C3 AI Installation Guide – Google Cloud Platform

Version 8.3

27 September 2023

Legal Notice

C3.ai products and services are sold subject to the C3.ai terms and conditions agreed at the time of purchase. Except as expressly permitted in that agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means the C3.ai products, services, or documentation.

The information contained herein is subject to change without notice, and is not warranted to be error-free. The information is provided by C3.ai "as-is" for informational purposes only, without representation or warranty of any kind, and C3.ai or its affiliated companies will not be liable for errors or omissions with respect to the information. The only warranties for C3.ai products and services are those that are set forth in the express warranty statements, if any, accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. If you find any errors, please report them to us in writing.

If this software or documentation is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, then use the following notice: "U.S. GOVERNMENT END USERS: C3.ai programs, including any integrated software, any programs installed on any hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government."

C3.ai materials are not intended for use in the operation of nuclear facilities, aircraft navigation or communication systems, air traffic control systems, life support machines or other equipment in which the failure the C3.ai materials could lead to death, personal injury, or severe physical or environmental damage. C3.ai disclaims any and all liability arising out of, or related to, any such use of the C3.ai materials.

Information contained in this document regarding third party product or services does not constitute a license from C3.ai to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party. C3.ai is not responsible for and expressly disclaims all warranties of any kind with respect to third-party content, products, and services. C3.ai is not responsible for any loss, costs, or damages incurred due to the access to or use of third-party content, products, or services, except as set forth in a written agreement between you and C3.ai.

Any software coding samples included in this documentation are examples only and are not intended to be used in a production environment. The code is provided "as-is" and use of any code is at your own risk. C3.ai does not warrant the correctness or completeness of the code given herein, and C3.ai is not liable for errors or damages caused by usage of the code.

The business names used in this documentation are fictitious and are not intended to identify any real companies currently or previously in existence.

C3 AI, C3.ai, and the C3.ai logos are trademarks or registered trademarks of C3.ai, Inc. in the United States and/or other countries. All other product names, trademarks, and registered trademarks are the property of their respective owners.

Table of Contents

Overview

C3 AI Applications support flexible deployment options including C3 AI-managed cloud deployments or customer-managed deployments. The choice of deployment option will have implications on project timelines, service-level agreement (SLA), and RACI requirements.

C3 AI-managed cloud deployment

For C3 AI-managed cloud deployment, C3 AI provides a dedicated project with dedicated compute, storage, and networking resources. All cloud resources are dedicated to your project and will not be shared with any other customer. C3 AI employs industry leading cyber security and access control practices to protect your applications and data.

C3 AI-managed cloud deployments are typically completed within two (2) days of the scheduled deployment start. All the services are hosted in C3 AI's Google Cloud Platform (GCP) cloud account.

C3 AI customer-managed deployment

You can optionally deploy the C3 AI cluster in your own GCP Virtual Private Cloud (VPC), a feature known as customer-managed deployment. You can use a customer-managed deployment to exercise additional control over your network configurations to comply with specific cloud security and governance standards your organization may require.

A GCP VPC allows you to provision a logically isolated section of the GCP Cloud where you can launch GCP resources in a virtual private secure network. The VPC is the network location for your C3 AI clusters.

A deployment start schedule for a customer-managed deployment is dependent on the customer.

C3 AI customer-managed deployment overview

In C3 AI, a cluster is a C3 AI deployment in the cloud that functions as the environment for developing and deploying C3 AI Applications. Your organization can choose to have multiple clusters or just one, depending on your needs.

A customer-managed deployment is a good solution if you have:

- Security policies that prevent Platform-as-a-Service (PaaS) or Software-as-a-Service (SaaS) providers from creating VPCs in your own GCP projects.
- An approval process to create a new VPC, in which the VPC is configured and secured and well-documented by internal information security or cloud engineering teams.

- A team with Terraform expertise and a change management system that are available for on-going management of infrastructure for the C3 AI Cluster.

