# Introduction

The MoneyGram team is aiming to automate customer validation and transaction release processes, which involve verification from data sources. Quantiphi will build a GCP environment in which users can upload various document types via an API and receive a classification or entity extraction response for the specific document type. This project aims at developing an end-to-end pipeline for document classification and entity extraction.

# Resource Management

This section covers how the MGO organization should be structured on the Google Cloud Platform. The standard naming convention for the GCP project and its resources. How to manage the billing account, monitor all your GCP charges in one place, set alerts for the defined budgets, and how to export your bills to BigQuery for detailed reporting and visual representation.

## 2.1 Resource Hierarchy

Description

GCP resources are organized hierarchically. This hierarchy allows MGO to map its organization's operational structure to GCP, and to manage access control and permissions for groups of related resources.

| **Organization** | An **Organization** is the root node of the Google Cloud Platform hierarchy of resources. All resources that belong to an organization are grouped under the Organization node, which provides insight into and access control over every resource in the organization.  An organization is associated with exactly one [Domain](https://cloud.google.com/billing/docs/onboarding-checklist#domain_and_identity). |
| --- | --- |
| **Domains** | The **Domain** allows you to manage the users in the MGO organization. An organization can have one primary domain and 599 non-primary (secondary domain).  Regardless of the secondary domain, users will still have the possibility to collaborate on the resources of the organization |
| **Folder** | **Folders** are resources in the organization hierarchy used to logically group projects by business needs. They are used as targets to set permission boundaries and support inheritance.  One benefit of using folders is that MGO can enforce different Cloud IAM policies on each folder level. |
| **Project** | At the bottom of the hierarchy, **Project**. Projects contain the service-level resources (such as computing, storage, and networking resources) that process your workloads and constitute your apps. |
| **Resource Labels** | **Labels** are key/value pairs that provide a way to group together with resources that are related or associated with each other. Leverage labels to easily search for resources that belong to each operating unit when necessary. |

## 2.2 Billing

Cloud Billing accounts are linked to one or more GCP projects and are used to pay for the resources you use, such as compute, networking, and storage. Cloud Identity and Access Management (Cloud IAM) roles control access to Cloud Billing accounts.

MGO currently owns and manages the billing account for the ‘mgo.com’ domain/organization.

* Additional Information

<https://cloud.google.com/billing/docs/concepts>

## 2.3 Naming conventions

Having a consistent naming practice is an important discipline. This ensures that administrators and end-users understand the purpose of each resource, where it is located, and how it is differentiated from other resources.

Recommendations

| **Token** | **Length (chars)** | **Details** | **Examples** |
| --- | --- | --- | --- |
| **Business unit prefix** | 3-5 | Defines the name of the business operating unit | **MGO:** mgo |
| **Environments** | 1-5 | The environment where the resource is deployed | **Development**: dev  **Staging**: stage  **User Acceptance Testing**: uat  **Production**: prod |
| **Initiatives** | 2-10 | Name of the initiatives (as an abbreviation) associated with operating unit | **Document Analysis:** da  **Shared Services:** ss |
| **Resource** | 1-5 | An abbreviation for the resource type | **Google Cloud Storage:** gcs **Bigquery:** bq  **Google Kubernetes Engine:** gke |
| **Location** | 1-5 | An abbreviation of where the resource is located | **Zone**: us-east4-b -> use4b  **Region**: us-east4 -> use4  **Global**: g |
| **Description** | 10 | [Optional] Description of the resource to be deployed | Cluster, Dataset, etc |

| **Services** | **Naming Convention** | |
| --- | --- | --- |
| **Quantiphi's Understanding** | **MoneyGram's Response** |
| Cloud Storage | Syntax : <ou>-<env>-<init>-<resource>-<location>-<description>  mgo-np-da-gcs-use4-<description>  *eg. mgo-np-da-gcs-use4-labelled-documents* |  |
| Service Account | sa-<mg>-<env>-<resourcedescription>@<projectid>.iam.gserviceaccount.com  *eg. sa-mg-dev-ai-notebook*  *@mgi-edw-dev.iam.gserviceaccount.com* |  |
| Pub/Sub Topics  &  Cloud Run | <ou>-<env>-<init>-<resource>-<location>-<description>  quantiphi-np-dp-topic-g-<description>  *eg. mg-dev-cloudrun-job-preprocessing* |  |
| BigQuery | <ou>-<env>-<init>-<resource>-<location>-<description>  *eg. mgo-np-da-bq-ml-datasetname* |  |
| Artifact Registry | <ou>-<env>-<init>-<resource>-<location>-<description>  e.g mgo-np-da-ar-use4-repository |  |
| GKE | <ou>-<env>-<init>-<resource>-<location>-<description>  e.g mgo-np-da-gke-use4-docai |  |

