.4.3 Data Security

1. Consent collection, logging, tracking, and monitoring by end-users for organizational access to their PII, PHI, or other sensitive data
2. Data governance lifecycle and data management strategy
3. Data classification, labeling, and handling as per regulatory requirements, service level agreements and operational continuity requirements
4. Policies and procedures for reuse, disposal, and deletion of resources (e.g., like data, equipment, and digital media)
5. Data destruction, obfuscation, and archival standards, including supporting tools and technologies.
6. Key management, if customers choose to use the Customer-Managed Encryption Keys or Customer-Supplied Encryption Keys (CSEK) capabilities of specific products.
7. Communication of incidents and notifications about data breaches, including reporting to regulatory authorities and customers

A.4.4 Infrastructure Security

1. Configuration and validation of firewall rules for both egress and ingress traffic
2. Penetration testing, black-box testing, and red teaming exercises

A.4.5 Security Operations

1. Audit logs for security events, faults, exceptions, and data access violations
2. Enabling of data-read and data-write logs for all critical environments and projects, which are reviewed on a periodic basis
3. Event logs stored and backed up at a centralized storage location and which are protected against tampering and unauthorized access
4. Clocks for all resources are in sync with the approved time sources like the network time protocol (NTP) servers/domain controllers.