Benefits include:

- Lower privilege level: You maintain more control of your own GCP project. And you do not need to grant C3 AI as many permissions as you do for a C3 AI-managed deployment. For example, there is no need for permission to create VPCs.
- Maintain more control of your own GCP account and limit outgoing connections.

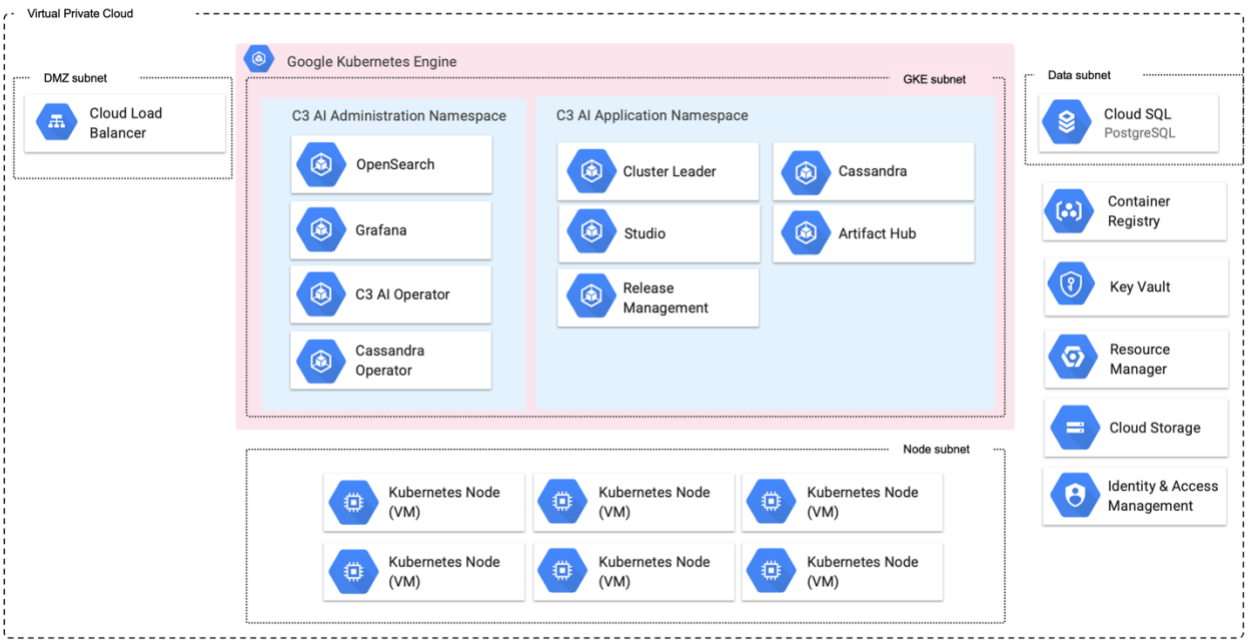

Figure 1. GCP Architecture with C3 AI Platform Deployment

Requirements

The C3 AI Platform requires specific Google Cloud Platform (GCP) cloud services and infrastructure for successful deployment, as well as specific access requirements for C3 AI Operations to install, administer, and upgrade the C3 AI Platform and C3 AI Applications.

The following sections describe the specific services and access needs, including network configurations and subnet requirements, security group egress and ingress rules, and subnet-level access control lists (ACLs).

Required GCP Cloud services

The table below describes the GCP Cloud infrastructure services required by the C3 AI Platform. You are required to provide the services below configured to C3 AI specifications as documented in the HashiCorp Terraform scripts.

