

Diseño de un sistema de gestión de archivos escalable y resiliente

Por: Daniel Valencia Backhoff
Senior Software Engineer @ HELIX RE 

 @DanValencia
 danvalencia

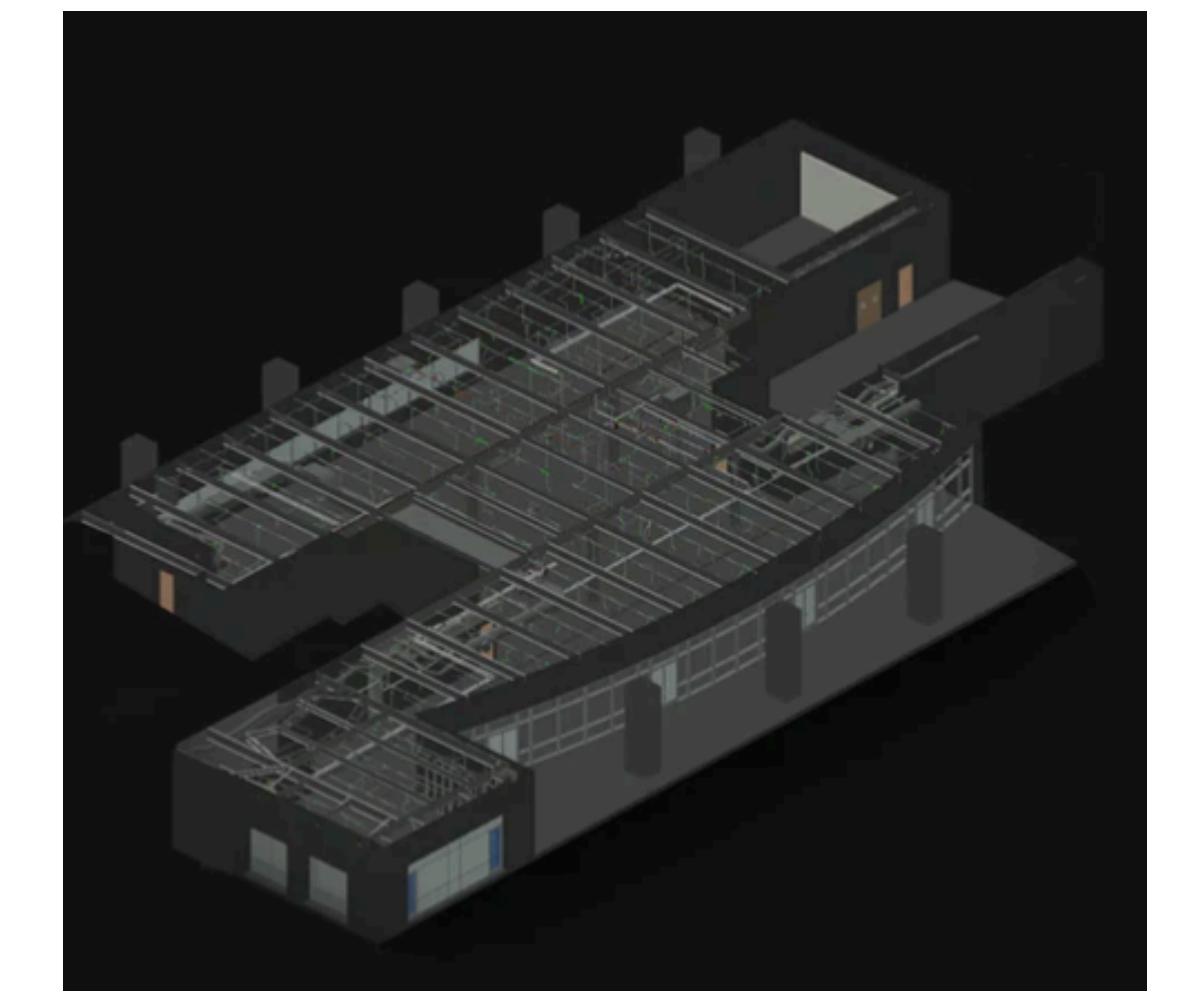
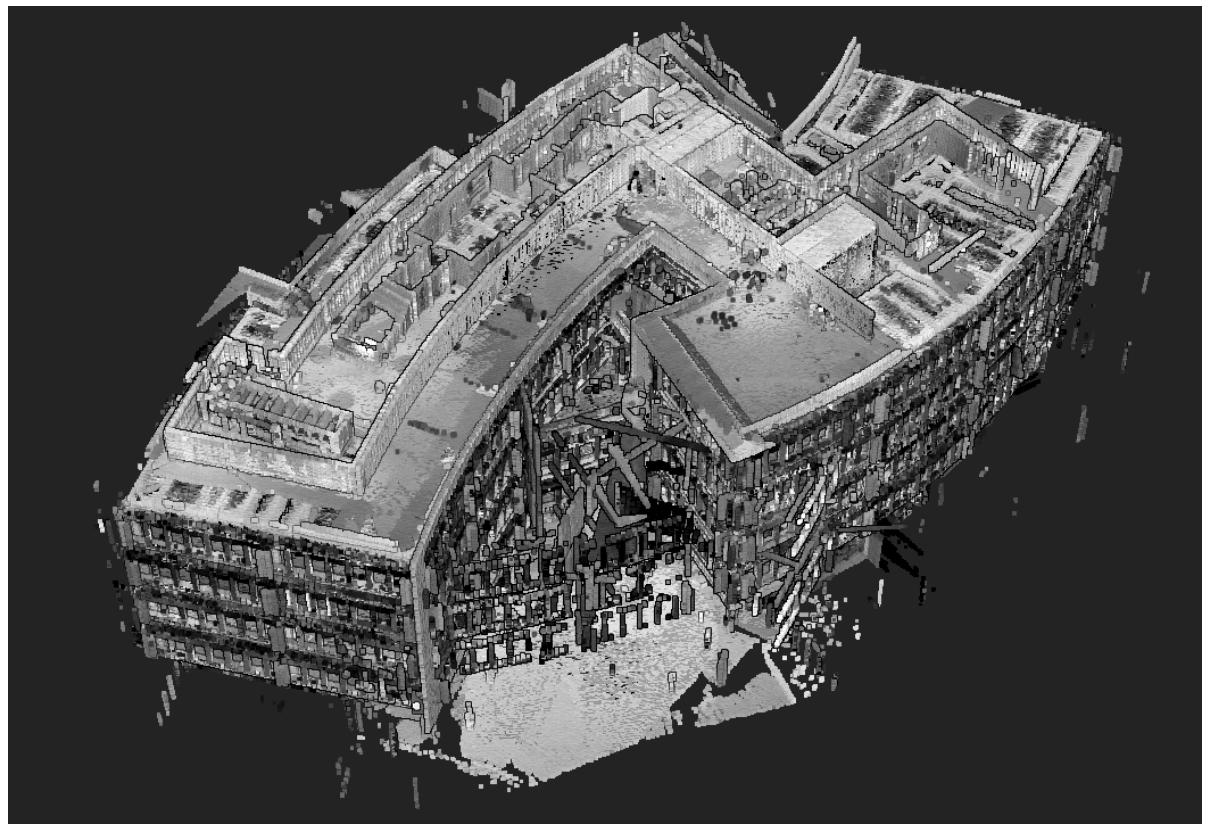
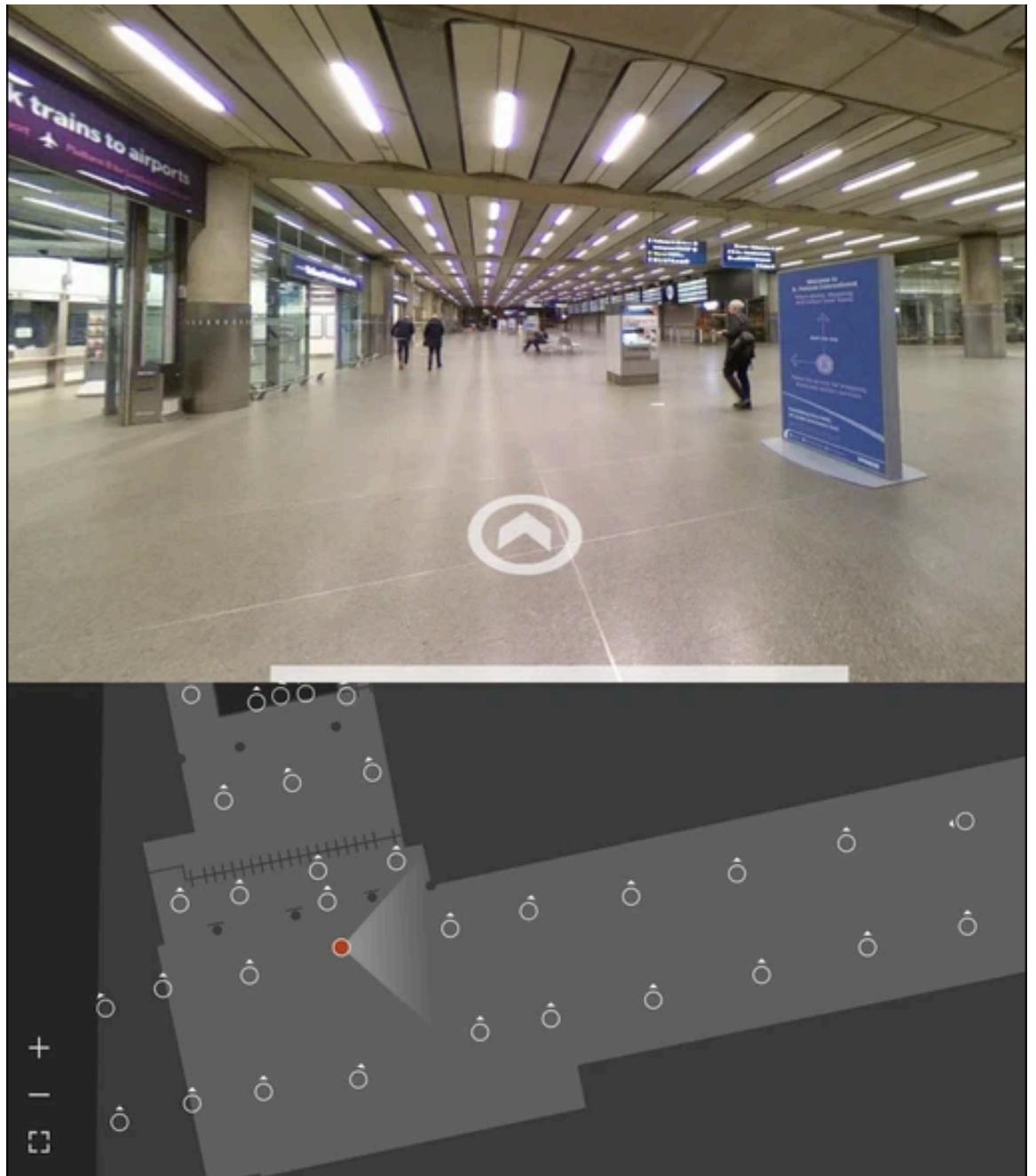
Agenda

