



# Google Cloud

---

Construyendo un  
DataWarehouse

# Temario

---

## El almacén de datos moderno

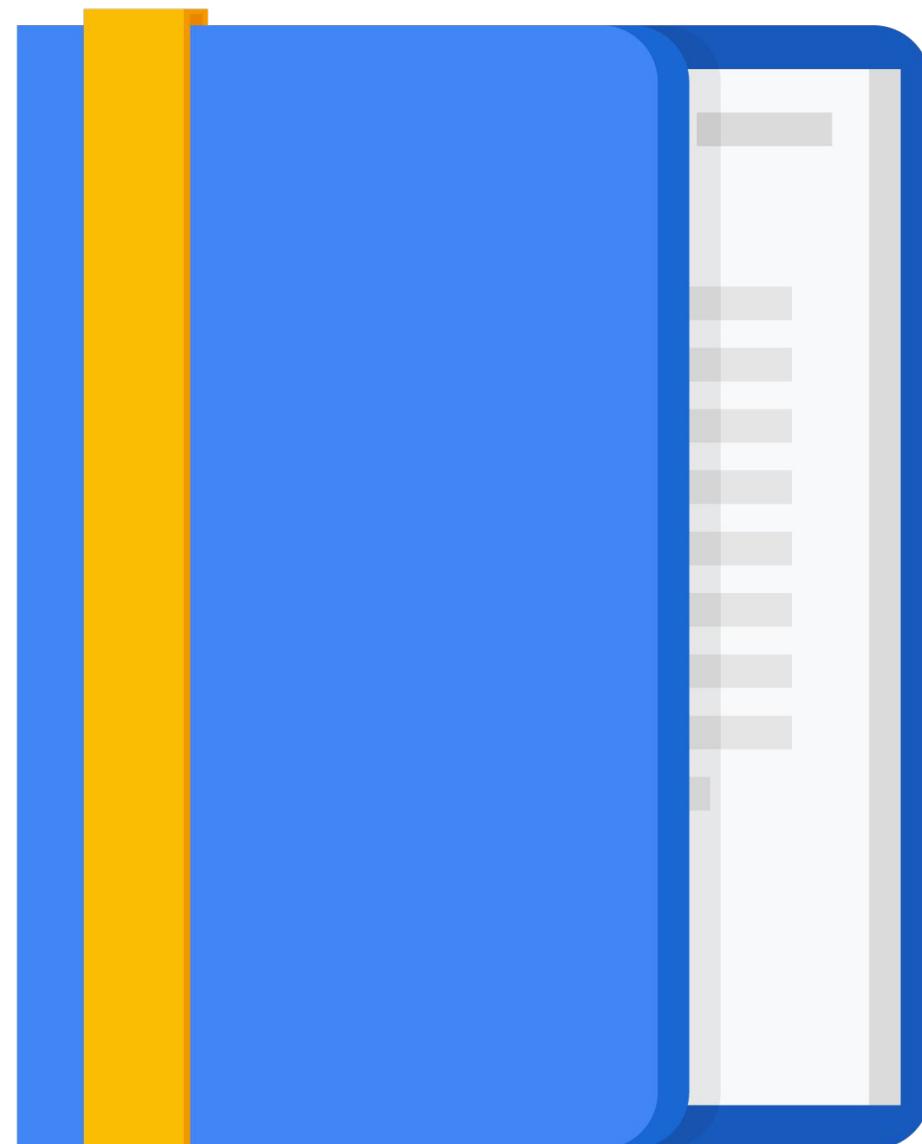
### Introducción a BigQuery

- Introducción
- Carga de datos
- Lab: Cómo cargar datos con Console y CLI

### Exploración de esquemas

- Diseño del esquema
- Optimización mediante la partición y el agrupamiento en clústeres

### Vista previa: Transformación de datos por lotes y datos de transmisión



# ¿Qué es BigQuery?

El almacén de datos empresariales para análisis de Google Cloud

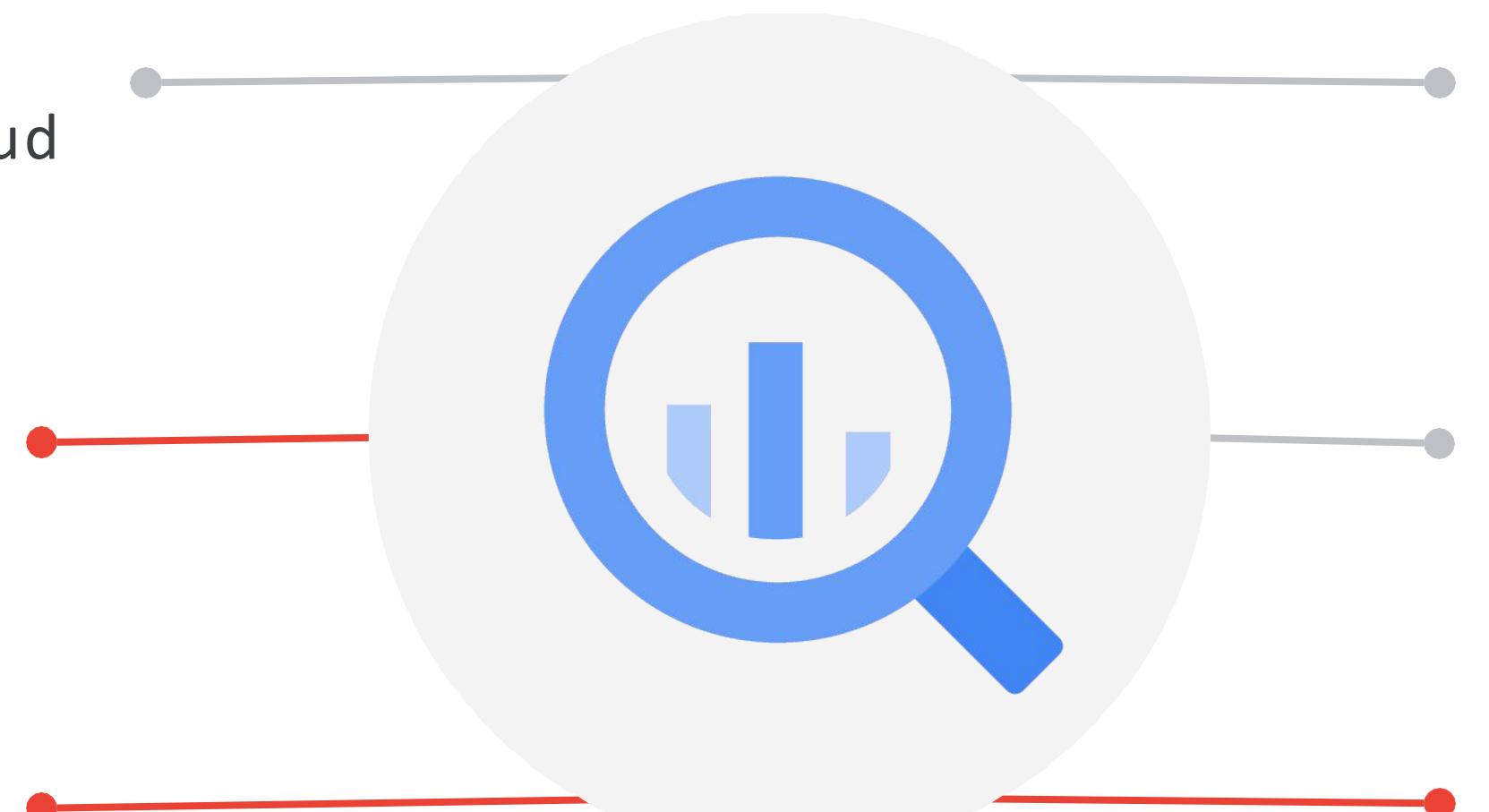
AA y GIS incorporados  
Función exclusiva