Google Cloud Platform Service	Version	Description
Kubernetes Engine (GKE)	1.2.5	Operating environment responsible for the deployment, scaling, and management of the C3 AI Platform and C3 AI Applications.
Cloud Key Management Service	Current version	Scalable, centralized, fast cloud key management.
Cloud SQL for PostgreSQL	11.14.5	Relational data required for internal operations of the C3 AI Platform.
Cloud Storage	Current version	Reliable and secure object storage used for the management of application and platform configuration and other ancillary tasks.
Identity and Access Management (IAM)	Current version	Fine-grained access control and visibility for centrally managing cloud service account resources.
Virtual Private Cloud (VPC)	Current version	Dedicated, isolated network for inter-C3Cluster communication.

Google Cloud access requirements

The table below describes the access requirements for C3 AI Operations to install, administer, and upgrade the C3 AI Applications and C3 AI Platform.

Access Requirements	Description
A dedicated GCP project	When creating the project, C3 AI requires: (1) Subscription identifier, (2) Cloud region.
Secure internet access to the GCP project	Secure, remote access via internet (VPN access is acceptable) to a bastion host from which C3 AI Operations personnel can administer cloud infrastructure and C3 AI services.
A bastion host accessible by C3 AI Operations to manage the cluster	The bastion host will be used by C3 AI Operations to administer the C3 AI Applications and C3 AI Platform. Software utilities required to be installed on the bastion host must include: RedHat 8, Google Cloud Command-Line Interface (gcloud CLI) v347 or greater, kubectl v1.2.5, Helm v3.8.
Access to C3 AI and third-party library and image repositories	Access to C3 AI and third-party repositories for the container images, Python libraries, NodeJS libraries, and runtime billing data collection. If connecting to remote C3 AI, Python, and NodeJS artifact repositories violates security standards, the C3 AI Platform can be configured to connect to local artifact repositories such as GCP Artifact Registry, JFrog, and Anaconda Enterprise.
X.509 certificate for terminating network encryption	A fully qualified domain name for C3 AI Cluster ingress configuration (for example, c3project.customer.com). You are responsible for providing the private and public key (and the certificate chain if necessary) from the x509 certificate to C3 AI. These are placed in a Kubernetes secret and used by C3 AI cluster ingress controller.

Network configuration

To deploy the C3 AI Platform in your own VPC, you must create the VPC following the requirements enumerated in VPC requirements section below.

VPC requirements

Your VPC must meet the requirements described in this section to host a C3 AI cluster.

VPC region

The GCP region where the deployment will occur. Refer to GCP documentation for a list of available regions.

VPC sizing

The C3 AI Platform requires two (2) CIDR blocks.

IP Address Range	Association
10.0.0.0/22	Routable space, used by Cloud SQL (Postgres), EKS Cluster, and node pools.
172.0.0.0/16	Non-routable space, used by EKS pods.

The VPC must have DNS hostnames and DNS resolution enabled.

Subnets

C3 AI must have access to at least two (2) subnets for each cluster, with each subnet in a different availability zone. The subnets are:

- A subnet for the GKE cluster and node pools
- A subnet for Kubernetes pods

Each subnet must have a netmask between /16 and /22.

Subnet route table

The route table for workspace subnets must have quad-zero (0.0.0.0/0) traffic that targets the appropriate network device.

Additional subnet requirements

- Subnets must be private.
- Subnets must have outbound access to the public network using a NAT gateway and internet gateway, or other similar customer-managed appliance infrastructure.
- The NAT gateway must be set up in its own subnet that routes quad-zero (0.0.0.0/0) traffic to an internet gateway or other customer-managed appliance infrastructure.

Security groups

C3 AI must have access to at least one GCP security group and no more than five security groups. You can reuse existing security groups rather than create new ones.

Security groups must have the following rules.

Egress (outbound)

- Allow all TCP and UDP access to the workspace security group (for internal traffic)
- Allow TCP access to 0.0.0.0/0 for these ports:
    - 443: for C3 AI infrastructure, cloud data sources, and library repositories

Ingress (inbound)

- 443: for C3 AI application access
- 22: for SSH access to a bastion host

Subnet-level network ACLs

Subnet-level network ACLs must not deny ingress or egress to any traffic.

