
Computação na nuvem

ISEL – LEIRT / LEIC / LEIM

- Serviço *Google Cloud Storage*
- Controlo de acessos e chamadas via API Java

José Simão jsimao@cc.isel.ipl.pt ; jose.simao@isel.pt

Luís Assunção lass@isel.ipl.pt ; luis.assuncao@isel.pt

Fernanda Passos fpassos@isel.ipl.pt ; fernanda.passos@isel.pt

Sumário

- Armazenamento distribuído de dados em ficheiros
 - Um pouco de evolução histórica
- Requisitos de escalabilidade, disponibilidade
 - Dimensões *BigData*
 - *Google File System*
- Armazenamento Flat com agrupamentos (*buckets*) de objetos
- Serviço *Google Cloud Storage* (GCS)
 - Classes de armazenamento no GCS
 - Criação e controlo de acesso a *buckets*
 - Espaço de nomes dos *buckets*
 - Metadados dos objetos
 - Consistência, boas práticas e limites

Armazenamento de objectos binários

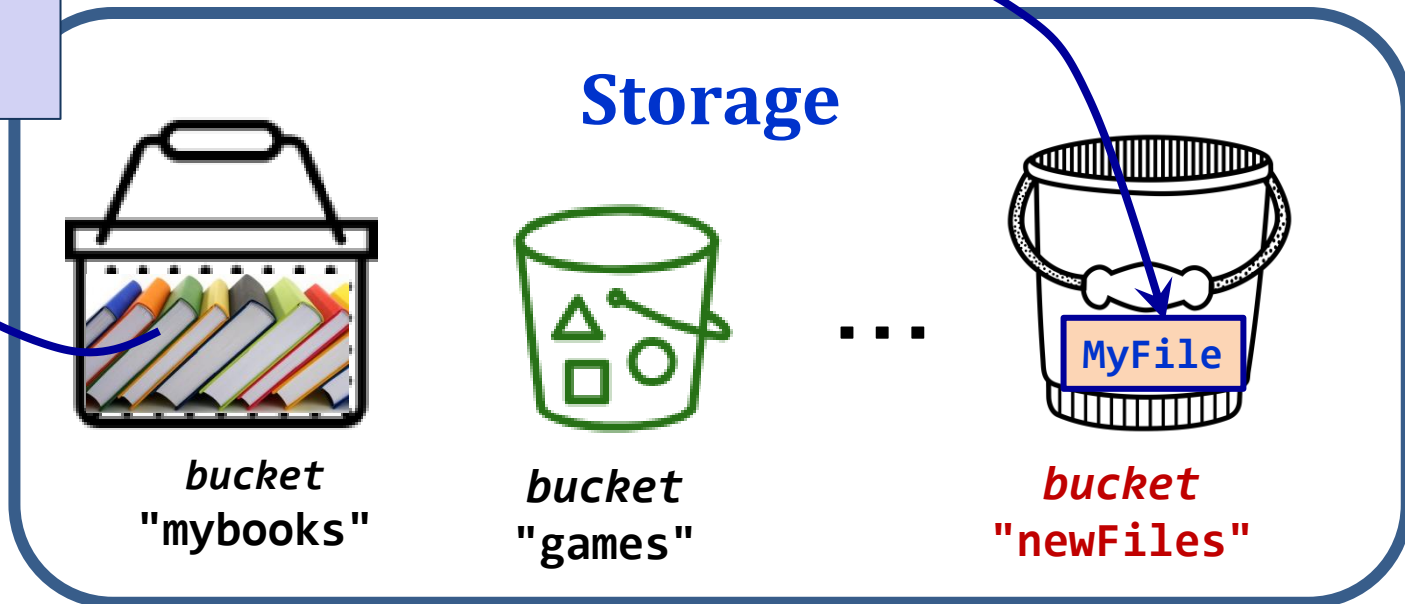
- Nos *file system* hierárquicos os ficheiros organizam-se em diretorias (*folders*)
- O armazenamento na Cloud usa um modelo *Flat*, ou *data lake*, onde os ficheiros, no seu formato nativo, são armazenados como objetos imutáveis, designados por: **(BLOB - *Binary Large Object*)**
- Os BLOB são identificados por identificadores únicos e um conjunto de *tags* ou *labels* que definem metadados sobre o ficheiro
- Por exemplo, um ficheiro JPG com uma foto pode ter os seguintes metadados: Local e data da foto, Resolução em pixels, etc.