Completamente administrado  
y sin servidores  
Función exclusiva

Almacenamiento en escala de gigabytes a petabytes y consultas de SQL

Encriptado, duradero y con alta disponibilidad

Análisis en tiempo real para datos de transmisión Función exclusiva



# BigQuery es un almacén de análisis de datos a escala de petabytes



BigQuery

## 1. Almacén de datos completamente administrado

No-ops, escala de petabytes

## 2. Confiable

Respaldado por los centros de datos de Google

## 3. Económico

Solo paga por el procesamiento y la infraestructura que usa

# BigQuery es un almacén de análisis de datos a escala de petabytes



BigQuery

## 4. Seguro

LCA de funciones,  
datos encriptados en  
tránsito y en reposo

## 5. Auditabile

Cada transacción se  
registra y se puede  
consultar

## 6. Escalable

El modelo de  
procesamiento muy  
paralelo permite  
consultas rápidas

# BigQuery es un almacén de análisis de datos a escala de petabytes



BigQuery

## 4. Flexible

Combine datos a través de múltiples conjuntos de datos

## 5. Fácil de usar

SQL conocido, sin índices, estándares abiertos

## 6. Conjuntos de datos públicos

Explore y practique con conjuntos de datos reales (NOAA, IRS, GitHub, NYC Taxi, etcétera)

# Tres formas de crear interfaces con BigQuery

## 01

### IU web

Cree, valide y ejecute consultas rápidamente a través de la IU web.

Este será el enfoque principal del curso.

## 02

### Interfaz de línea de comandos (CLI)

Use Cloud Shell o el SDK de Google Cloud (`gcloud`) para interactuar a través de una terminal.

```
bq mk [DATASET_ID]
```

## 03

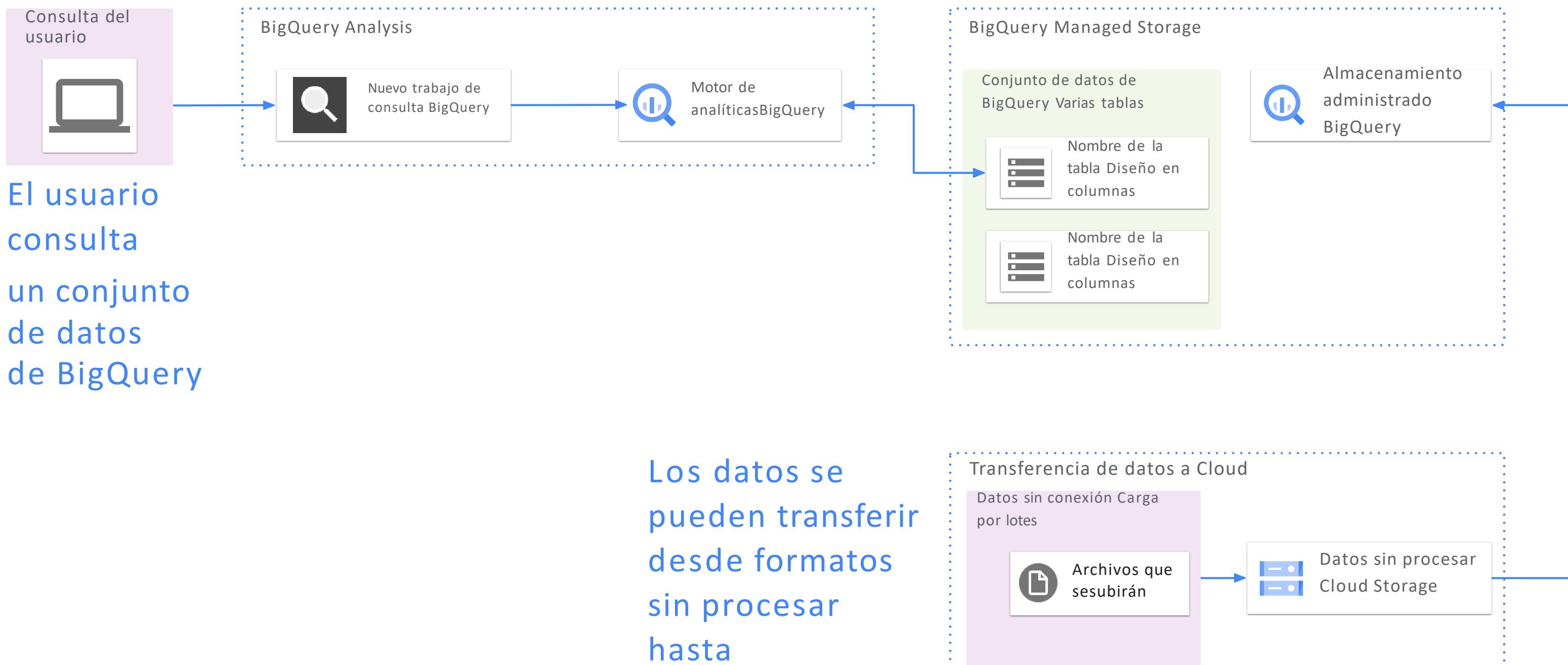
### API de REST

Ejecute consultas de forma programática en lenguajes como Java y Python por HTTP.