- ALLOW ALL from Source 0.0.0.0/0. This rule must be prioritized.
- Egress:
    - Allow all traffic to the C3 AI cluster VPC CIDR, for internal traffic.
    - Allow TCP access to 0.0.0.0/0 for these ports:
        - 443: for C3 AI infrastructure, cloud data sources, and library repositories.

HashiCorp Terraform Configuration

HashiCorp Terraform is a popular open-source tool for creating safe and predictable cloud infrastructure across several cloud providers. Terraform scripts are used to create the cloud infrastructure required by the C3 AI Platform and automate the deployment of the C3 AI Platform in your Google Cloud Platform (GCP).

NOTE: For C3 AI customer-managed deployments, any customization performed on the Terraform scripts must be reapplied with each version of the Terraform scripts from C3 AI.

Getting started

In this section, you install and configure requirements to use Terraform. You then configure Terraform authentication. Following completion of this section, you go to "Installation Steps" section below to deploy and configure the cloud infrastructure required by the C3 AI Platform.

Requirements

To use Terraform to create cloud infrastructure resources required by the C3 AI Platform in your GCP account, you must have the following:

- A Google Cloud account.
- A Google Cloud project in the account. The GCP project name must adhere to the following restrictions: name must include not include a hyphen and must start with a letter; only lowercase letters are allowed with no other special characters or diacritics (accented letters); and, should be less than 18 characters in length.
- On your local development machine, you must have:
    - The HashiCorp Terraform CLI. See Install Terraform on the Terraform website to download the binary of the required Terraform version specified in the `main.tf` file example in the "Installation Steps" section below. Select AMD64 or ARM64 depending on the which matches the client hardware from which you will run the Terraform scripts.
    - The gcloud CLI. See Install the gcloud CLI on the Google Cloud website.
    - The gcloud CLI, signed in through the gcloud auth application-default login command to obtain user access credentials via a web flow and put them in the well-known location for Application Default Credentials (ADC).

```
gcloud --project="<project-id>" auth application-default login
```

NOTE: Replace:

- <project-id> with the specific Google Cloud project ID to use for this deployment. If omitted, then the current project is assumed.

For more details, see Installing Google Cloud SDK and Authorize the gcloud CLI on the Google Cloud website.

Installation Steps

Installation of the C3 AI Platform on the Google Cloud Platform (GCP) is a multi-step process due to limitations of HashiCorp Terraform and GCP-specific configuration requirements. The installation process is the following:

1. Create the VPC and required GCP services.
2. Validate the configuration of the VPC and required GCP services and provide C3 AI Operations access to the cluster.
3. C3 AI Operations completes the installation of the C3 AI Platform.

To create a VPC, C3 AI requires the use of HashiCorp Terraform and will provide a set of Terraform scripts to assist you in the creation of the VPC and required GCP Services.

NOTE: See the "HashiCorp Terraform Requirements" section above to ensure all requirements are met prior to completing the installation steps below.

A description of the Terraform modules is below.

Terraform Module	Description
bootstrap	Configures the necessary Identity and Access Management (IAM) roles and policies to allow a Terraform orchestrator to deploy all services required for C3 AI Platform on GCP.
c3cluster	Coordinates the execution of all other Terraform modules.
gke-cluster	Configures GCP Kubernetes Engine (GKE), including VPC configuration, endpoint access, authorized IP addresses, and the version of Kubernetes used by the cluster.
gke-nodepool	Configures the GKE node groups, including default instance size, required subnet, and permissions assigned to each node.
firewall	Configures ingress and egress security rules.
iam	Configures the required IAM roles and policies.
kms	Configures the GCP Key Management service.
network	Configures the VPC, including public and private subnets, internet gateway, CIDR blocks, DHCP, and NAT.
postgres	Creates an GCP Cloud SQL database (PostgreSQL) and assigns the database to the database subnet.

