

# Google cloud platform

Programa de formación MLDS - Módulo Big Data



# Google cloud platform (GCP)

- Servicio de computación en la nube de Google.
- Ofrece la misma tecnología usada para soportar sus productos internos (GMail, Youtube, GDrive, etc.)
- APIs disponible para múltiples lenguajes de programación.
- Ventajas de CC:
  - Seguridad
  - Networking
  - Elasticidad
  - Aprovisionamiento automático

# Google cloud platform (GCP)



Compute



Networking



Management Tools



Storage & Databases



Cloud AI



Big Data



Identity & Security



Developer Tools

# Google cloud platform (GCP)



Compute



Storage & Databases



Identity & Security

## Compute

- PaaS
- Ejecución de máquinas virtuales
- Containers (Docker, Kubernetes)



Management Tools



Big Data

# Google cloud platform (GCP)



Compute



Storage & Databases



Identity & Security

## Networking

- Administración de los recursos de GCP
- Firewall
- Enrutamiento
- Balanceo de cargas



Management Tools



Big Data

# Google cloud platform (GCP)



Compute



Storage & Databases



Identity & Security

## Big data

- Análisis de bodegas de datos (BigQuery)
- Sistemas distribuidos “ready-to-use” (Hadoop, Spark)
- Workflows de procesamiento de datos (Dataflow)



Management Tools



Big Data

# Google cloud platform (GCP)



Compute



Storage & Databases



Identity & Security

## Cloud AI

- Procesamiento de lenguaje natural
- Vision por computador
- Sistemas de reconocimiento de voz
- Plataformas para entrenar modelos propios



Management Tools



Big Data

# Google cloud platform (GCP)



Compute



Storage & Databases



Identity & Security

## Identity & security

- Sistemas de autenticación y autorización
- Escaneo automático de vulnerabilidades
- Administración de llaves de encriptación



Management Tools



Big Data



# Google cloud platform (GCP)



Compute



Storage & Databases



Identity & Security

## Management tools

- Monitoreo
- Registro (logging)
- Alertas
- Análisis de rendimiento
- Facturación



Management Tools



Big Data

# Google cloud platform (GCP)



Compute



Storage & Databases



Identity & Security

## Developer tools

- Librerías para administración de los recursos GCP
- Plugins para IDEs en diferentes lenguajes
- Sistemas de control de versiones
- Sistemas de integración continua



Management Tools



Big Data

# Google cloud platform (GCP)



Compute



Storage & Databases



Identity & Security

## Storage and databases

- Bases de datos relacionales
- Bases de datos NoSQL
- Almacenamiento de archivos
- Discos duros para uso de VMs





Management Tools





Big Data

# Storage and databases

PRODUCT	DESCRIPTION	GOOD FOR	COMMON WORKLOADS
 <u>Persistent Disk</u>	Fully-managed, price-performant block storage that is suitable for virtual machines and containers.	<ul style="list-style-type: none"><li>• Block storage for Google Compute Engine and Google Container Engine</li><li>• Snapshots for data backup</li></ul>	<ul style="list-style-type: none"><li>• Disks for virtual machines</li><li>• Sharing read-only data across multiple virtual machines</li><li>• Rapid, durable backups of running virtual machines</li></ul>
 <u>Google Cloud Storage</u>	A scalable, fully-managed, highly reliable, and cost-efficient object / blob store.	<ul style="list-style-type: none"><li>• Images, pictures, and videos</li><li>• Objects and blobs</li><li>• Unstructured data</li></ul>	<ul style="list-style-type: none"><li>• Storing and streaming multimedia</li><li>• Storage for custom data analytics pipelines</li><li>• Archive, backup, and disaster recovery</li></ul>



# Storage and databases

PRODUCT	DESCRIPTION	GOOD FOR	COMMON WORKLOADS
 <u>Google Cloud</u> <u>Bigtable</u>	A scalable, fully-managed NoSQL wide-column database that is suitable for both real-time access and analytics workloads.	<ul style="list-style-type: none"><li>• Low-latency read/write access</li><li>• High-throughput analytics</li><li>• Native time series support</li></ul>	<ul style="list-style-type: none"><li>• IoT, finance, adtech</li><li>• Personalization, recommendations</li><li>• Monitoring</li><li>• Geospatial datasets</li><li>• Graphs</li></ul>
 <u>Google Cloud</u> <u>Datastore</u>	A scalable, fully-managed NoSQL document database for your web and mobile applications.	<ul style="list-style-type: none"><li>• Semi-structured application data</li><li>• Hierarchical data</li><li>• Durable key-value data</li></ul>	<ul style="list-style-type: none"><li>• User profiles</li><li>• Product catalogs</li><li>• Game state</li></ul>