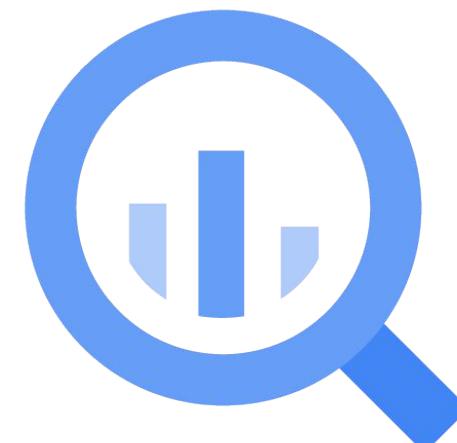
GET

<https://www.googleapis.com/bigquery/v2/projects/projectId/queries/jobId>

# Crear y consultar conjuntos de datos: terminología de BigQuery



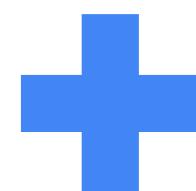
# BigQuery: dos servicios en uno



BigQuery

## BigQuery Managed Storage

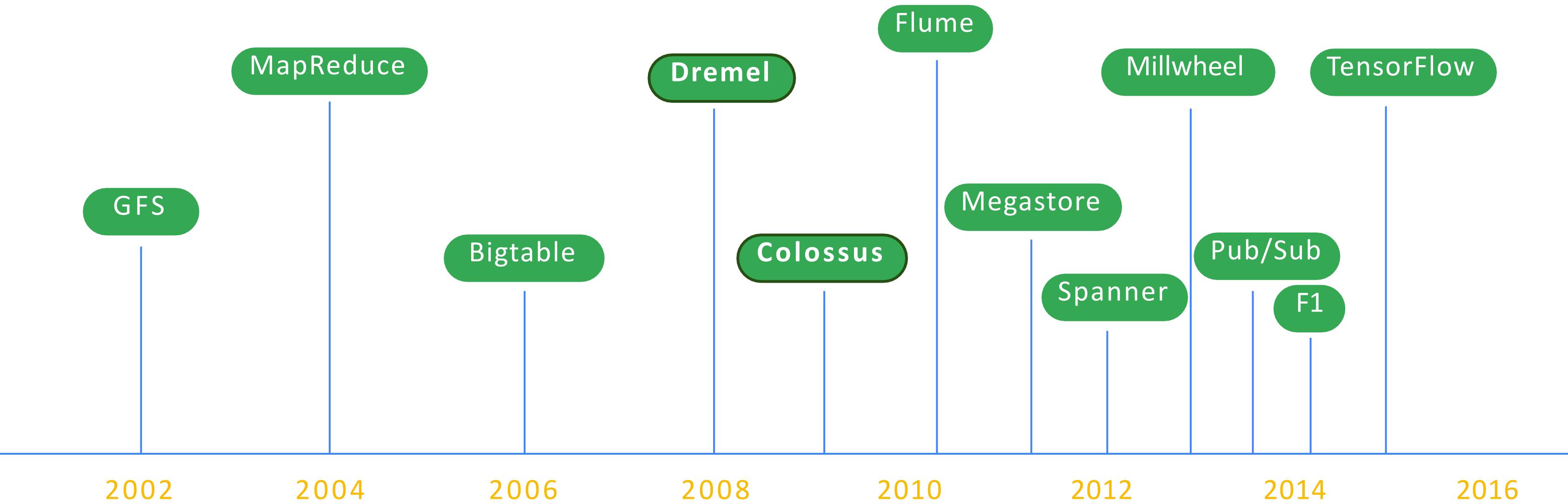
*Almacenamiento de datos completamente administrado y escalable* basado en la misma tecnología que almacena los datos de los productos de Google (Ads, Gmail, etcétera)



## BigQuery Analysis

*Motor de SQL rápido y muy paralelo basado en la tecnología de Dremel, el motor interno de consultas de Google*

# Google innova en las tecnologías de datos



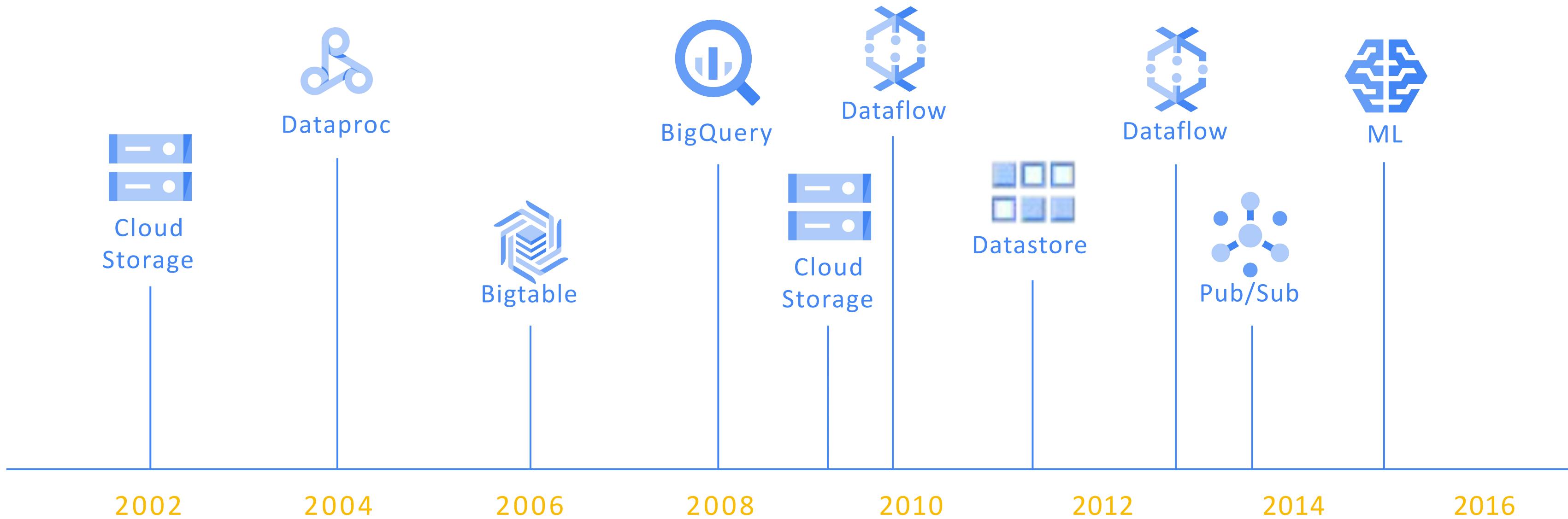
Las publicaciones de Google Research que se usaron como referencia se pueden encontrar aquí:

<http://research.google.com/pubs/papers.html>

The Datacenter as a Computer: [An Introduction to the Design of Warehouse-Scale Machines, 2009](#)

<http://research.google.com/pubs/pub35290.html>

# Google Cloud le ofrece a usted esa innovación



Las publicaciones de Google Research que se usaron como referencia se pueden encontrar aquí:

<http://research.google.com/pubs/papers.html>

The Datacenter as a Computer: [An Introduction to the Design of Warehouse-Scale Machines, 2009](#)