Terraform Module	Description
gke-sa	Configures the workflow identity to be used by the C3 AI cluster.
gcs	This module configures the GCP cloud storage resources to be used with the C3 AI cluster.
v8	This module configures the GCP project, availability zones, IAM roles, and storge accounts to be used by the C3 AI cluster.

In addition to the required tools listed in the “HashiCorp Terraform Requirements” section, install TFSwitch, which is a tool used to switch easily between Terraform versions.

See Install TFSwitch and TFSwitch Quick Start on the TFSwitch website for more information.

1. Create the VPC and required GCP services

This guide shows you how to create the cloud infrastructure services required by the C3 AI Platform using HashiCorp Terraform on GCP.

1.1 Run the bootstrap module

This module creates the necessary IAM roles and policies to configure the VPC and required GCP services. Configure a new `main.tf` file below, replacing the CAPITALIZED variable names with your values.

NOTE: The project name must adhere to the following restrictions: name must include not include a hyphen and must start with a letter; only lowercase letters are allowed with no other special characters or diacritics (accented letters); and, should be less than 18 characters in length.

```
module "bootstrap" {

    ##Link will be provided by C3
    source = "https://<c3_url>/gcp-bootstrap-x.y.z.zip"

    project_id = "YOUR_PROJECT_ID"

    ##Note your project cannot contain hyphens
    project_name = "YOUR_PROJECT_NAME"
}

provider "google" {}

terraform {
    required_version = "= 1.4.6"
    required_providers {
```

```
        google = {
           source = "hashicorp/google"
           version = "4.34.0"
         }
         google-beta = {
           source = "hashicorp/google-beta"
           version = "4.34.0"
         }
    }
}
```

1.2 After configuring the main.tf file, run the following Terraform commands from the same directory

```
tfswitch
terraform init
terraform plan --out out.plan
terraform apply "out.plan"
```

NOTE: If you receive a "Command not found: Terraform" after running the commands above, the `terraform` binary might not be in your path. See the Get Started in GCP – Install Terraform page on the HashiCorp Terraform website for more information.

1.3 Run the c3cluster module

This module coordinates execution of all other Terraform modules. Configure a new `main.tf` in a separate directory from the bootstrap module, replacing the CAPITALIZED variable names with your values. Note that you must assume the role created by bootstrap module.

Contact your account manager for the list of IP addresses required by C3 AI. These values will be used to update the `ip_allowlist` section below.

NOTE: The project name must adhere to the following restrictions: name must include not include a hyphen and must start with a letter; only lowercase letters are allowed with no other special characters or diacritics (accented letters); and, should be less than 18 characters in length.

```
module "c3cluster" {

  ##Link will be provided by C3
  source = "https://<c3_url>/gcp-c3cluster-x.y.z.zip"
  c3_region = "YOUR_REGION"

  project_id = "YOUR_PROJECT_ID"
  project_name = "YOUR_PROJECT_NAME"
```

```
  ip_allowlist = [
    {
      "cidr_block" : "X.X.X.X/32",
      "display_name" : "C3 Control IP"
    },
    {
      "cidr_block" : "X.X.X.X/32",
      "display_name" : "C3 Control IP"
    }
  ]

  gke_node_pools = {
    "c3": {
      "autoscaling": {
          "max_node_count": 9,
          "min_node_count": 1
      },
      "local_ssd_count": 0,
      "machine_type": "n2-standard-8",
      "node_count": 1
    }
  }
}

provider "google" {}

terraform {
    required_version = "= 1.4.6"

    required_providers {
       google = {
           source = "hashicorp/google"
           version = "4.34.0"
        }

       google-beta = {
          source = "hashicorp/google-beta"
          version = "4.34.0"
        }
    }
}
```

1.4 After configuring the main.tf file, run the example below from the same directory as the new main.tf file

```
tfswitch
terraform init
terraform plan --out out.plan
terraform apply out.plan
```

2. Validate the configuration of the VPC and required GCP services and provide C3 AI Operations access to the cluster

After the VPC and required cloud services are configured, you are required to execute the C3 AI Cluster Validation Utility and provide the results to C3 AI. If all checks performed by the C3 AI Cluster Validation Utility pass, the VPC is suitable for C3 AI Operations to deploy the C3 AI Platform on the Kubernetes cluster.