- Qué hacemos en HELIX ?
- Requerimientos del sistema de carga/descarga de archivos
- Posibles Arquitecturas y la Nube
- Ventajas y Desventajas
- Demo

Qué hacemos en HELIX ?

QUÉ hacemos en HELIX ?

- Digitalización de edificios utilizando LiDAR y fotografías 360°
- Nuestros entregables son modelos 3d y tour virtual de los espacios.
- Plataforma para visualizar/colaborar con dichos modelos sin necesidad de herramientas externas.
- Aplicación interna que nuestro equipo de Reality Capture utiliza para automatizar ciertos procesos y crear los tours virtuales.

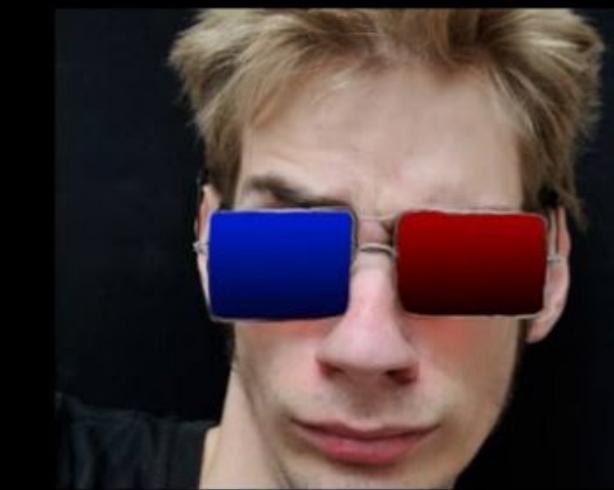


QUÉ hacemos en HELIX ?

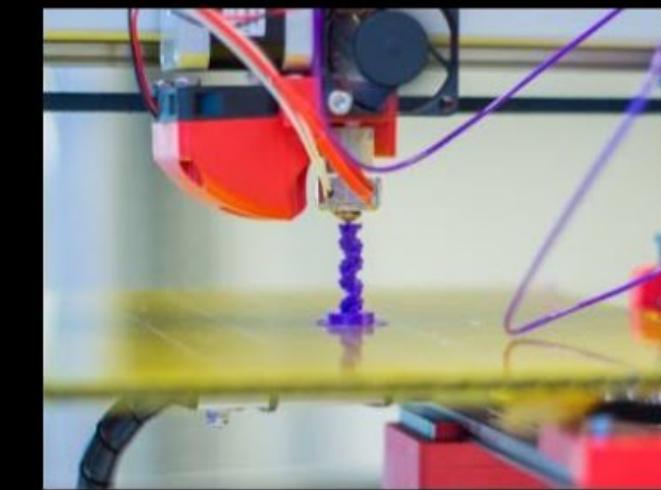
3D LASER SCANNING



What my friends think I do



What my Mom think I do



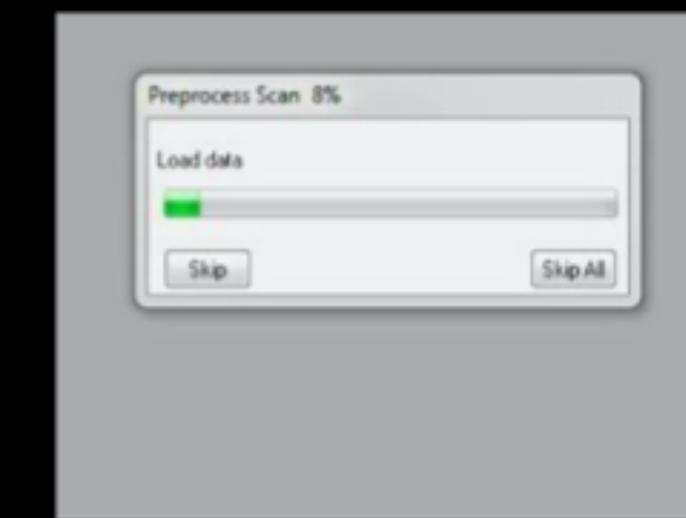
What society think I do



What my co workers
think I do



What I think I do



What I actually do

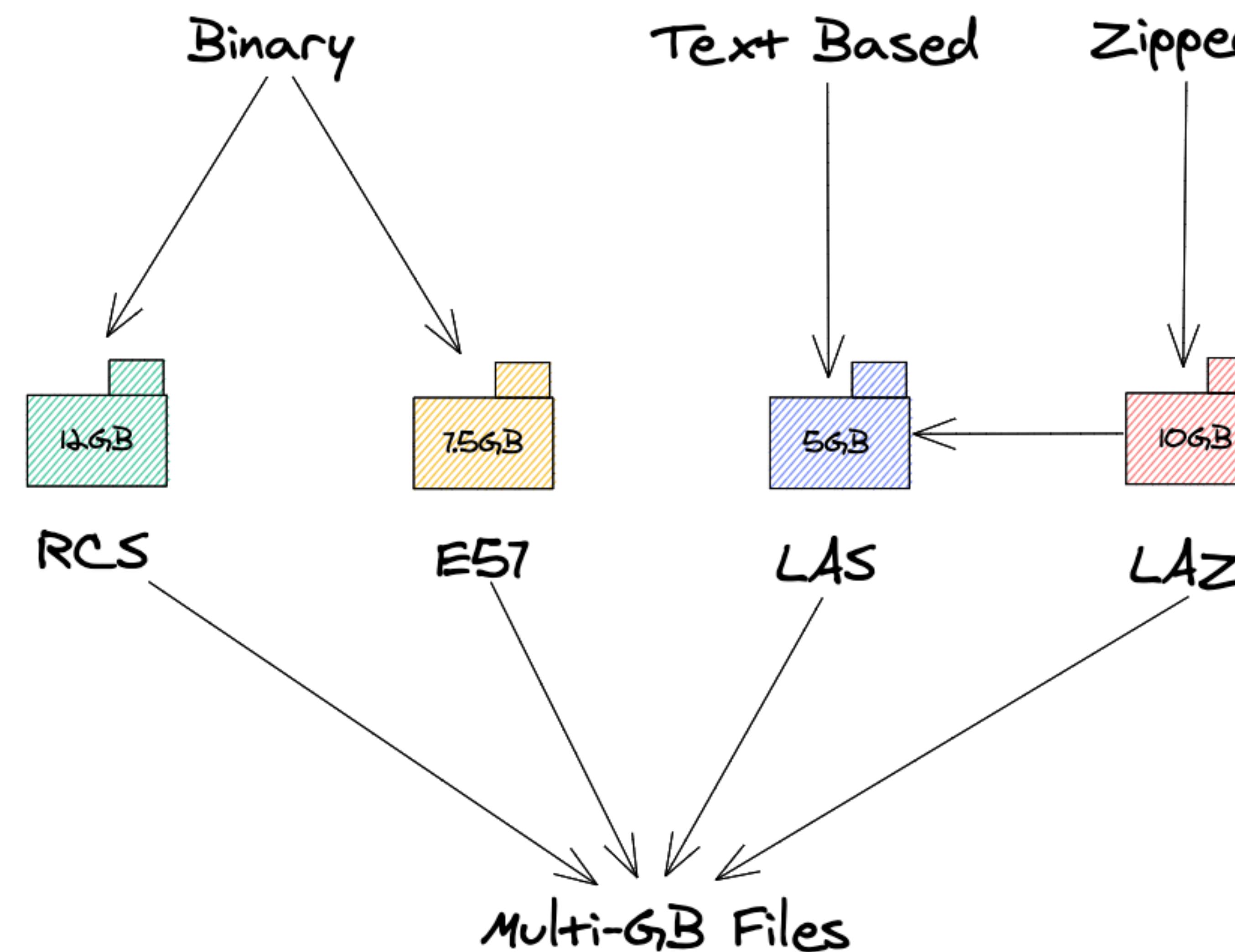
LiDAR

Per Wikipedia:

"...is a surveying method that measures distance to a target by illuminating the target with laser light and measuring the reflected light with a sensor.