<http://research.google.com/pubs/pub35290.html>

# BigQuery no requiere servidores. No debe preocuparse por:

X Vencimiento de los datos

X Copias de seguridad

X Administración del almacenamiento

X Recuperación ante errores

X Optimización del motor de consultas

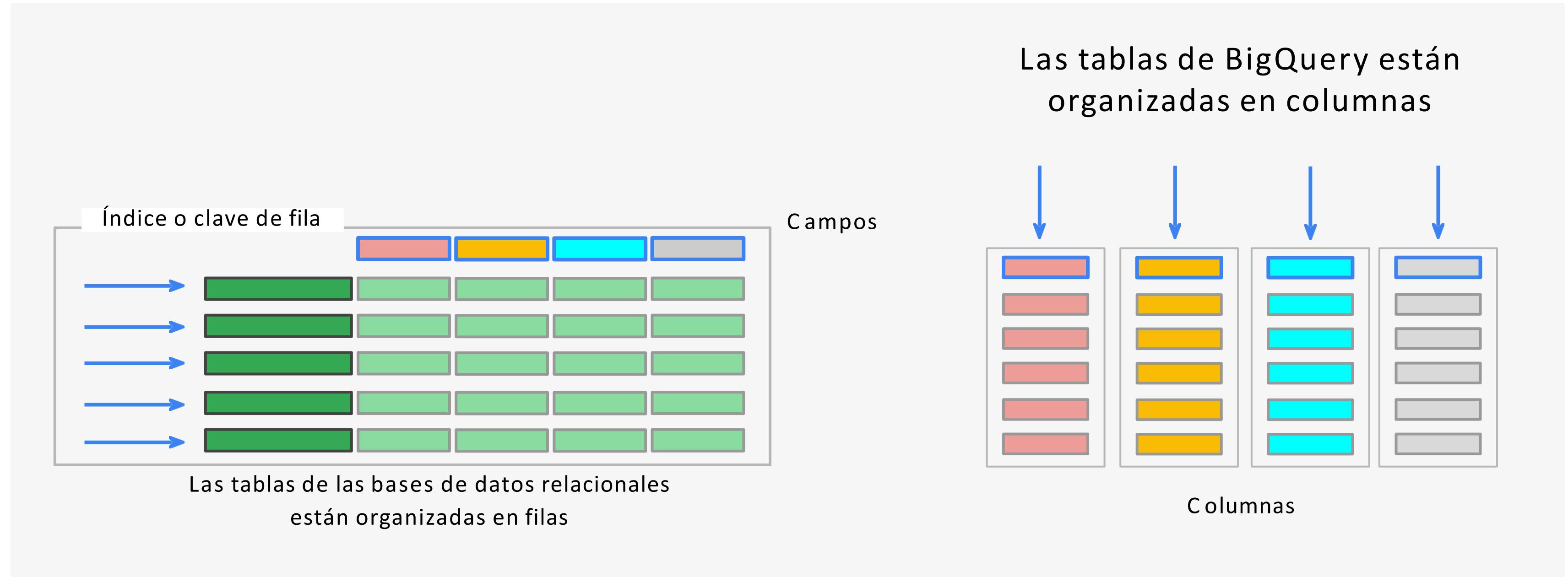
X Hardware

X Actualizaciones

Tendrá más horas de trabajo reales, ya que  
no deberá preocuparse por tareas comunes