NOTE: If the cluster validation utility fails, you must remediate all exceptions. All checks must pass for C3 AI Operations to be able to deploy the C3 AI Platform on your Kubernetes cluster. See the next section for details.

Once the checks are successfully completed, provide C3 AI Operations access to the cluster. Refer to the subsequent section for more information.

2.1 Run the C3 AI Cluster Validation Utility and provide results to C3 AI Operations

To run the C3 AI Cluster Validation Utility, do the following:

1. Download and unzip the c3prereqs module from `https://<c3_url>/c3prereqs-x.y.z.zip`
2. Edit the `config.env` to only include the Kubernetes cluster, removing the cloud deployment types that are not applicable.

```
#KUBECONFIG=""

# Minimum helm version required
MINIMUM_HELM_VERSION="3.12"

# Minimum kubernetes version required. Used for client and server
MINIMUM_KUBERNETES_VERSION="1.25"

# Timeout for the c3prereqs job
JOB_TIMEOUT="300s"

## Select Cloud to gcp or aws or azure ; uncomment appropriate cloud-
specific sections below
C3_CLOUD=""

C3_CLUSTER_NAME=""

## AWS SPECIFIC
# C3_AWS_ACCOUNT=""
# C3_AWS_SERVICE_ACCOUNT="c3"

## Set to true if cloud is AWS
S3_BUCKET_NAME=""
```

```
## GCP SPECIFIC
# C3_GCP_PROJECT_ID=
# C3_GCP_SERVICE_ACCOUNT="c3"


## AZURE SPECIFIC
# C3_AZURE_CLIENT_ID=


# Jfrog auth info
C3_REGISTRY_URL="registry.c3.ai"
C3_REGISTRY_USER=""
C3_REGISTRY_PASS=""

# Postgres validation
POSTGRES_USERNAME=""
POSTGRES_PASSWORD=""
POSTGRES_HOST=""
POSTGRES_PORT=
POSTGRES_DATABASE=""
```

3. Run the c3prereqs script. The script will record the results in the reports folder.
4. Review the report.

If the report indicates the VPC is ready for C3 AI Operations to deploy the C3 AI Platform on the Kubernetes cluster, provide the report to C3 AI Operations.

If the report fails, remediate all exceptions and rerun the C3 AI Cluster Validation Utility.

2.2 Provide C3 AI Operations access to the cluster

In addition to the output of the validation utility, you must provide C3 AI Operations with the following.

Title	Description
C3 AI Operations credentials	Credentials for C3 AI Operations team members.
GKE cluster name	The name of the EKS cluster where the C3 AI Platform will be installed. **NOTE:** The cluster name must adhere to the following restrictions: name must include not include a hyphen and must start with a letter; only lowercase letters are allowed with no other special characters or diacritics (accented letters); and, should be less than 18 characters in length.
Region	The GCP region of the GKE cluster.

Title	Description
Cloud SQL Postgres endpoint	Endpoint value for the instance. From GCP Console, select your project, go to SQL and look for the SQL resource <C3_CLUSTERNAME>-pg-shared. Get the "Private IP address" and share it with C3 AI Operations.
Cloud SQL Postgres credentials	Credentials required for the C3 AI Platform to connect to PostgreSQL. From GCP Console, select your project, go to SQL and look for SQL resource <C3_CLUSTERNAME>-pg-shared. Then, click Users and click three (3) dots (…) next to Postgres user and change the password. Share this with C3 AI Operations securely.
Service Account Names	From GCP Console, go to the desired GCP project, navigate to Cloud IAM dashboard, then select Service Accounts. Get the principal of the service account starting with c3aiops and c3server and share it with C3 AI Operations.
GCP Bucket Name	Within the created project in the GCP Console, go to the Cloud Storage Buckets page. There should only be one bucket listed.
Domain name	A fully qualified domain name for C3 AI cluster ingress configuration (for example, c3project.customer.com).
Public and private key	The private and public key (and the certificate chain if necessary) from the x509 certificate. This will be required for ingress configuration.