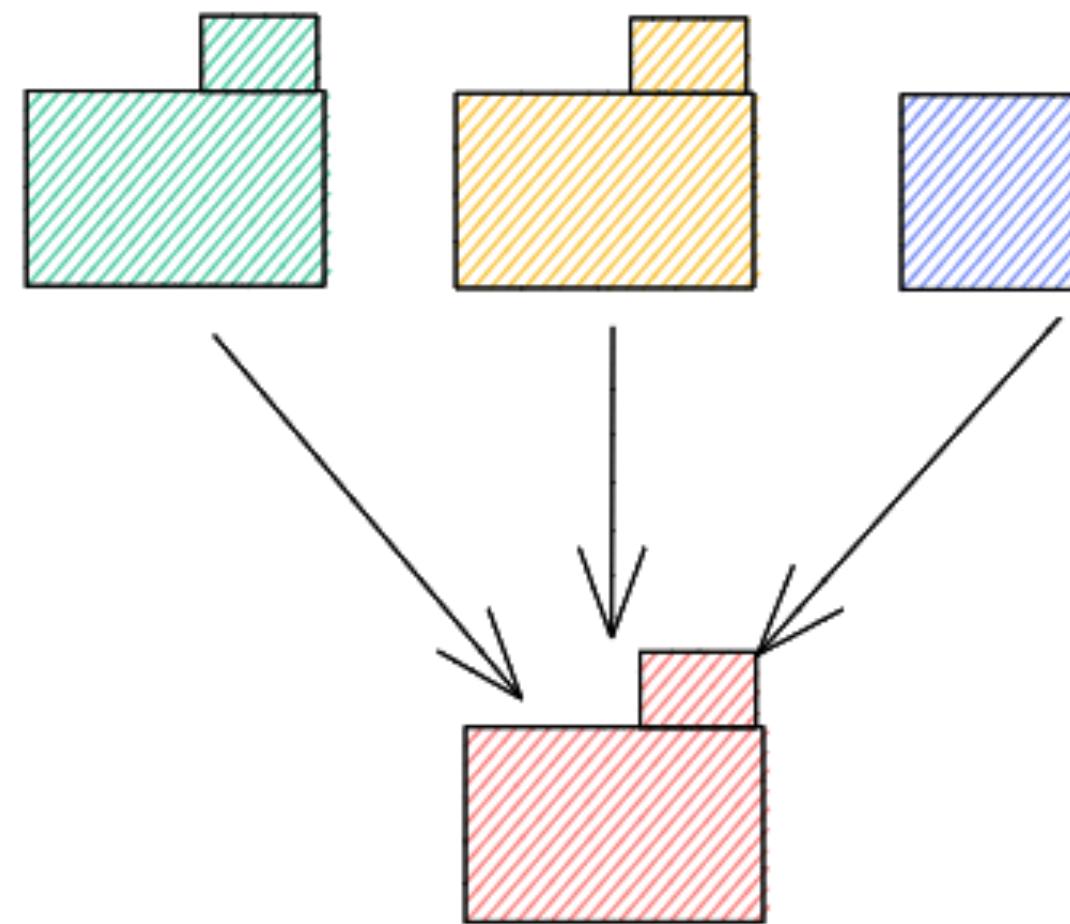
Differences in laser return times and wavelengths can then be used to make digital 3-D representations of the target."