Armazenamento como agrupamentos de objetos

Storage GCP: armazenamento baseado em coleções/agrupamentos (*buckets*) de objetos (BLOB) de qualquer tipo

```
PutBLOB("MyFile", Bucket="newFiles", Metadata={ (type, CSV), ...})
```

Metadata:
author: "John"
title: "GCP"
type: "PDF"



Requisitos do serviço *Google Cloud Storage*

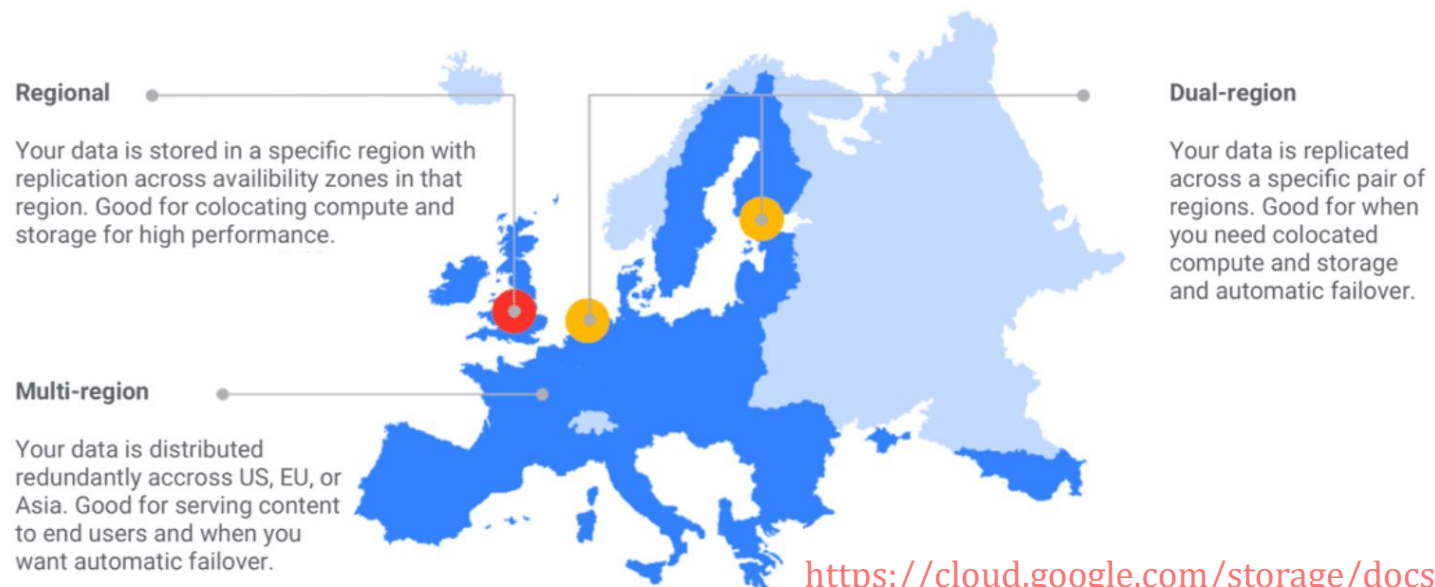
- Diferentes tipos de localização (*Regional, Dual-region, Multi-region*)
- Classes de acesso aos dados, desde alta frequência a menos frequente (*Standard, Nearline, Coldline, Archive*)
- Alta disponibilidade (> 99%) com replicação distribuída geograficamente
- Tolerância a falhas (de energia, de hardware e humanas)
- Tempos de resposta na ordem dos milissegundos
- Um objeto pode ter dimensões até 5TB
- Consistência ao nível do objeto BLOB

Tipos de localização

Regional Dados armazenados numa localização regional específica (ex. *us-central1* ou *asia-east1*) sem ter redundância em largas áreas geográficas