# GCP Components and Explanations

This solution will comprise a number of GCP components. This section will highlight the major components and describe those components in detail. For each component links to GCP, documentation are included where more information can be found about each component. Lastly, for platform certifications relating to security, data handling, and privacy please find details here:

<https://cloud.google.com/security/compliance/financial-services>

This table includes a broad summary of the resources leveraged on this project and additional details can be found in Section

| **Type** | **Resource** | **Purpose in this application** |
| --- | --- | --- |
| Identity & Access Management | IAM | To ensure principles of ‘least privilege’ are followed and only the required accesses are granted to team members working on DocAI project  [Link](https://cloud.google.com/iam/docs/concepts) |
|  | MFA | Quantiphi users have MFA as part of their Google account logins  [Link](https://cloud.google.com/identity/solutions/enforce-mfa) |
|  | SSO | MGO users to have SSO  AD and SSO - ADFS are on for Quantiphi team  [Link](https://cloud.google.com/identity/solutions/enable-sso) |
| Networking | Cloud VPN | To establish a secure connection between MGO on-prem environment and GCP  [Link](https://cloud.google.com/network-connectivity/docs/vpn/concepts/overview) |
| Virtual Private Cloud | To provide secure connectivity among cloud resources  [Link](https://cloud.google.com/vpc/docs/vpc) |
| Cloud Internal Load Balancer (HTTPS) | To improve availability and scalability of our application by intelligently distributing work-loads to compute resources based on incoming traffic and location of requests  [Link](https://cloud.google.com/load-balancing/docs/l7-internal/setting-up-l7-internal) |
| Cloud DNS | Cloud DNS is a high-performance, resilient, global Domain Name System (DNS) service that publishes your domain names to the global DNS in a cost-effective way  [Link](https://cloud.google.com/dns/docs/overview) |
|  | Pub/Sub | To provide eventing architecture and to make for more efficient scaling in response to increased load  [Link](https://cloud.google.com/pubsub/docs/overview) |
| Security | Google Security Scanning | To proactively identify vulnerabilities in any of the resources used on GCP for this project |
| Data Encryption | By default data is encrypted at Rest and in Transit  [Link](https://services.google.com/fh/files/misc/security_whitepapers_march2018.pdf) |
| Logging | Logging - Cloud Operations | To provide cloud monitoring, log management and analysis, error reporting, and application performance management  [Link](https://cloud.google.com/logging/docs) |
| API | Google Kubernetes Engine (GKE) | To host model inference scripts and to automate the deployment, scaling, and operations for container clusters  [Link](https://cloud.google.com/kubernetes-engine/docs/concepts/kubernetes-engine-overview) |
| Google Artifact Registry | To store manage and maintain Docker container images [Link](https://cloud.google.com/artifact-registry/docs/integrate-gke) |
| Storage/Database | Firestore | Firestore will be used to maintain transaction states of the Application and extracted data from the Doc AI API endpoints  [Link](https://cloud.google.com/firestore/docs) |
| BigQuery | BigQuery will be used to store logging and transaction information for the API and additional analytics data as per MGO business needs - such that they are accessible with SQL queries.  [Link](https://cloud.google.com/bigquery/docs) |
| Cloud Storage | To store objects passively (e.g. currently being used to store documents for model training)  [Link](https://cloud.google.com/storage/docs/json_api/v1/buckets) |
| AI/ML/Document Understanding AI | Form Parser API (Doc AI) | To extract all textual entities present in the form of a Key-Value pair or a Table (with bounding boxes) from input PDF documents  [LInk](https://cloud.google.com/document-ai/docs) |
| Cloud Vision API (Doc AI) | To extract all textual data present in a document (Optical Character Recognition) which will be used by the document classification model  [Link](https://cloud.google.com/vision/docs) |

Note on Data Privacy: Google may only access data in MGO’s account in strict compliance with Google’s Privacy Policy and the Customer Agreement with MGO. Google also offers a detailed Data Processing Amendment that further describes Google's commitment to protecting your data. For purposes of providing technical support, an administrator from your domain may choose to grant the Google Support team permission to access accounts in order to resolve a specified issue. Further details can be found on Google’s Privacy Policy and data processing and security terms.

Users will only be able to access the resources and data within MGO’s org if MGO grants access. According to GCP Access Transparency, data access can only be done when using GCP managed encryption, and is only done to support the customer if requested. If requested, this access is approved by the customer and logged in Access Transparency.