LiDAR Output == Point Clouds

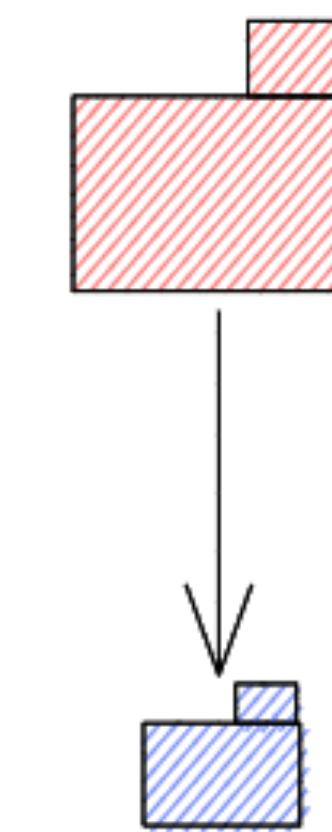


Tareas con Point Clouds

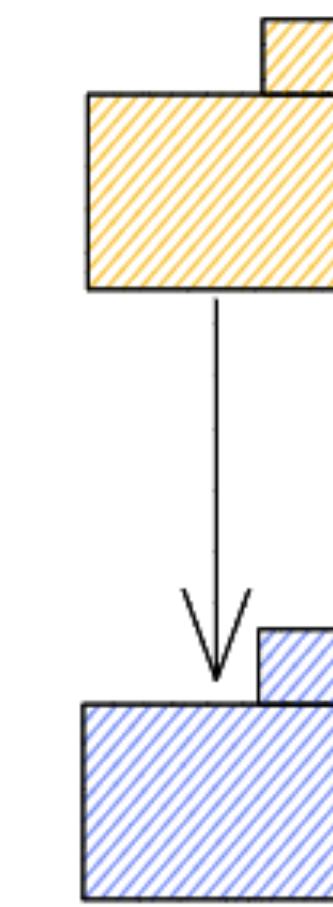
Combinación



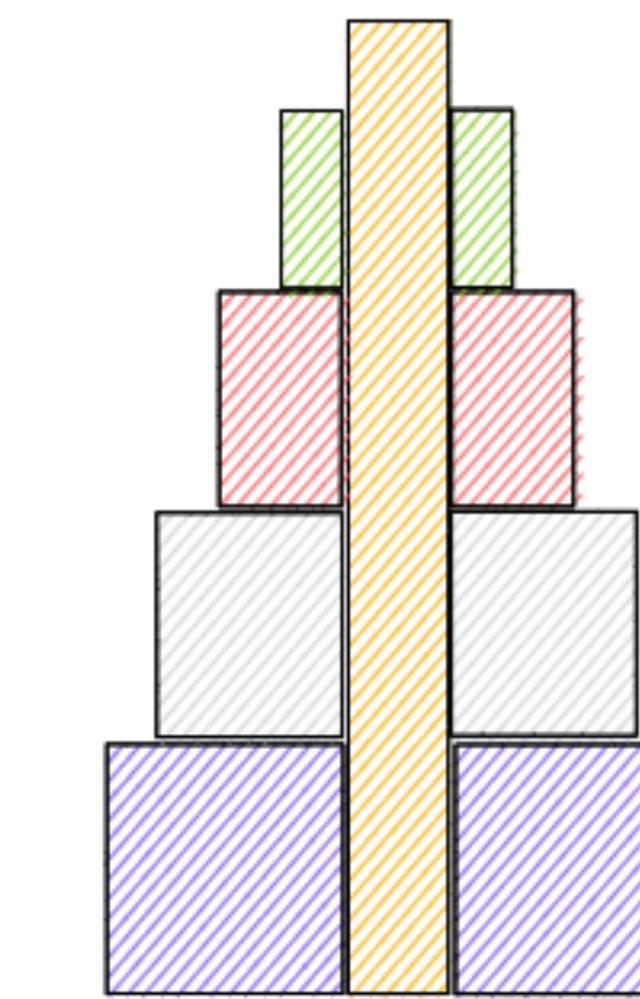
Extracción



Conversión

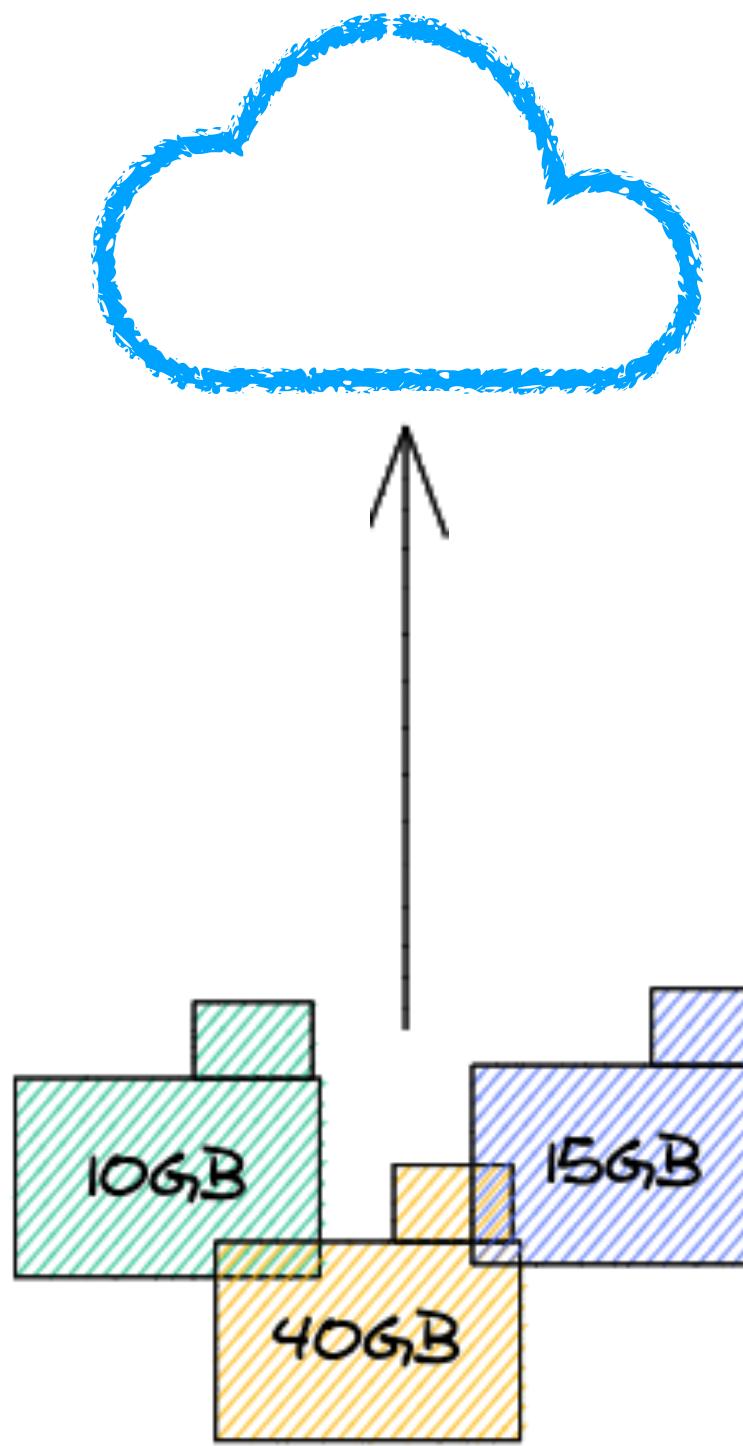


Clasificación



Requerimientos del Sistema

Requerimientos para carga/descarga de archivos

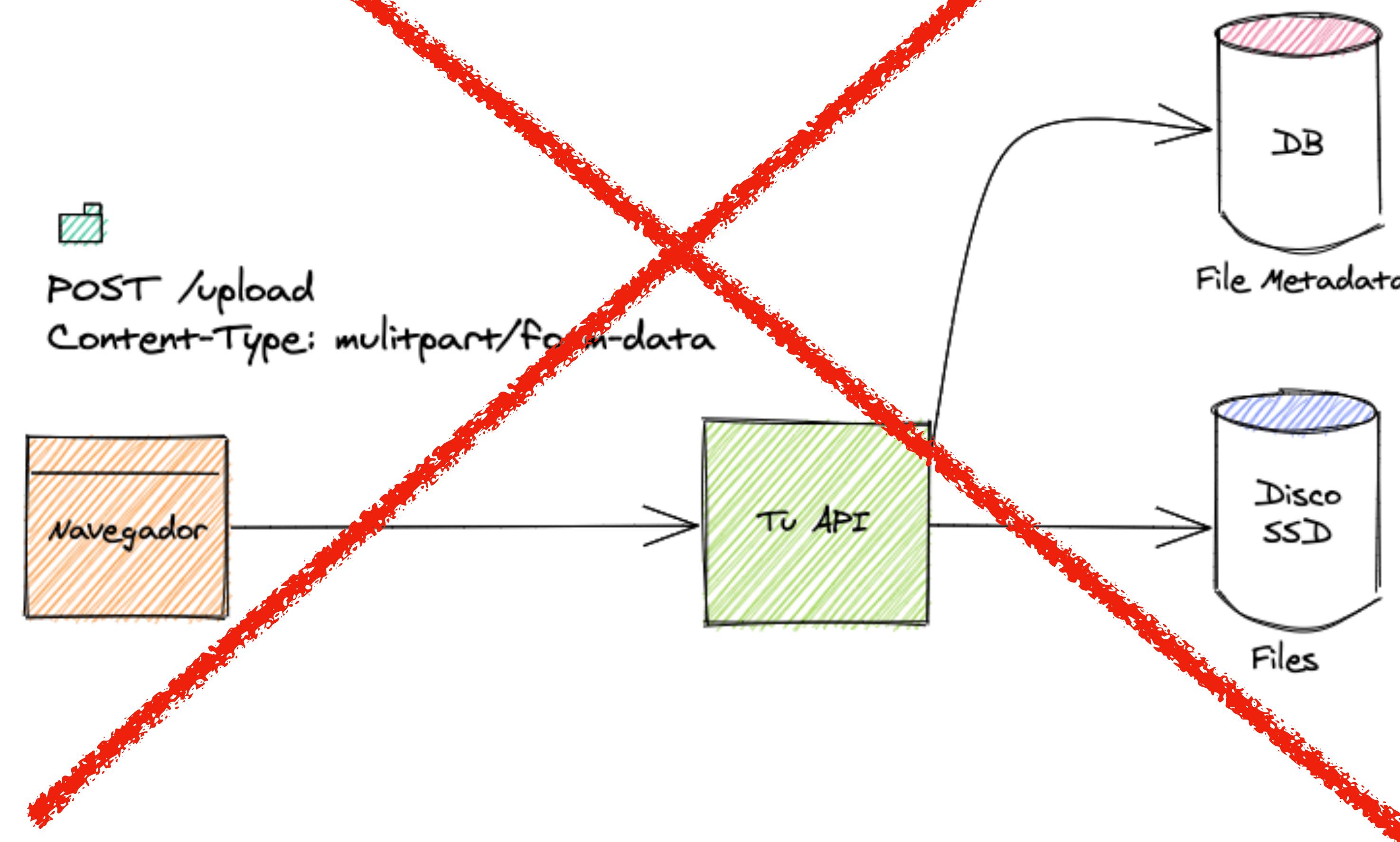


Cargas Multi-GB
Confiables

- * Habilidad de reponerse ante:
 - * Fallas de Red
 - * Fallas de Usuarios
- * Almacenamiento de Archivos Escalable
- * Aprovechar soluciones de la Nube
- * Distribución Global