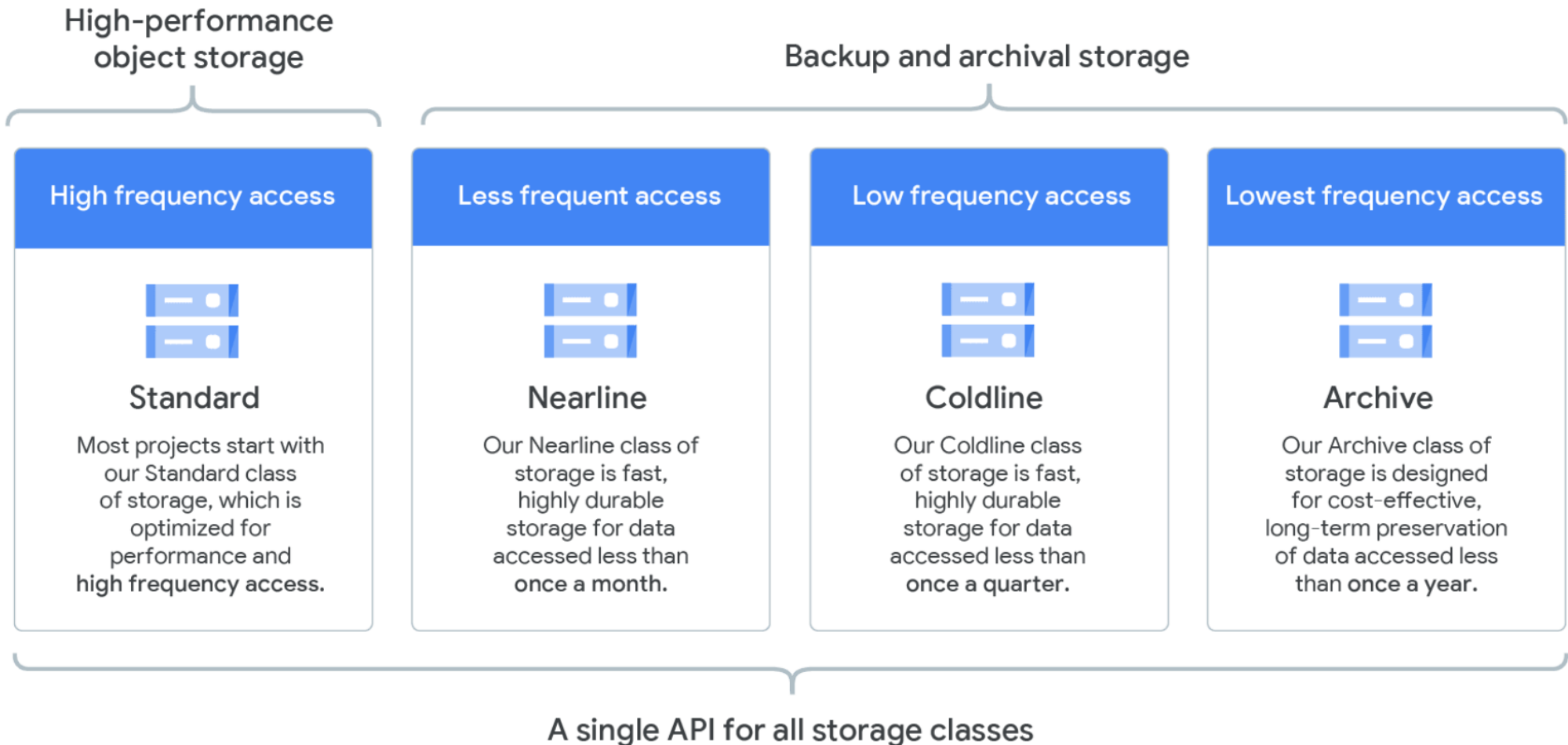
Dual-Region Dados armazenados em duas regiões (ex. *eur4* (Netherlands and Finland)). Disponibilidade na presença de falhas ou *disaster recovery*

Multi-region (*geo-redundancy*) Dados armazenados com redundância em múltiplas regiões (ex: *usa*, *eu*, *asia*). Disponibilidade na presença de falhas ou *disaster recovery*. Ideal para dados acedidos frequentemente (*web sites*, mobile Apps ou jogos)



<https://cloud.google.com/storage/docs/locations>

Tipos de classe de Acesso



<https://cloud.google.com/storage/docs/storage-classes>

Pricing

Iowa (us-central1) ▾

Standard Storage
(per GB per Month)