It is recommended that the sharing of Postgres credential and certificates occur using GCP Vault.

To grant C3 AI Operations GKE cluster administration permissions to the GCP project, run the following code snippet.

```
gcloud auth login

gcloud projects add-iam-policy-binding PROJECT_ID \
    --member=user:USER \
    --role=roles/container.admin
```

NOTE: Replace:

- PROJECT_ID with the ID of the project or fully qualified identifier for the project
- USER with the principal to add the binding for. Should be of the form user:email.

3. C3 AI Operations completes the installation of the C3 AI Platform

With the infrastructure properly configured and EKS node pool configuration updated, C3 AI Operations will continue with the installation of the C3 AI Platform.

At the conclusion of the VPC creation and the deployment of the C3 AI Platform, the Google Cloud Platform (GCP) environment will resemble the following.

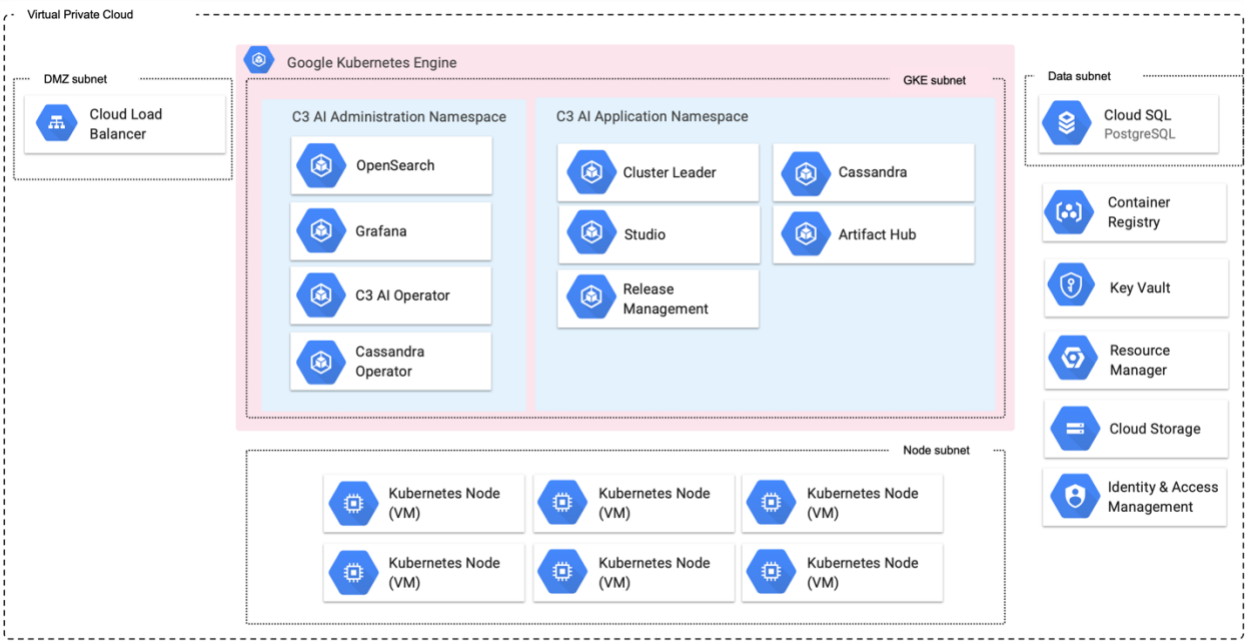

Figure 2. GCP Architecture with C3 AI Platform Deployment