Posibles Arquitecturas

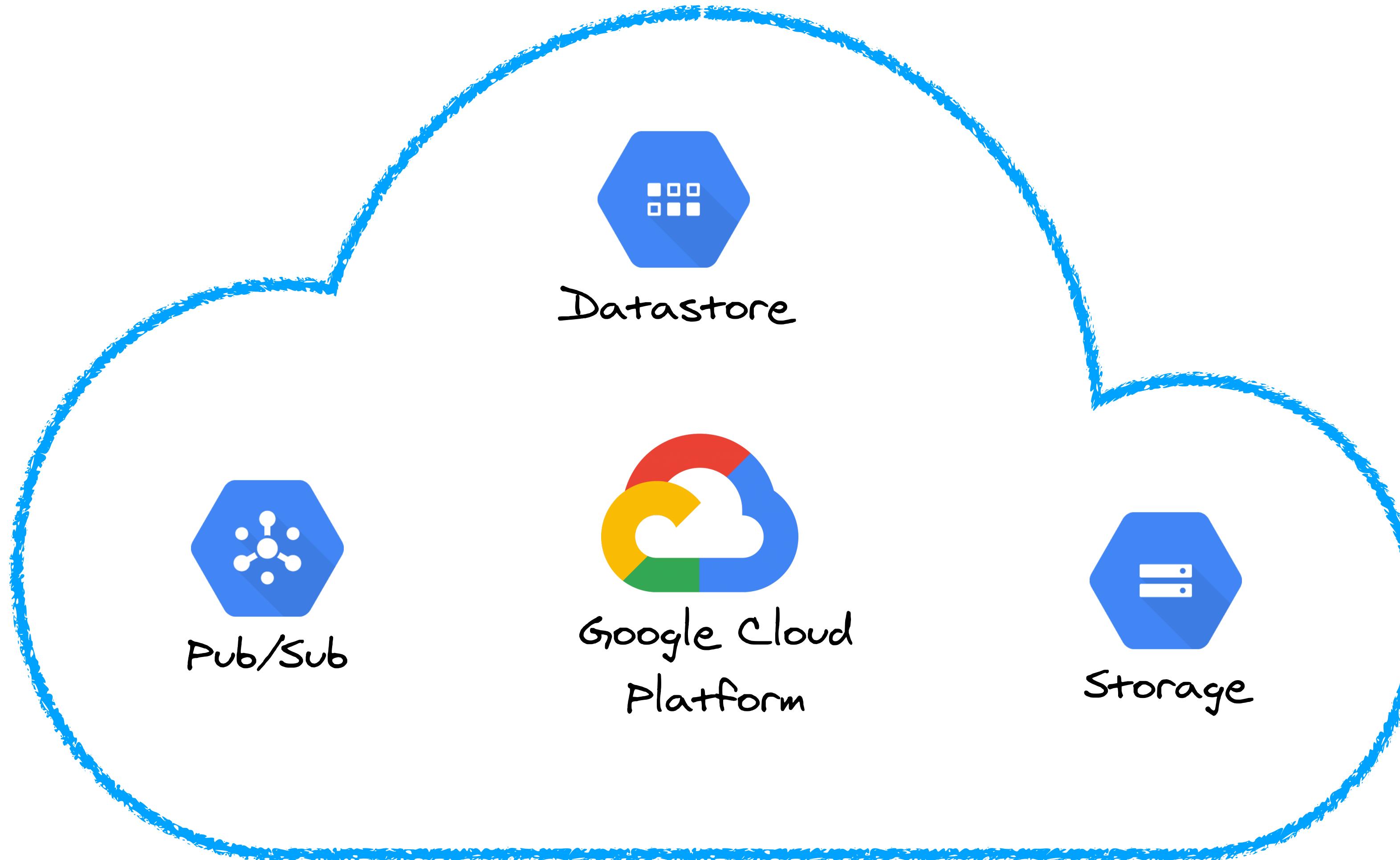
Opción 1: Solución 100% a la medida



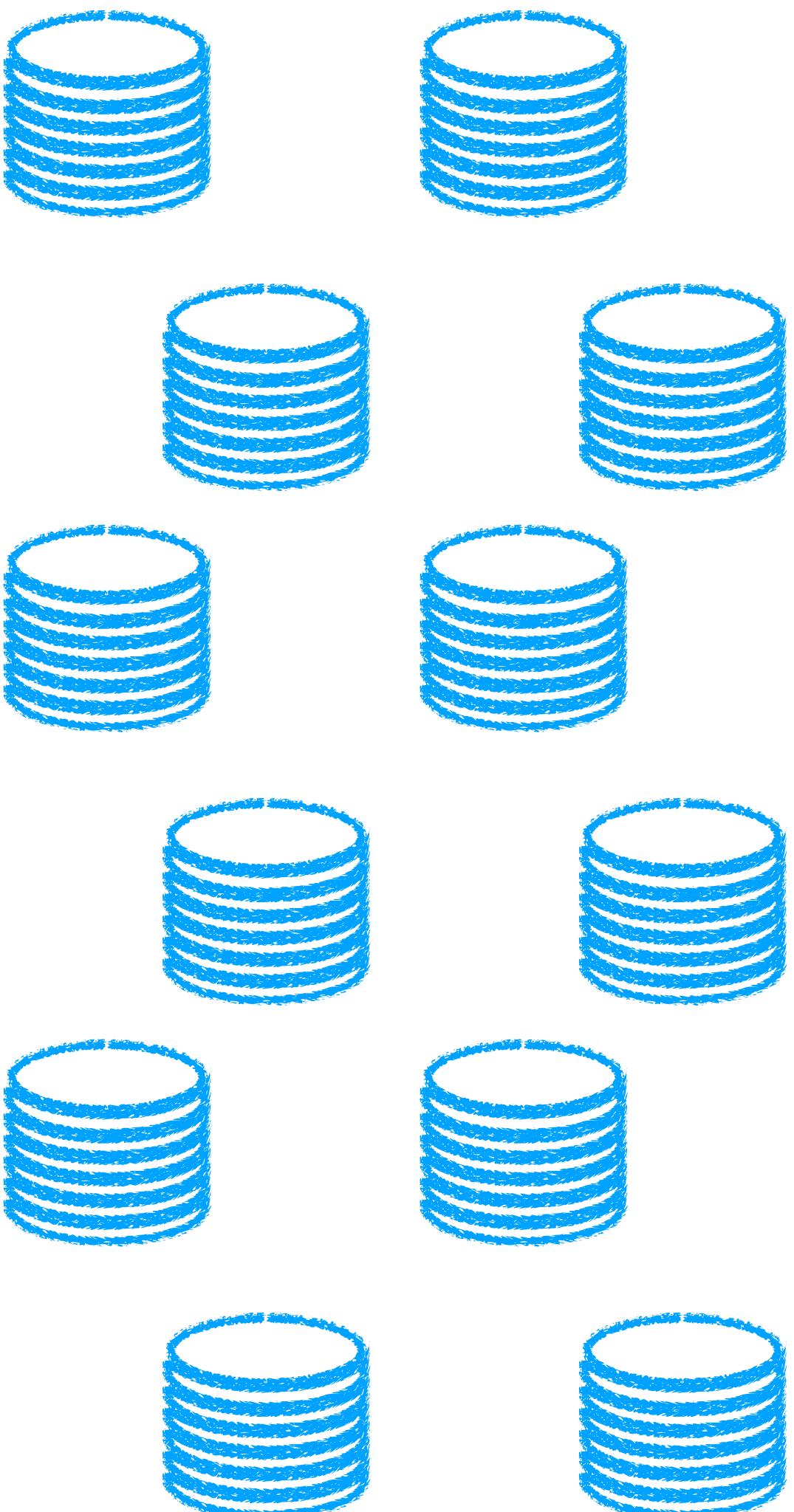
Opción 2: Aprovechar "La Nube"



Opción 2: Aprovechar "La Nube"

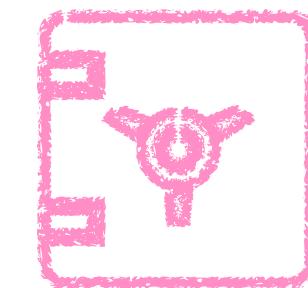
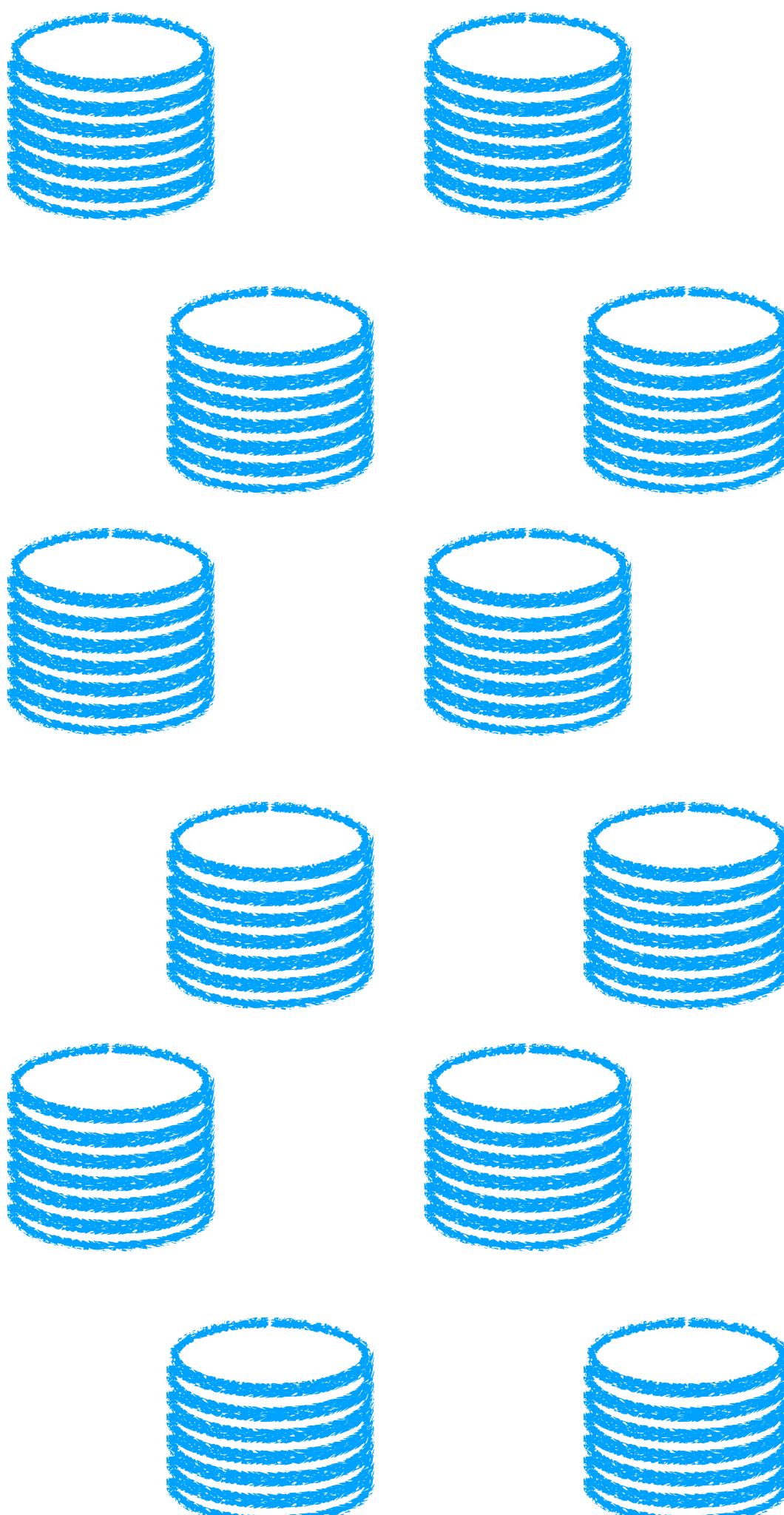


Google Cloud Storage

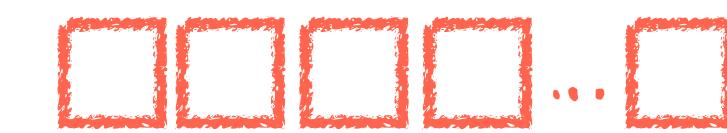


- * Almacenamiento de Objetos (i.e. Archivos)
- * Diferentes clases de almacenamiento
- * Integración con CDN
- * Distribuido globalmente (múltiples regiones)
- * Notificaciones por Pub/Sub
- * Registro auditable
- * Versionamiento
- * Respaldos automáticos

Google Cloud Storage



URLs Cifradas
(Signed URLs)



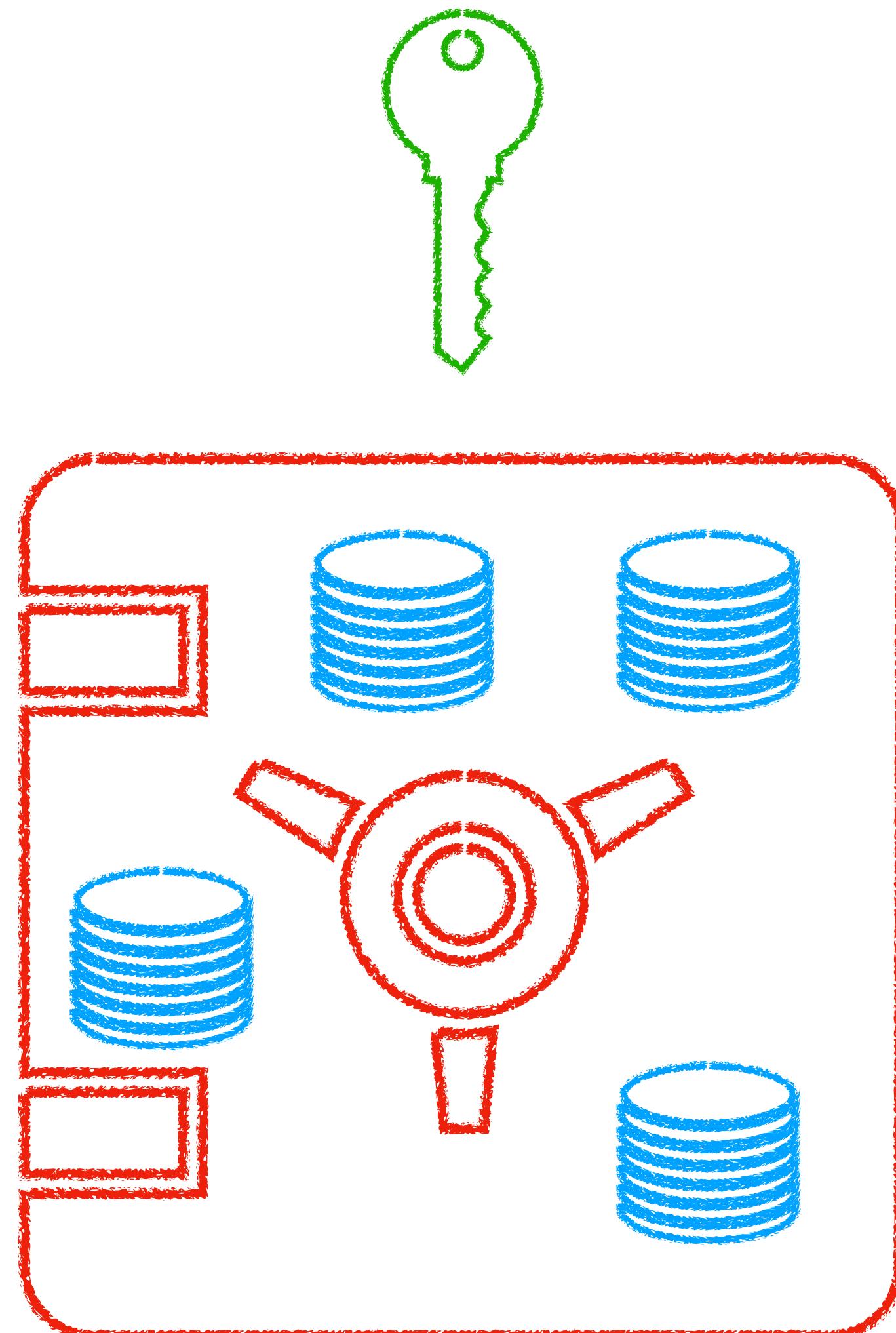
Cargas en pedazos
(Chunked Uploads)