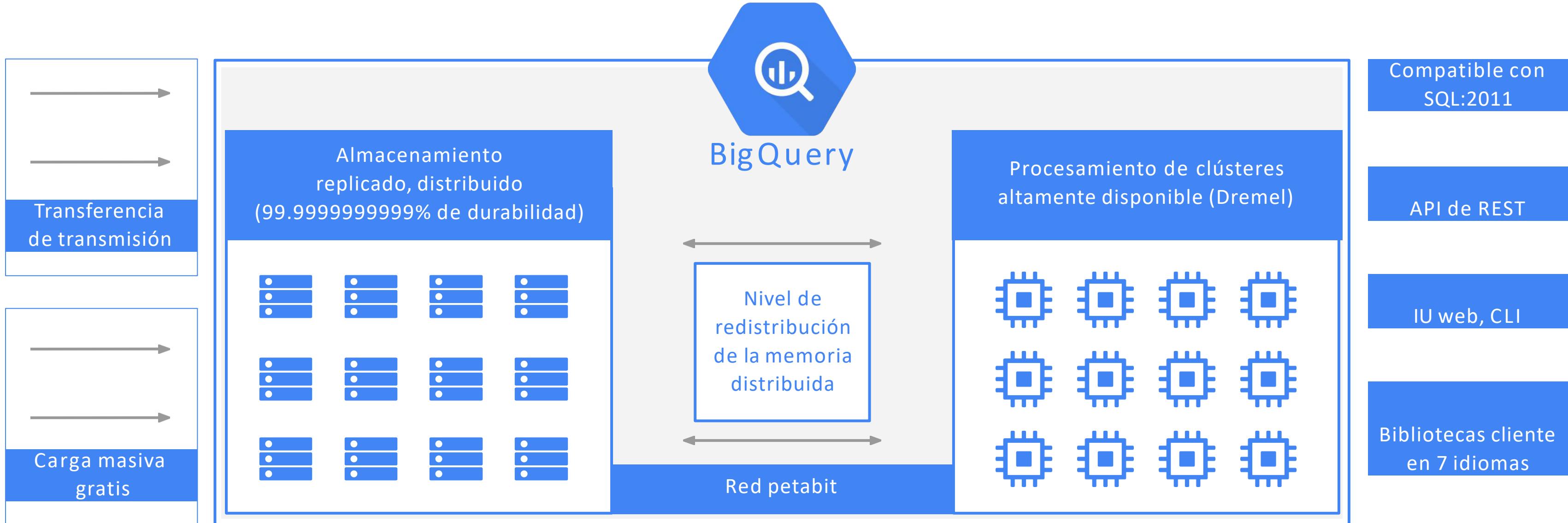


# ¿Qué hace que BigQuery sea rápido?

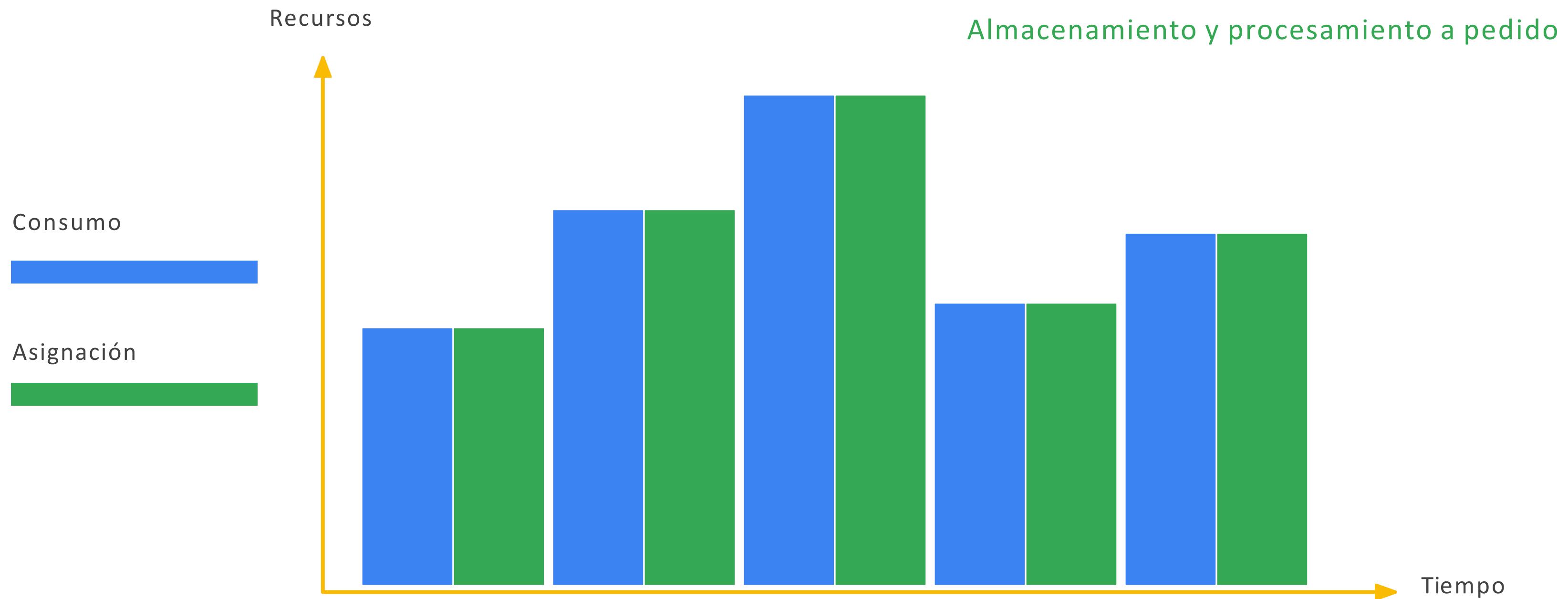


# Los datos se almacenan físicamente de forma redundante y separados del clúster de proceso

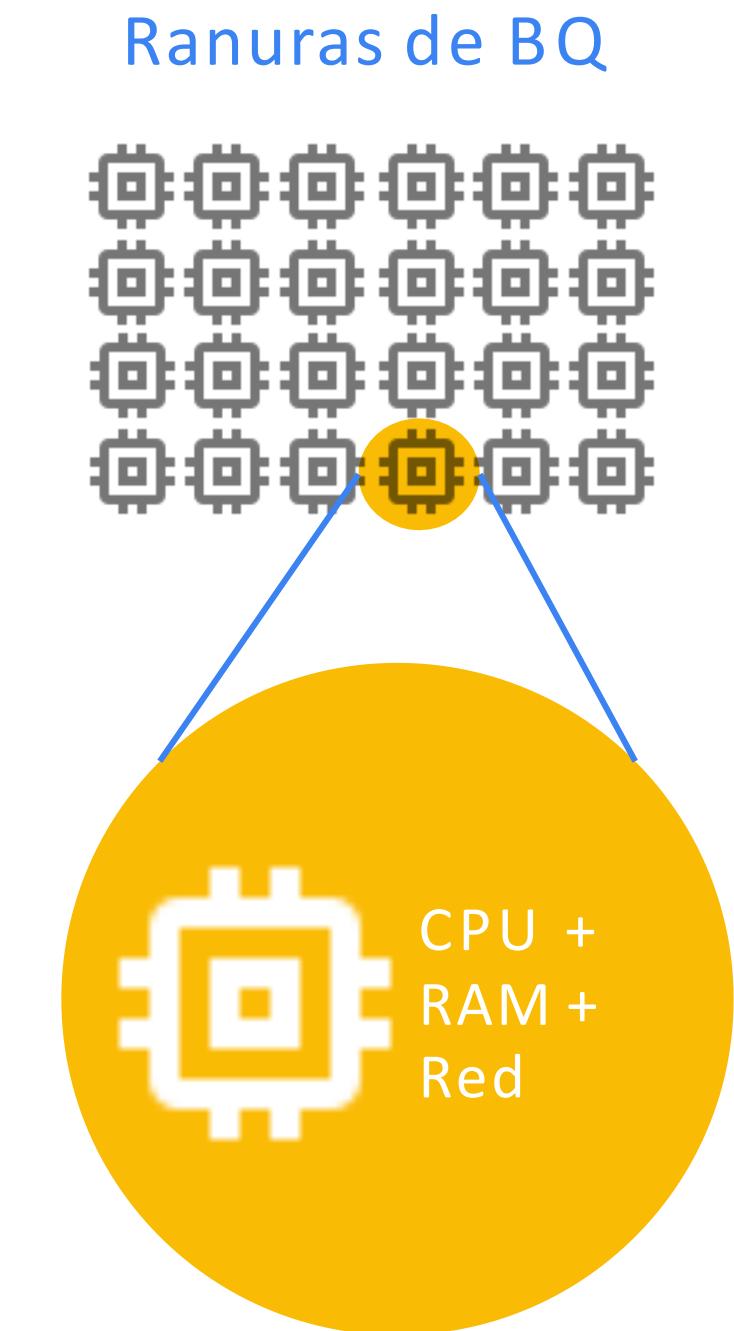
## Colossus Jupiter



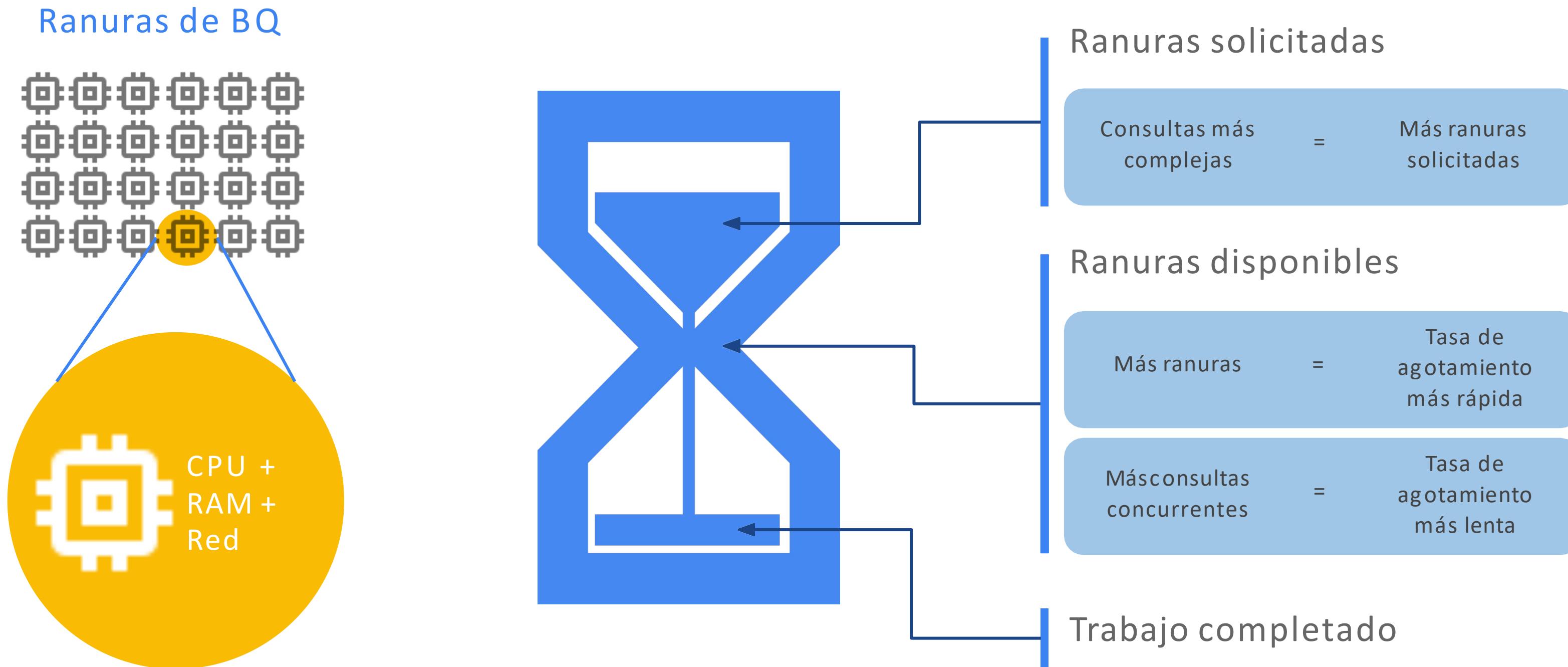
# No necesita aprovisionar recursos antes de usar BigQuery