\$0.020

Nearline Storage
(per GB per Month)

\$0.010

Coldline Storage
(per GB per Month)

\$0.004

Archive Storage
(per GB per Month)

\$0.0012

Finland and Netherlands (eur4) ▾

Standard Storage
(per GB per Month)

\$0.036

Nearline Storage
(per GB per Month)

\$0.020

Coldline Storage
(per GB per Month)

\$0.009

Archive Storage
(per GB per Month)

\$0.005

US (multi-region) ▾

Standard Storage
(per GB per Month)

\$0.026

Nearline Storage
(per GB per Month)

\$0.010

Coldline Storage
(per GB per Month)

\$0.007

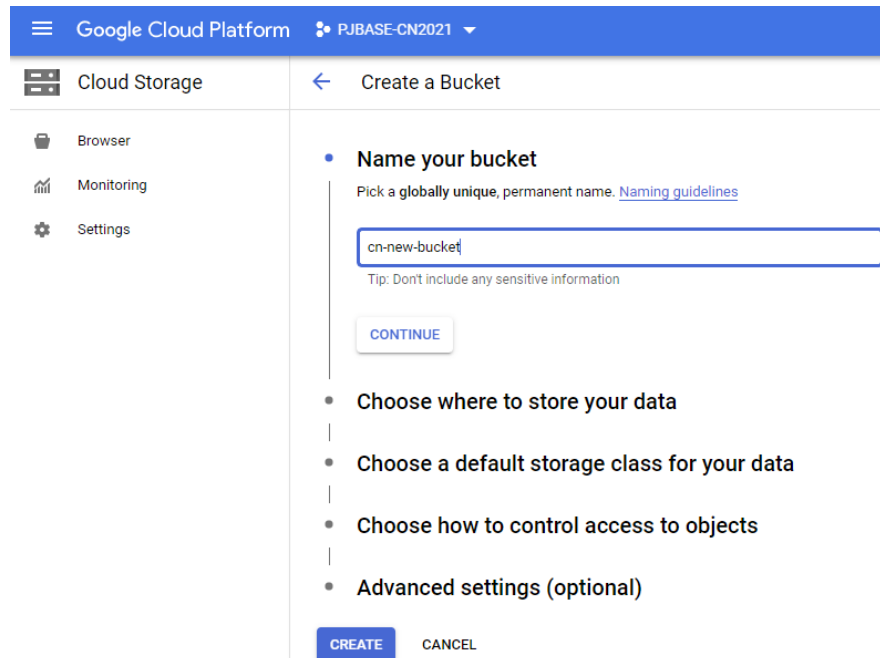
Archive Storage
(per GB per Month)

\$0.004

<https://cloud.google.com/storage/pricing#price-tables>

Criação e acesso a *buckets* e objetos

- Os *buckets* e objetos (BLOBs) podem ser criados ou acedidos via:
 - Consola web do GCP
 - URL de objetos (BLOBs) públicos
 - Ferramenta `gsutil` (do Google Cloud SDK - <https://cloud.google.com/sdk/>)
 - APIs gRPC e REST com clientes em várias linguagens de programação



Google Cloud Platform PJBASE-CN2021

Cloud Storage

Create a Bucket

- Name your bucket
Pick a globally unique, permanent name. [Naming guidelines](#)

Tip: Don't include any sensitive information
CONTINUE
- Choose where to store your data
- Choose a default storage class for your data
- Choose how to control access to objects
- Advanced settings (optional)

CREATE CANCEL

Os nomes dos *buckets* têm de ser únicos globalmente no *GCP storage*

Criação e acesso aos *buckets* e objetos

Google Cloud CN2223-LA Search (/) for resources, docs, products, and more

Cloud Storage Buckets Monitoring NEW Settings

Bucket details

cn-new-bucket

Location	Storage class	Public access	Protection
us-east1 (South Carolina)	Standard	Subject to object ACLs	None