Cargas Reanudables
(Resumable Uploads)

URLs Cifradas

Signed URLs

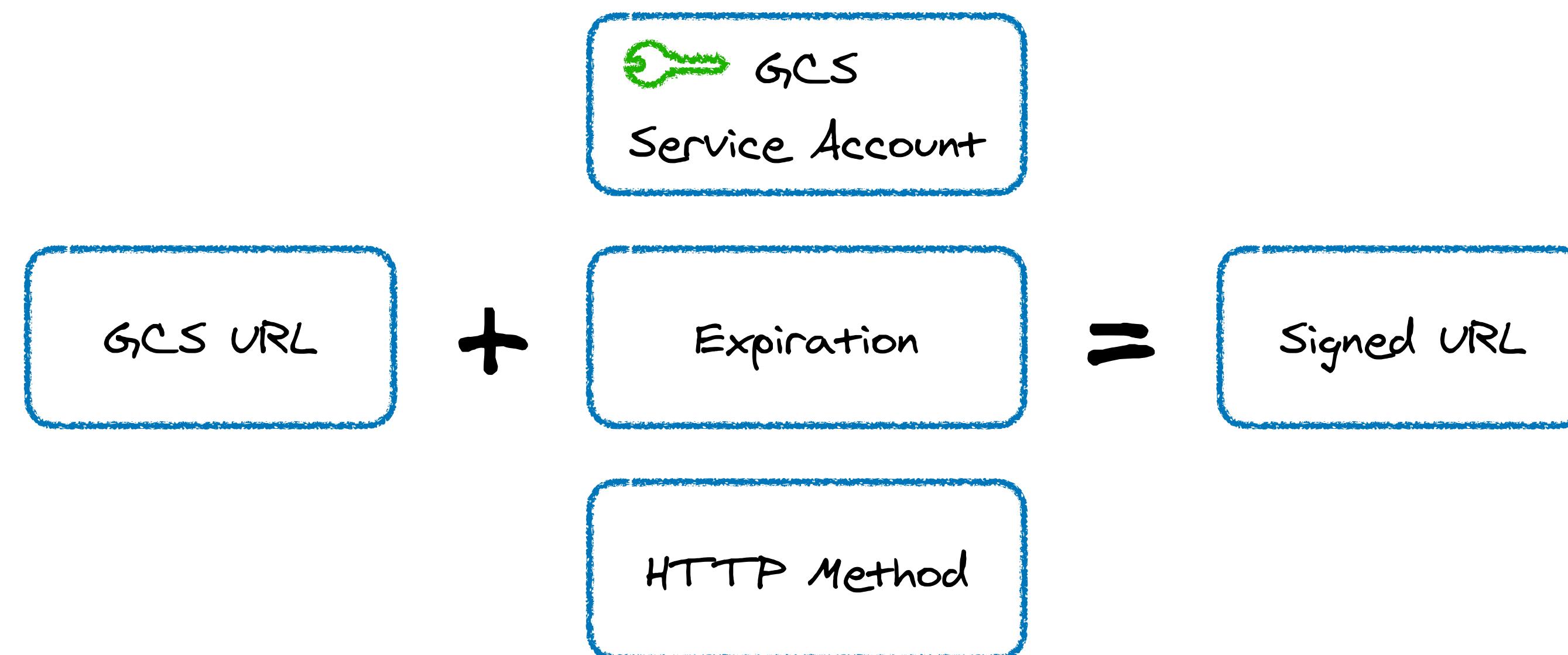


- * URL que provee acceso limitado temporal a objetos en GCS.
- * Contienen información de autenticación en el Query string, permitiendo a usuarios sin credenciales operar sobre un recurso.

URL en GCS

gs://helix-dev-drive/drive/de/e8/
dee8f2c8351441439d2e41529d6b95fe/big_one.laz

[https://storage.googleapis.com/helix-dev-drive/drive/de/e8/
dee8f2c8351441439d2e41529d6b95fe/big_one.laz](https://storage.googleapis.com/helix-dev-drive/drive/de/e8/dee8f2c8351441439d2e41529d6b95fe/big_one.laz)



Anatomía de una URL Cifrada

`https://storage.googleapis.com/helix-dev-drive/drive/de/e8/dee8f2c8351441439d2e41529d6b95fe/big_one.laz?GoogleAccessId=helix-dev-k8s@helix-dev-195819.iam.gserviceaccount.com&Expires=1587278741&Signature=RHUM5Eet%2BV0JyKFSxb2uT3lLBUj7RXE4ow5H8v356f%2F2F4g8q5jYYLi8kabIutQkUJAibnJ7SrbLv4eSF434b5bSnELLBORT97XFcUQLWGE8QK23IAWGQI%2FLx8SnhN74BeiJSaDTy2NK52rnUlzASRPAKDZeihcY2j6GwVTNH6vSXCtMvASkPN24sI5ixIF%2FXbepWuBaSWMZvE7vwiAPHlX%2F6FA9RCZ0T0QezZC6wTk%2BFiHLzDlBupVOxU4Wnr31b%2B%2BvhCCV8dfCPi1lXyq0gWhAk9FLN%2Fvib%2FNppzi0LhW3bcW9Ral1Gxa9oMNB4AF4q8JrjhZr6MtoGaS3FRW0Fg%3D%3D&upload_id=AAANsUk5J68kg08DTYZFSNsxIZSK3rIW0wp9sYKnuqDFTWI6GgIdKdhtFPsxN9Qv2athzB-pdNH5RcEJjMnYdERWReU`

Anatomía de una URL Cifrada

`https://storage.googleapis.com/helix-dev-drive/drive/de/e8/dee8f2c8351441439d2e41529d6b95fe/big_one.laz`