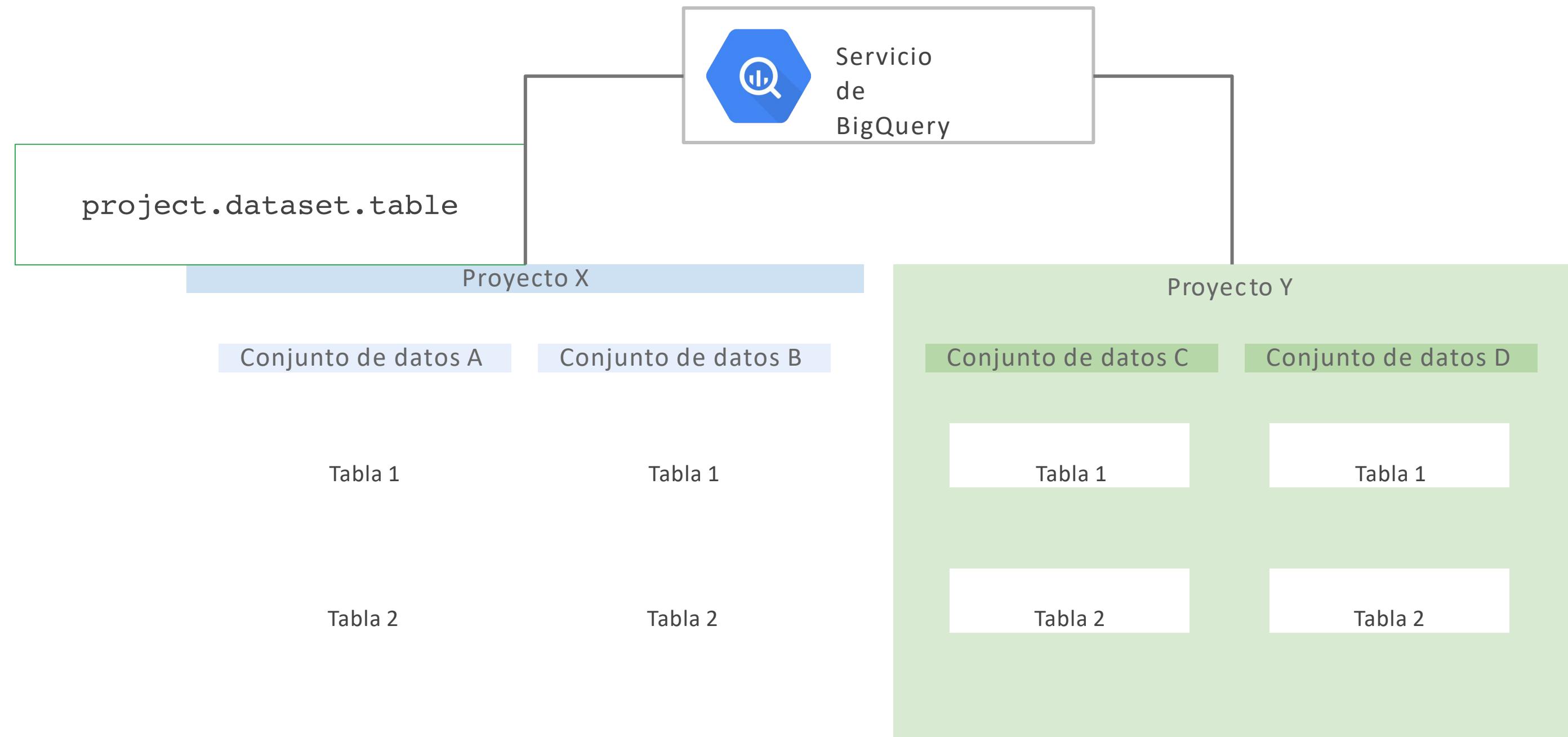
Una ranura de BigQuery es una combinación de CPU, memoria y recursos de red