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY

Overview

Created	April 6, 2023 at 12:58:14 PM GMT+1
Updated	April 6, 2023 at 1:05:35 PM GMT+1
Location type	Region
Location	us-east1 (South Carolina)
Replication	—
Default storage class	Standard
Requester Pays	OFF
Tags	None
Labels	None
Cloud Console URL	https://console.cloud.google.com/storage/browser/cn-new-bucket
gsutil URI	gs://cn-new-bucket

Permissions

Access control	Fine-grained
Public access prevention	Not enabled by org policy or bucket setting
Public access status	Subject to object ACLs

Protection

Object versioning	Off
Retention policy	None
Encryption type	Google-managed key

Object lifecycle

Lifecycle rules	None
-----------------	------

Antes da criação é possível prevenir acesso público

Public access will be prevented

This bucket is set to prevent exposure of its data on the public internet.

Keep this setting enabled unless you have a use case that requires public access (such as static website hosting). You can change it now or later. [Learn more](#)

- ☐ Enforce public access prevention on this bucket
- ☒ Don't show this message again

CANCEL CONFIRM

Edit access control

Choose how to control object access in this bucket.

- ☐ Uniform
Ensure uniform access to all objects in the bucket by using only bucket-level permissions (IAM). This option becomes permanent after 90 days. [Learn more](#)
- ☒ Fine-grained
Specify access to individual objects by using object-level permissions (ACLs) in addition to your bucket-level permissions (IAM). [Learn more](#)

CANCEL SAVE

Objetos (BLOBs) de um bucket

**BLOBs no
bucket**

Google Cloud Platform PJBASE-CN2021 Search products

Cloud Storage

Browser

Monitoring

Settings

Bucket details

cn-new-bucket

OBJECTS CONFIGURATION PERMISSIONS RETENTION LIFECYCLE

Buckets > cn-new-bucket

UPLOAD FILES UPLOAD FOLDER CREATE FOLDER MANAGE HOLDS DOWNLOAD DELETE

Filter by name prefix only Filter Filter objects and folders

<input type="checkbox"/>	Name	Size	Type	Created time ?
<input type="checkbox"/>	A_NGC1333_APOD1024.jpg	773.1 KB	image/jpeg	22 Apr 2021, 14:08:54
<input type="checkbox"/>	The-Google-File-System-sosp2003.pdf	269.5 KB	application/pdf	22 Apr 2021, 14:08:53
<input type="checkbox"/>	fotos/	—	Folder	—

Existem no *bucket* BLOBs com nome iniciado com fotos/

Como funciona o espaço de nomes

- O espaço de nomes dos *buckets* é global no GCP e por isso não pode haver *buckets* com nomes iguais
- Os nomes dos objetos (*blobs*) podem conter '/' para dar a ilusão de uma divisão lógica do espaço de nomes (*folders*), e enriquecer a interface das aplicações cliente



Exemplo: `$ gsutil ls gs://cn-new-bucket/fotos/`

Listas de controlo de acessos - permissões

- Os *buckets* podem ter as seguintes políticas:
 - Uniform*: Lista de controlo de acessos (ACL) única para todos os objetos
 - Fine-grained*: Lista de controlo de acessos específica para cada objeto
- A definição da ACL para objetos (com *bucket fine-grained*) pode incluir:

Entity 1 * User ▼	Name 1 lassuncao.cn1920@g	Access 1 * Owner ▼
Entity 2 * Public ▼	Name 2 * allUsers ▼	Access 2 * Reader ▼
Entity 3 * User ▼	Name 3 jose.m.simao@gmail.	Access 3 * Owner ▼
Entity 4 * Domain ▼	Name 4 maria@domain.pt	Access 4 * Reader ▼

+ ADD ENTRY

- ❖ “*allUsers*” dá acesso a todos os utilizadores, mesmo que não tenham conta no GCP
- ❖ “*allAuthenticatedUsers*” dá acesso a utilizadores autenticados com conta GCP:”

Alterar permissão num objeto para acesso público

The screenshot illustrates the process of making a Google Cloud Storage object public. It is divided into two parts: the main console view on the left and a context menu on the right.