`GoogleAccessId=helix-dev-k8s@helix-dev-195819.iam.gserviceaccount.com`

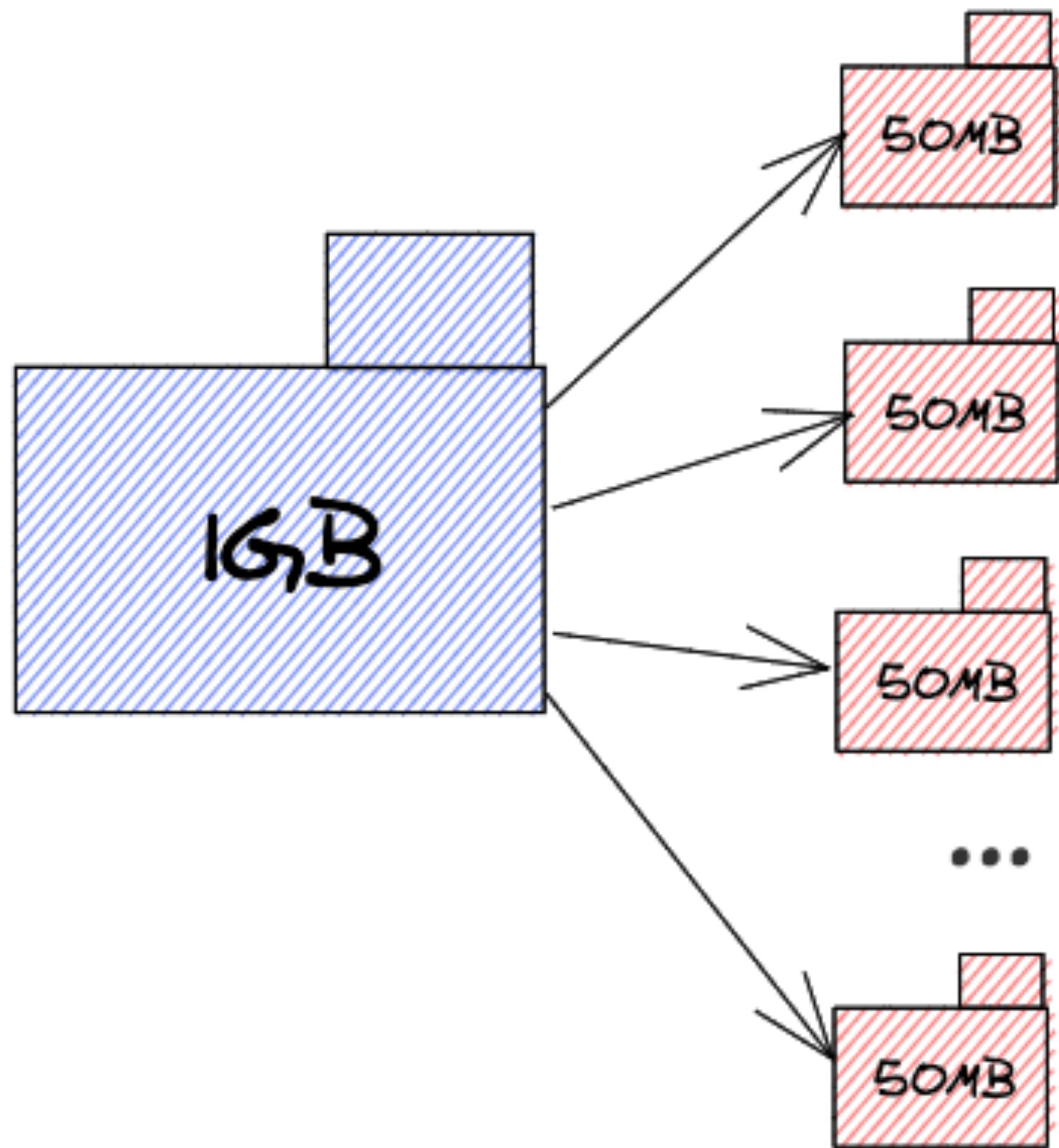
`Expires=1587278741`

`Signature=RHUM5Eet%2BV0JyKFSxb2uT3lLBuj7RXE4ow5H8v356f%2F2F4g8q5jYYLi8ka
bIutQkJJAibnJ7SrbLv4eSF434b5bSnElLBORT97XFcUQLWGE8QK23IAWGQI%2FLx8SnhN74
BeiJSaDTy2NK52rnUlzASRPAKDZeihcY2j6GwVTNH6vSXctMvASkPN24sI5ixIF%2FXbepWu
BaSWMZvE7vwiAPHlX%2F6FA9RCZ0T0QezZC6wTk%2BfIHLzDlBupVOxU4Wnr31b%2B%2BvhC
CV8dfCPi1lXyq0gWhAk9FLN%2Fvib%2FNppzi0LhW3bcW9Ral1Gxa9oMNB4AF4q8JrjhZr6M
toGaS3FRW0Fg%3D%3D`

`upload_id=AAANsUk5J68kg08DTYZFSNsxIZSK3rIwOwp9sYKnuqDFTWI6GgIdKdhtFPsxN9
Qv2athzB-pdNH5RcEJjMnYdERWReU`

