

## Trabalho final

Data limite para entrega do trabalho: **18 de junho 2020 às 23:59**

Para além do desenvolvimento que cada grupo realizará autonomamente, haverá aulas específicas, predominantemente nas aulas de 3 horas, para apoio à realização do trabalho.

### Componentes a entregar:

- 1) Ficheiro ZIP com as componentes desenvolvidas, incluindo ficheiros README com informações sobre configurações, pressupostos de execução, teste ou outros.
- 2) Documento em formato PDF com descrição da solução: Diagramas de arquitetura, contratos e pressupostos entre as partes envolvidas, formatos de dados e mensagens envolvidos nas interações, bem como os aspectos relevantes da implementação e eventuais pontos de falha.

**Objetivos:** Saber planear e realizar um sistema para submissão e execução de tarefas de computação na nuvem, com requisitos de elasticidade, utilizando de forma integrada serviços da Google Cloud Platform para armazenamento, comunicação e computação, nomeadamente, Cloud Storage, Firestore, Pub/Sub, Compute Engine e Cloud Functions.

**Descrição:** Desenvolva um sistema, designado *CNV2021TF*, para detetar características (*labels*), como por exemplo, *tree*, *street*, *night*, *cat*, *fish*, em ficheiros de imagem (JPG, PNG, etc.) e traduzir essas características de inglês para português. O sistema deve ter elasticidade, aumentando ou diminuindo a sua capacidade de processamento de imagens. As funcionalidades do sistema estão disponíveis para aplicações cliente através de uma interface gRPC com as seguintes operações:

- Submissão de um ficheiro imagem para deteção de características e respectiva tradução de texto. Esta operação recebe o conteúdo de um ficheiro em *stream* de blocos, guardando o mesmo no serviço Cloud Storage. No final, a operação retorna um identificador do pedido (por exemplo, uma composição única entre o nome do bucket e o blob) que será usado posteriormente para obter as características detectadas e a respetiva tradução.
- A partir de um identificador retornado na chamada à operação anterior, obter a lista de todas as características encontradas na imagem e respetivas traduções para português;
- Obter todos os nomes de ficheiros armazenados no sistema entre duas datas e que contêm uma determinada característica;
- [opcional] Outras operações que considere relevantes (por exemplo, *download* das imagens que foram encontradas na operação anterior).

As operações de submissão são disponibilizadas através de um servidor gRPC, o qual funciona como a fachada do sistema, isto é, a aplicação cliente não conhece nada sobre a plataforma GCP.

Para aumentar a disponibilidade e balanceamento de carga do sistema, devem existir várias instâncias do servidor gRPC, cada uma a executar-se numa VM de um *instance group*. Para obter o endereço IP do servidor (*lookup*), o cliente acede por HTTP a um URL pré-definido para obter a

lista de endereços IP das instâncias de servidores gRPC disponíveis, e escolhe um endereço IP aleatoriamente.

A arquitetura do sistema *CNV2021TF* usa vários serviços GCP:

- O serviço Cloud Storage armazena as imagens a processar;
- O serviço Firestore guarda a informação relevante sobre processamento de um ficheiro, nomeadamente o identificador do pedido, data do processamento, metadados do blob no storage, as características detetadas nas imagens e traduções, ou outros que achar convenientes;
- O serviço Pub/Sub é usado para troca desacoplada de mensagens entre os componentes do sistema;
- O serviço Compute Engine aloja máquinas virtuais onde se executam os servidores de gRPC e as aplicações (*labels app*) de deteção de características de imagens;
- O serviço Cloud Functions é usado para implementar duas *cloud function*, uma para *lookup* dos endereços dos servidores gRPC (*trigger* HTTP) e outra para tradução de texto de inglês para português (*trigger* Pub/Sub).

A visão geral do sistema, e as diferentes interações entre os seus componentes, é apresentada na Figura 1.