# La cantidad real de ranuras asignadas a una consulta depende de su complejidad y de la cuota del proyecto

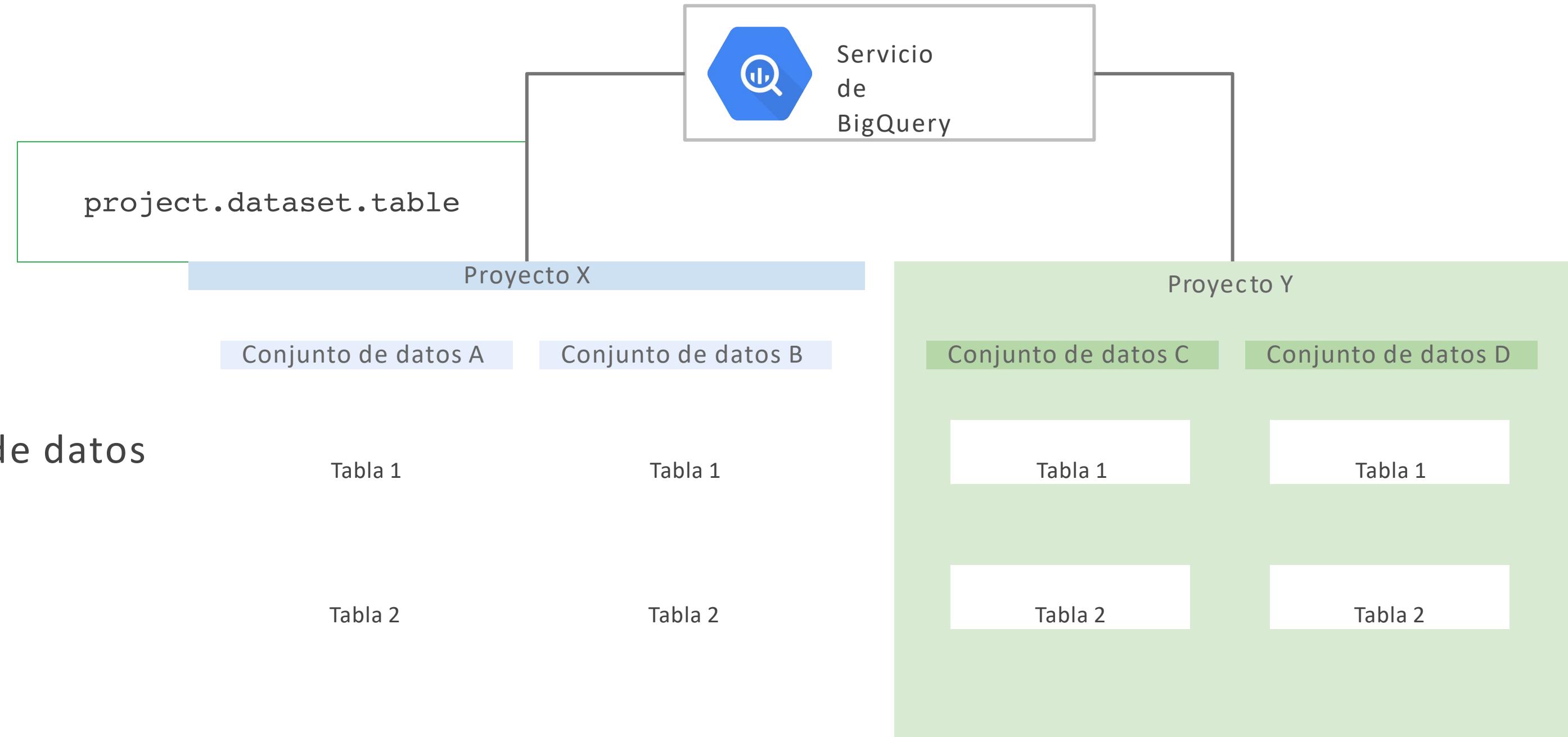


# BigQuery organiza las tablas de datos en unidades llamadas conjuntos de datos

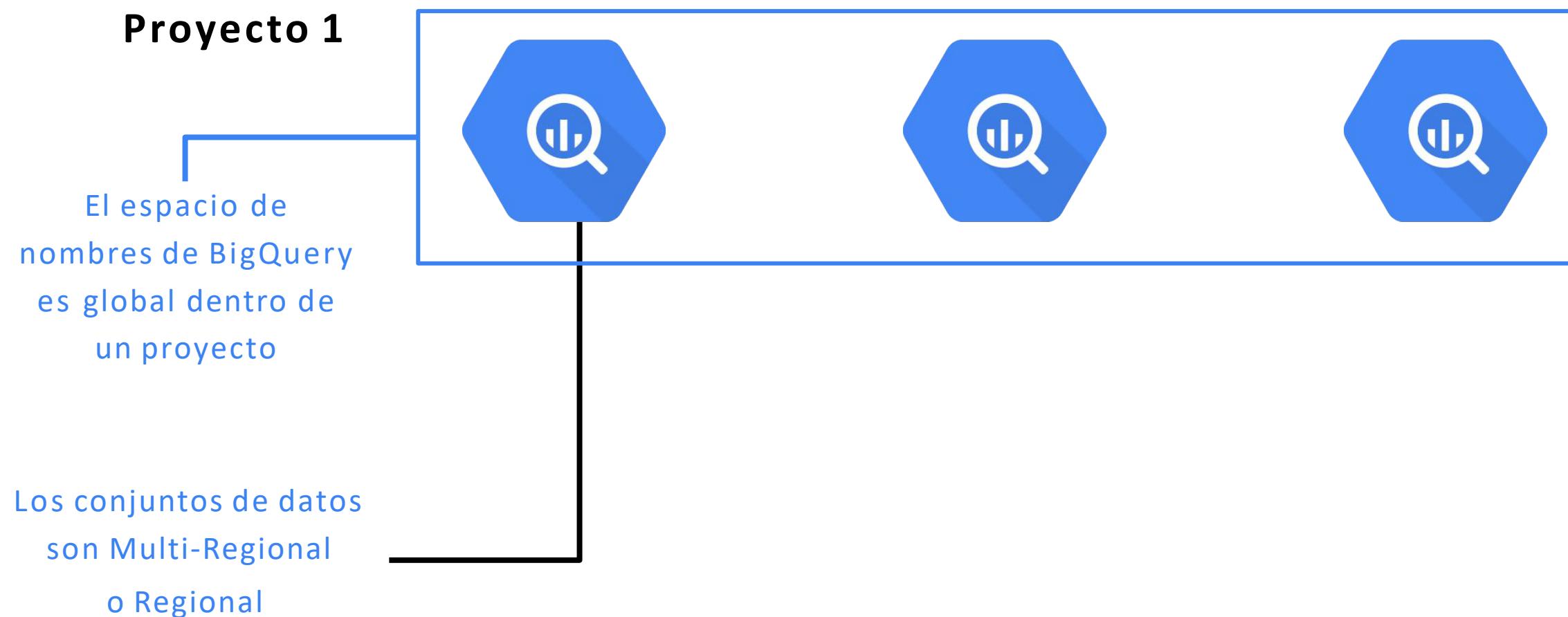


# ¿Cuáles son algunos de los motivos para estructurar su información?

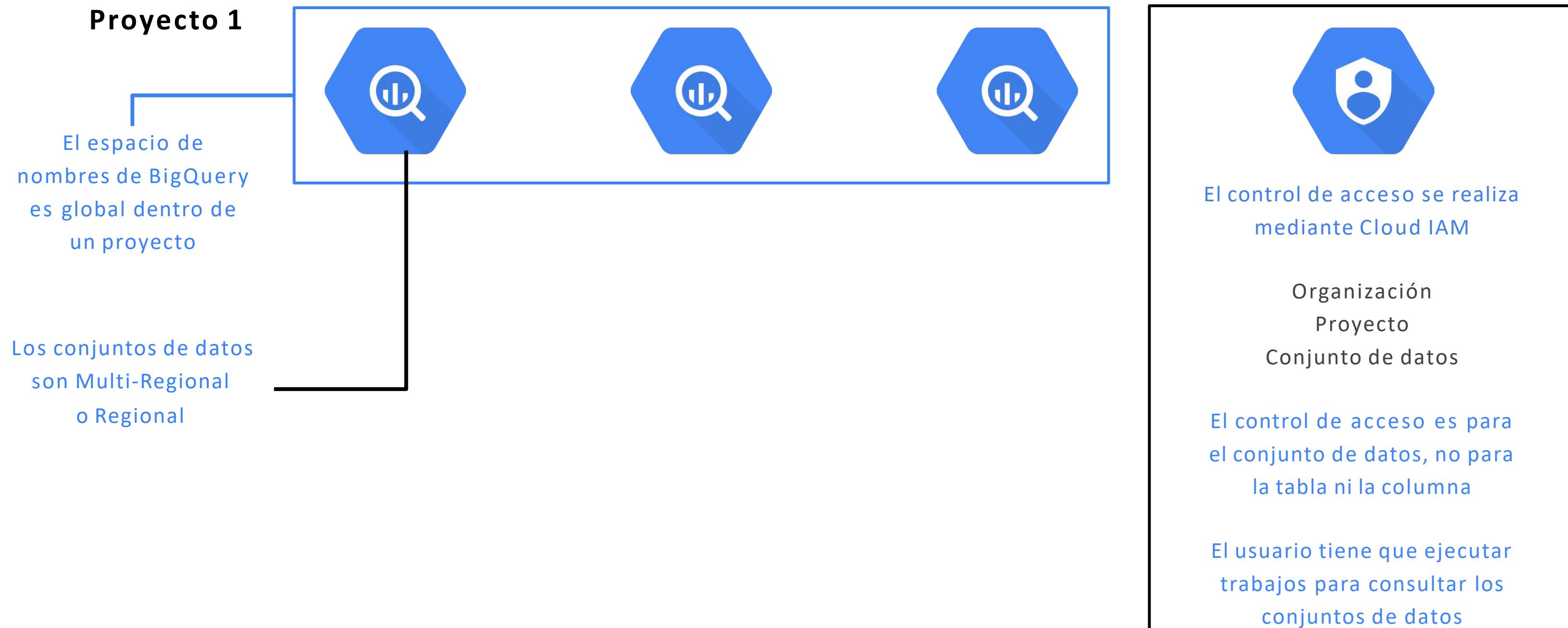
- Proyectos
- Conjuntos de datos
- Tablas