Cargas en Pedazos

chunked uploads



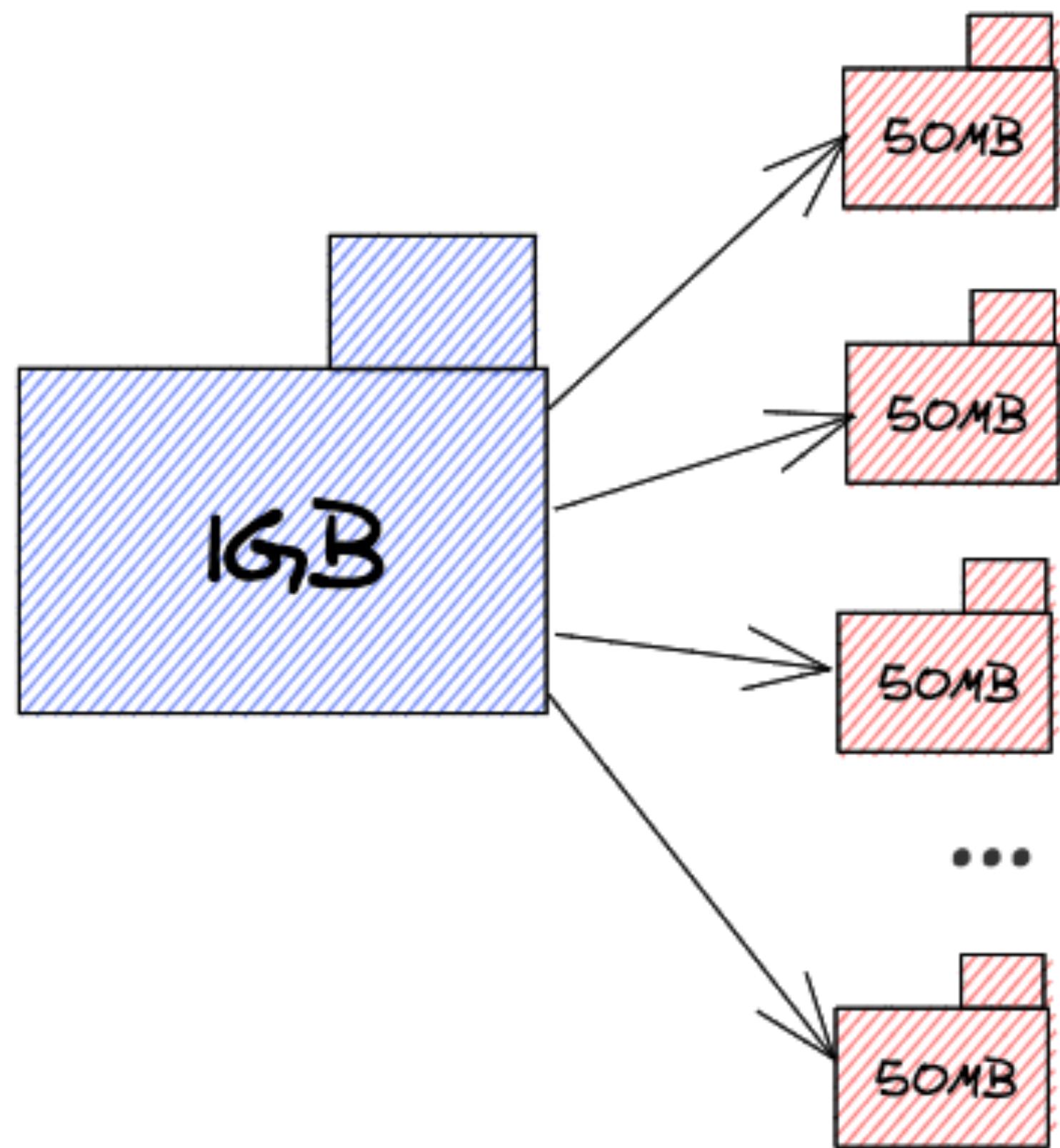
SPLIT FILE IN CHUNKS

FOR EACH CHUNK

SEND HTTP PUT W/ RANGE HEADER

Cargas en Pedazos

chunked uploads



Request

PUT \$SIGNED_URL

Content-Length: 19922944

Content-Range: bytes 0-19922943/166940429

Response

Http Status: 308

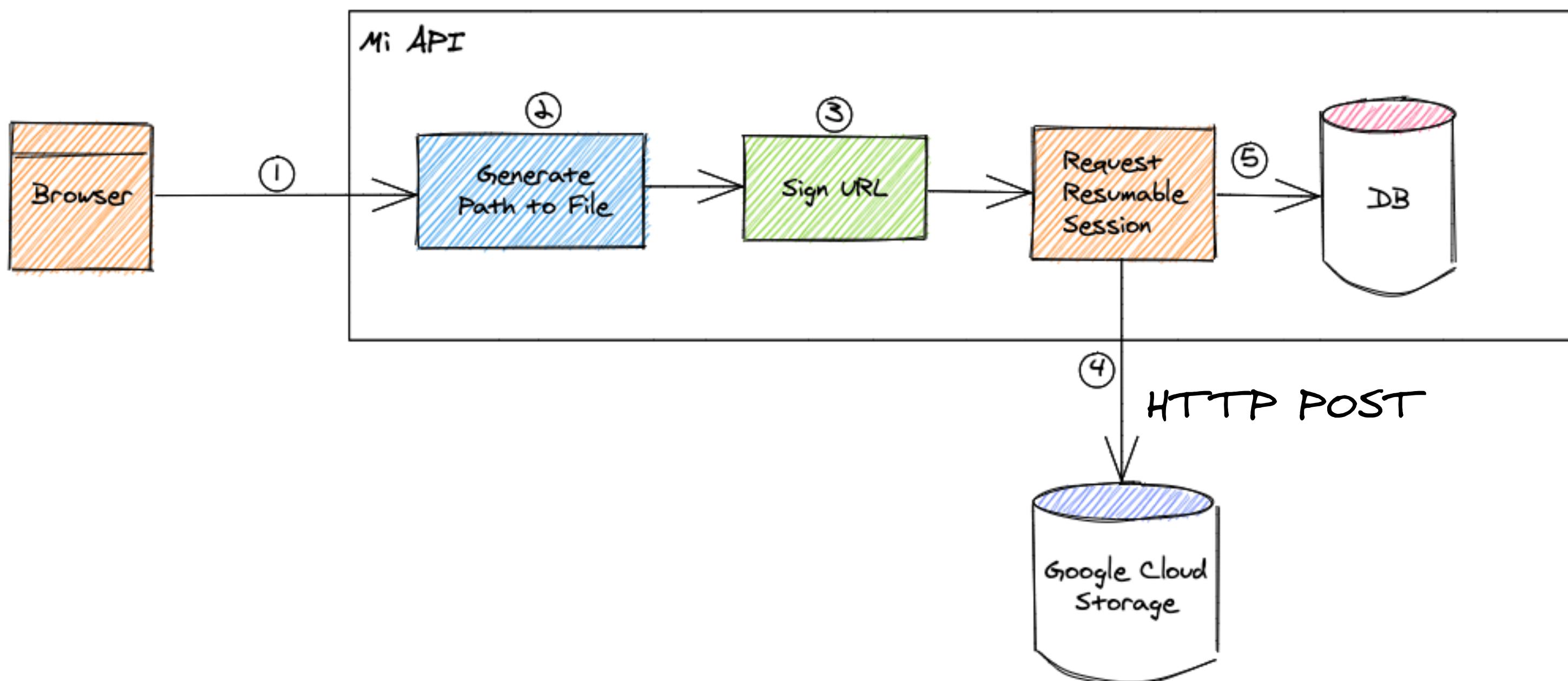
Range: bytes=0-99614719

||

Http Status: 200

Cargas Reanudables

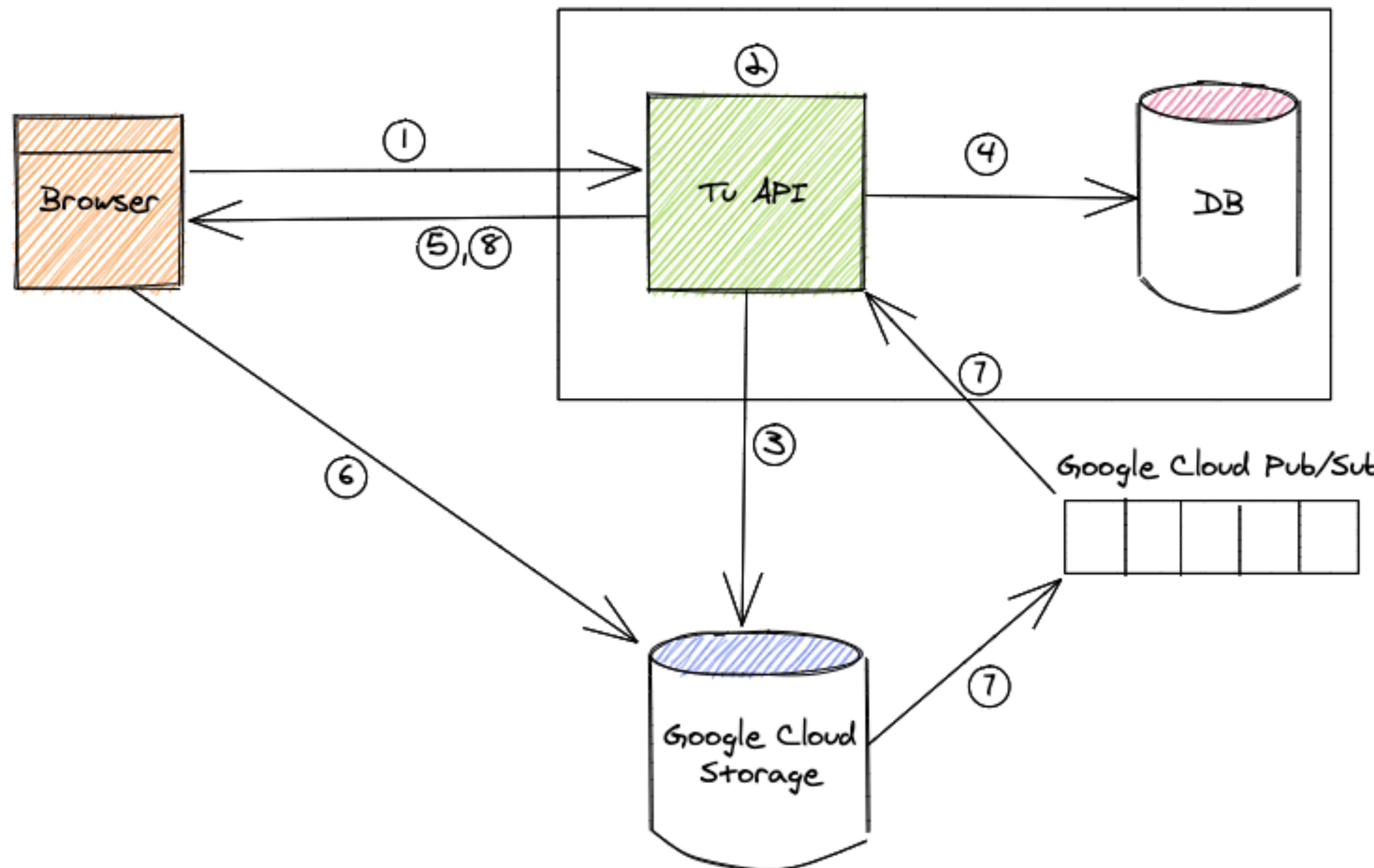
Resumable uploads



- 1) Iniciar petición para Carga de Archivo
OJO: NO se envía el archivo en este paso.
- 2) Generar ruta para tu archivo.
- 3) Cifrar URL
- 4) Petición POST para crear sesión reanudable (válida por 7 días)
- 5) Guardar URL de sesión renovable en DB

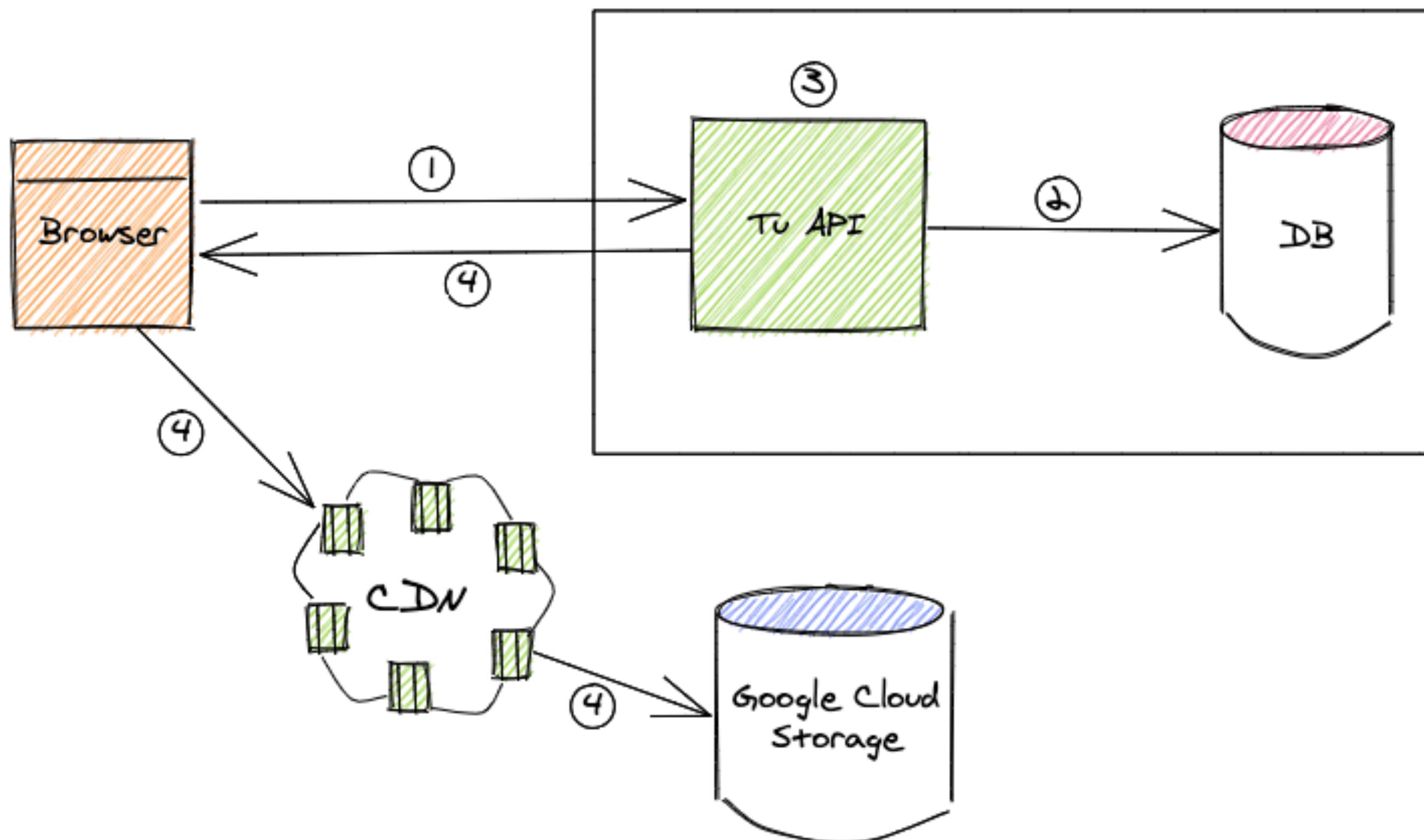
Conclusiones

Diseño Final (upload)



- 1) Petición para URL Cifrada
- 2) Cifrar URL
- 3) Petición para sesión renovable
- 4) Almacenar Metadatos
- 5) Respuesta con URL de sesión
- 6) Carga de archivo directa a GCS
- 7) Notificación de fin de carga
- 8) Notificación al cliente

Diseño Final (download)



1) Petición para URL Cifrada (Lectura)

2) Obtener ruta de archivo en GCS

3) Cifrar URL

4) Redirección (HTTP 302) a CDN

Ventajas

- Diseño altamente escalable a cualquier tamaño de archivo.
- Nuestra API sólamente se encarga de cifrar URLs y no de trasladar bytes.
- Latencia reducida al subir y bajar archivos directamente desde la nube..

Desventajas

- Flujo triangulado, más complicado.
- Cargas en pedazos introducen latencia.
- Notificaciones de Pub/Sub introducen latencia.

Demo

Fin