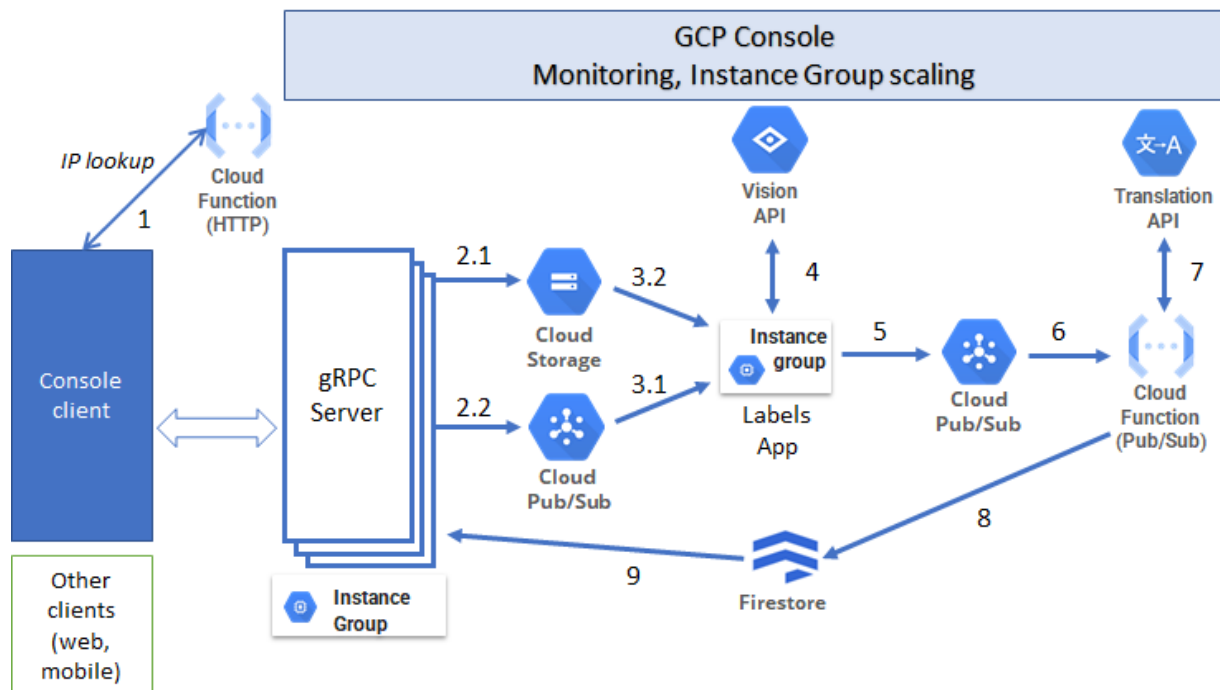


Figura 1: Componentes do *CNV2021TF* e respetivas interações

Para garantir disponibilidade e elasticidade devem ser usados os seguintes requisitos:

- O serviço *IP lookup*, usado pela aplicação cliente (1) para obtenção do endereços IP dos servidores gRPC, deve ser desenvolvido e alojado numa Cloud Function. O cliente escolhe um IP aleatoriamente e, em caso de falha de ligação ao servidor gRPC através do IP escolhido, tenta outro IP ou repete o processo de *lookup* para atualizar a lista dos IP e estabelece uma nova ligação;
- Devem existir em execução várias instâncias do servidor gRPC e do programa para deteção de características (*worker*), alojadas em VM no serviço Compute Engine. Para que estes dois conjuntos de VM possam ter um número diferente de instâncias, os servidores gRPC e os workers são alocados em *instance groups* diferentes. Através da consola GCP, um utilizador gestor do sistema poderá monitorizar o sistema e aumentar ou diminuir o número de instâncias de cada *instance group*, não tendo de haver qualquer outra alteração para o sistema se manter operacional;
- A operação de tradução é realizada numa função Cloud Function associada a um tópico do serviço Pub/Sub, sendo a elasticidade automaticamente garantida pelo GCP.

Fluxo de operações:

Tendo em conta os números apresentados na Figura 1, a lista seguintes descreve cada uma das funcionalidades:

- Após a submissão de uma imagem, a mesma é guardada no Cloud Storage (2.1) e é retornado ao cliente gRPC um identificador único para posteriormente ser possível pedir as características e a respetiva tradução. De seguida, é enviado para um tópico Pub/Sub com nome *topicworkers* o identificador do pedido, o nome do bucket e do *blob* para processamento de obtenção de *labels* (2.2);
- Associado ao tópico *topicworkers* está uma subscrição que é consumida por vários *workers* (*work-queue pattern*). Um *worker* de análise de imagem recebe o nome da imagem a processar (3.1) e obtém o seu conteúdo do Cloud Storage (3.2), interagindo depois com o serviço Vision (4);
- Após o processamento é enviada uma notificação para um tópico Pub/Sub (5) com as características de que se pretende obter tradução e o identificador do pedido. A *Cloud Function* de tradução reage a este evento (6) e interage com o serviço Translation (7), guardando o resultado no Firestore (8);
- Com base no descritor do pedido, a qualquer momento, as aplicações cliente podem pedir ao servidor gRPC informações sobre os ficheiros submetidos, nomeadamente as características e respetivas traduções encontradas nas imagens. Para retornar essa informação o servidor gRPC consulta a base de dados Firestore (9).

Aspetos de implementação:

- A API de visão faz deteção de características sobre imagens, retornando também para cada característica uma pontuação de confiança (entre 0 e 1).  
(<https://cloud.google.com/vision/docs/labels>)
- A API de tradução recebe, na forma de *strings*, um texto para traduzir, a língua desse texto (podendo ser também detetada automaticamente), e a língua do texto de destino. As línguas têm siglas pré-definidas, por exemplo, “pt”, “en”, “es” ou “it”.  
(<https://github.com/googleapis/java-translate>)

CrITÉrios de avaliação do trabalho:

- ❖ 30% - qualidade do relatório, que permita a um leitor entender claramente a arquitetura e as decisões de interação entre as partes, evitando apresentar código, exceto se o mesmo ajudar a explicar detalhes relevantes. O relatório deve indicar os pressupostos assumidos, indicando eventuais comparações com outras decisões possíveis. Deve constar no relatório qual a(s) parte(s) onde cada elemento do grupo teve mais ou menos responsabilidade.
- ❖ 60% - Operacionalidade, simplicidade e flexibilidade das soluções, nomeadamente na configuração e utilização da solução;
  - Nesta avaliação será ponderado o resultado da apresentação da funcionalidade da solução a toda a turma nas aulas da última semana de aulas. Para tal, será posteriormente estabelecido para cada grupo um calendário de apresentação, bem como um guião dos aspetos principais a demonstrar.
- ❖ 10% - participação individual de cada elemento do grupo durante as aulas afetas à realização do trabalho, bem como na apresentação do trabalho à turma.

*José Simão*

*Luís Assunção*