Main Console View (Left):

- Object name:** 04-Visao-geral-servicos-GCP-Intro.pdf
- Status:** This object is public and can be accessed by anyone on the internet. To remove public access, search for and remove all public entries from the object's permissions.
- Permissions Table:**

Entity *	Name *	Access *
User	lassuncao.cn1920@g	Owner
Public	allUsers	Reader
- Buttons:** + ADD ENTRY, CANCEL, SAVE

Context Menu (Right):

- Download
- Copy Authenticated URL
- Copy gsutil URI
- Edit metadata
- Edit access** (highlighted with a red box and a red arrow pointing from the menu icon)
- Copy
- Move
- Rename
- Export to Cloud Pub/Sub
- Scan with Cloud Data Loss Prevention

Bottom Bar:

- Object: 04-Visao-geral-servicos-GCP-Intro.pdf
- Size: 751.8 KB
- Type: application/pdf
- Created: 29 Apr 2021, 13:34:50
- Standard
- Last Modified: 29 Apr 2021, 18:04:15
- Public to Internet** (highlighted with a red box and a red arrow pointing to the URL below)
- Copy URL

Public URL: <https://storage.googleapis.com/cn2122-test-operations/slides2022/04-Visao-geral-servicos-GCP-Intro.pdf>

Exemplos com *gsutil* no *Google Cloud Shell*

- `wget https://apod.nasa.gov/apod/image/1903/A_NGC1333_APOD1024.jpg`
- `gsutil cp A_NGC1333_APOD1024.jpg gs://<bucket name>`
- `gsutil cp A_NGC1333_APOD1024.jpg gs://<bucket name>/pictures`
- `gsutil ls -lhr gs://<bucket name>`
 - lista de objetos, recursivo e com detalhes

<https://cloud.google.com/storage/docs/gsutil/commands/ls>

Metadados

- Cada objeto tem associado *metadados* na forma ***chave : valor***
- Existe um conjunto pré-determinado de chaves mas outras podem ser acrescentadas a cada objecto
 - **Access control metadata:** <https://cloud.google.com/storage/docs/access-control/lists>
 - **Cache-Control:** <https://cloud.google.com/storage/docs/metadata#cache-control>
 - **Content-Disposition:** <https://tools.ietf.org/html/rfc6266>
 - **Content-Encoding:** <https://cloud.google.com/storage/docs/transcoding>
 - **Content-Language:** Língua dos dados do Blob (ex: English, Portuguese,...)
 - **Content-Type:** MIME type do blob (ex: *application/pdf*; *image/jpg*)
- Acrescentar novas chaves tem custos extra no armazenamento e transporte (cada caracter de *chave* ou *valor* conta 1 byte)

Cache-Control

- A chave de *metadados* **cache-control** regula se os objetos são mantidos em *cache* nos nós da Google mais próximos do pedido



Centros de dados



Edge Points of Presence
para content delivery



Google Global Cache com
ligação à rede Google

<https://peering.google.com/>

Consistência

- A maior parte das operações têm consistência forte
 - *Read-after-write*: leituras após escritas
 - *Read-after-metadata-update*: leituras após alteração de *metadados*
 - *Read-after-delete* : insucesso nas leituras após remoção
 - *Bucket listing*: leitura de *bucket* após criação
 - *Object listing*: listagem do objeto após criação
- Operações com consistência eventual
 - *Granting or Revoking access from resources*: para acessos após dar ou revogar autorizações
 - Objetos com cache ativa na *metadata* de *cache-control*

<https://cloud.google.com/storage/docs/consistency>

<https://cloud.google.com/storage/docs/metadata#cache-control>

Boas práticas e Limites

- Boas práticas
 - Não usar como nome de *buckets* dados sensíveis (email, IDs de projeto, etc.)
 - Nomes diferentes podem ser gerados usando GUIDs
 - Perante erros, deve usar-se técnicas de *retry* segundo uma abordagem *exponential backoff* - <https://cloud.google.com/storage/docs/exponential-backoff>
- Limites
 - Objetos (BLOB) limitados a 5 TB
 - Atenção às regras de nomes dos *buckets* e objetos:
<https://cloud.google.com/storage/docs/naming#requirements>
 - Por projeto, a criação e destruição de *buckets* está limitada por operações por segundo
 - O ritmo de escritas pode variar ao longo do tempo (*Ramp up request rate gradually*)
 - Distribuição (*sharding*) dos BLOB por múltiplos servidores

<https://cloud.google.com/storage/docs/request-rate#indexing>