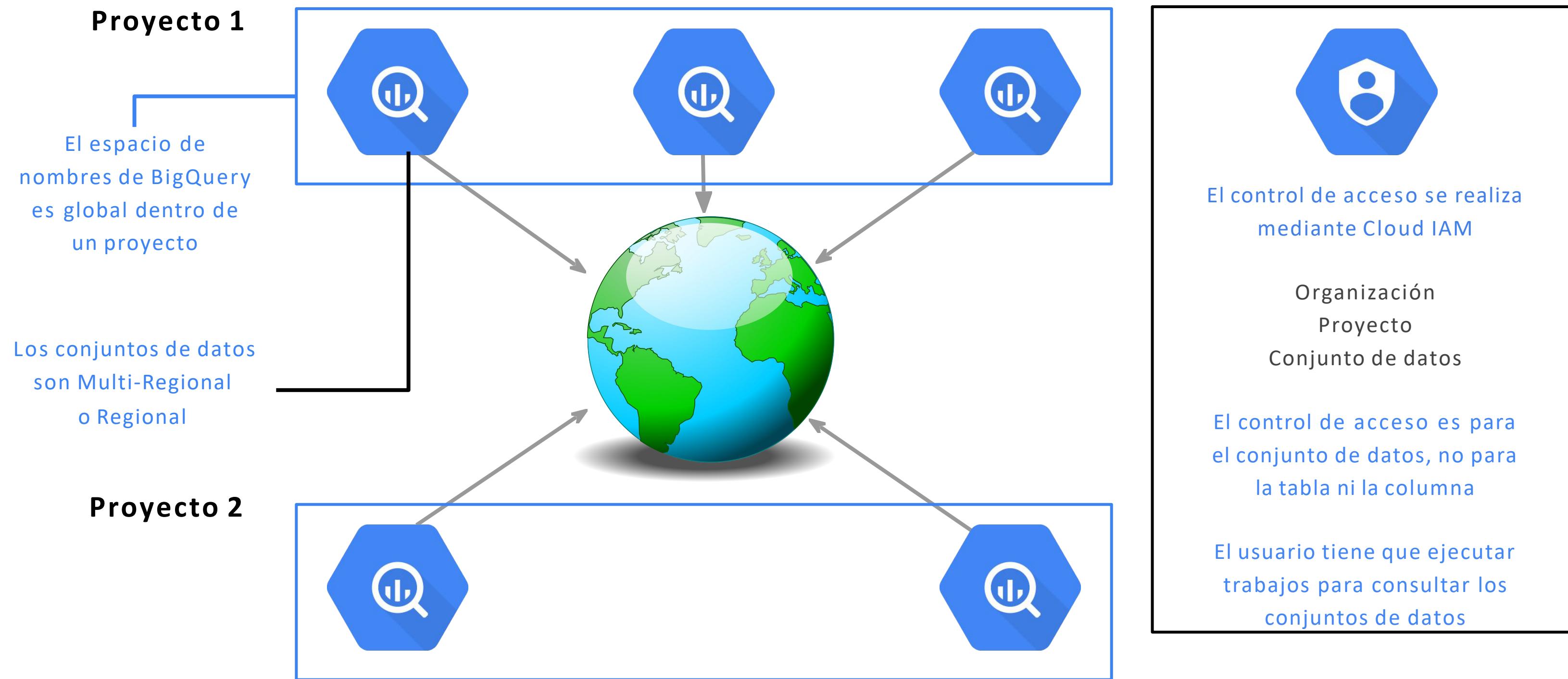
# Los conjuntos de datos de BigQuery pertenecen a un proyecto



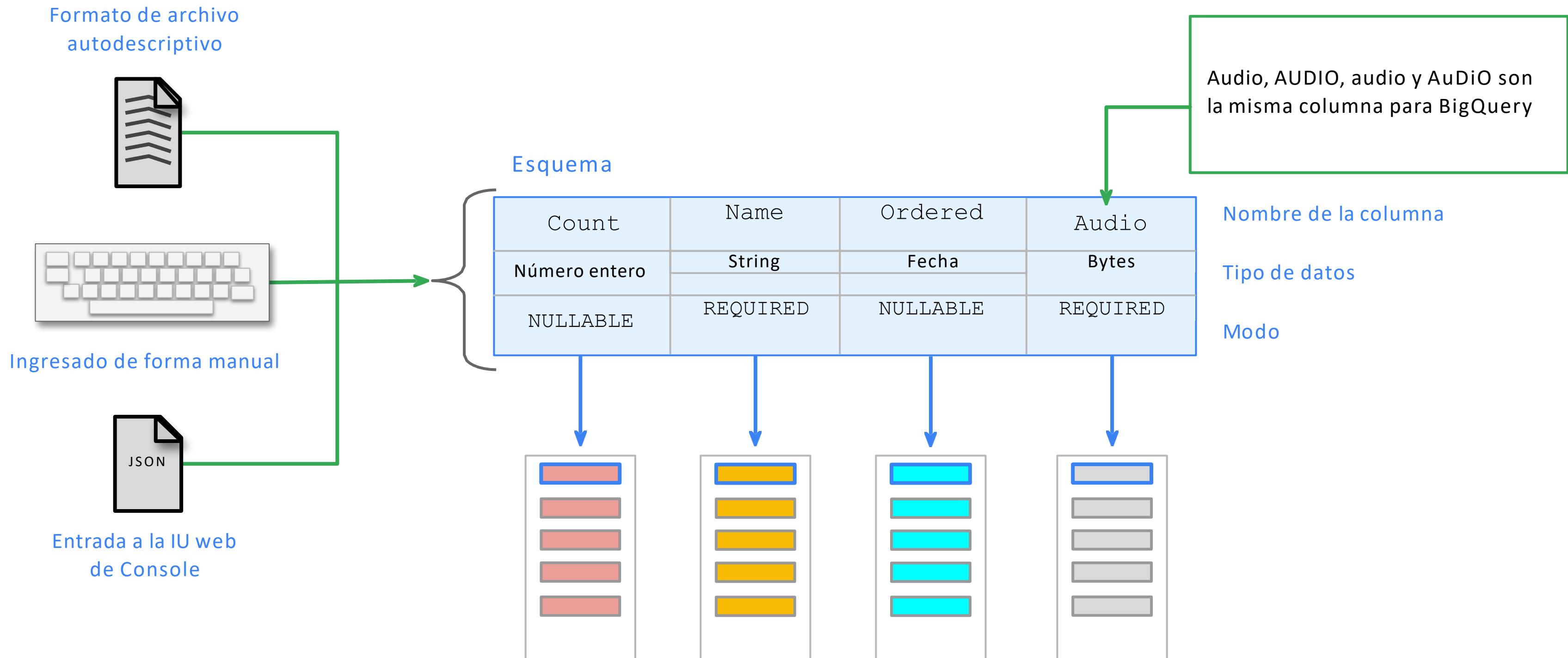
# El control de acceso para ejecutar una consulta se realiza mediante Cloud IAM



# Los conjuntos de datos de BigQuery pueden ser Regional o Multi-Regional



El esquema de la tabla proporciona una estructura a los datos



# El ciclo de vida de una consulta de SQL de BigQuery