# Storage and databases

PRODUCT	DESCRIPTION	GOOD FOR	COMMON WORKLOADS
 <u>Google Cloud SQL</u>	A fully-managed MySQL database service that is built on the strength and reliability of Google's infrastructure.	<ul style="list-style-type: none"><li>• Web frameworks</li><li>• Structured data</li><li>• OLTP workloads</li></ul>	<ul style="list-style-type: none"><li>• Websites, blogs, and content management systems (CMS)</li><li>• Business Intelligence (BI) applications</li><li>• ERP, CRM, and eCommerce applications</li></ul>
 <u>Google Cloud Spanner</u>	Mission-critical, relational database service with transactional consistency, global scale and high availability.	<ul style="list-style-type: none"><li>• Mission-critical applications</li><li>• High transactions</li><li>• Scale + Consistency requirements</li></ul>	<ul style="list-style-type: none"><li>• Adtech</li><li>• Financial services</li><li>• Global supply chain</li><li>• Retail</li></ul>

# Storage and databases

PRODUCT	DESCRIPTION	GOOD FOR	COMMON WORKLOADS
 <u>Google BigQuery</u>	A scalable, fully-managed Enterprise Data Warehouse (EDW) with SQL and fast response times.	<ul style="list-style-type: none"><li>• OLAP workloads up to petabyte-scale</li><li>• Big Data exploration and processing</li><li>• Reporting via Business Intelligence (BI) tools</li></ul>	<ul style="list-style-type: none"><li>• Analytical reporting on large data</li><li>• Data Science and advanced analyses</li><li>• Big Data processing using SQL</li></ul>
 <u>Google Drive</u>	A collaborative space for storing, sharing, and editing files, including Google Docs.	<ul style="list-style-type: none"><li>• End-user interaction with docs and files</li><li>• Collaborative creation and editing</li><li>• Syncing files between cloud and local devices</li></ul>	<ul style="list-style-type: none"><li>• Access files from anywhere through web, apps, and sync clients</li><li>• Create and work on documents with coworkers</li><li>• Backup photos and media</li></ul>





# Google Cloud DataStore (Cloud Firestore)

Google Cloud Datastore es una base de datos No-SQL basada en documentos.  
Incluye:

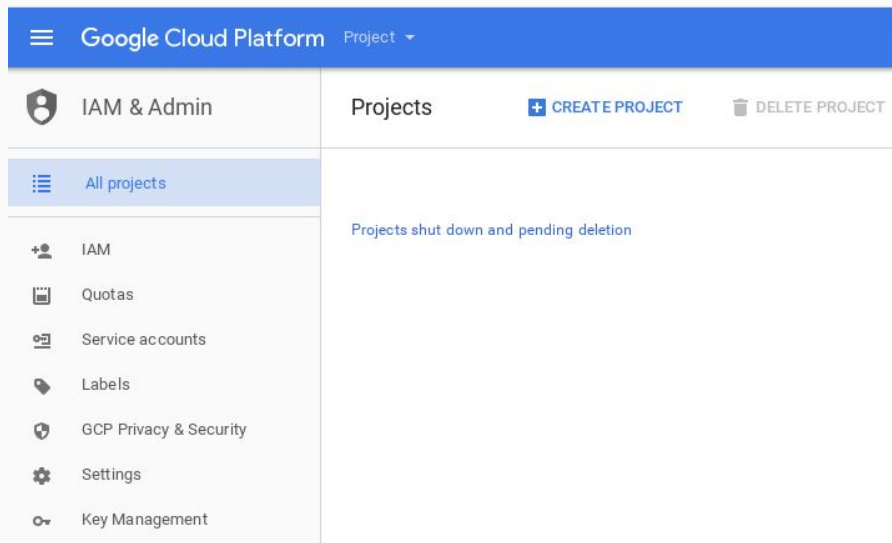
- Operaciones atómicas limitadas
- Escalamiento masivo
- Modificación dinámica del modelo de datos.

# Google Cloud DataStore

Concepto	Cloud Datastore	Base de datos relacional
Categoría de un objeto	Kind	Tabla
Un objeto	Entity	Registro
Dato de un objeto	Property	Campo
Identificador único de un objeto	Key	Llave primaria

# Crear el proyecto

Ingresa a <http://console.cloud.google.com>



The screenshot shows the Google Cloud Platform console. The top navigation bar is blue with the Google Cloud Platform logo and a 'Project' dropdown menu. Below this, there's a sidebar with various management tools: IAM & Admin, IAM, Quotas, Service accounts, Labels, GCP Privacy & Security, Settings, and Key Management. The main content area is titled 'Projects' and features a '+ CREATE PROJECT' button and a 'DELETE PROJECT' button. Below these buttons, it states 'Projects shut down and pending deletion'.

## New Project

Project name ?

Your project ID will be prueba-datastore-160320 ? Edit

CANCEL CREATE

# Seleccionar modo Cloud Firestore

You selected Cloud Firestore in Datastore mode. Now choose a database location.

The location of your database affects its cost, availability, and durability. Choose a regional location or a multi-region location (higher availability, higher cost). [Learn more](#)



Choose carefully: your location selection is permanent and will also apply to this project

Select a location

nam5 (United States)



To improve performance, store your data close to the users and services that need it

## Native mode

Enable all of Cloud Firestore's features, with offline support and real-time synchronization.

SELECT NATIVE MODE

## Datastore mode

Leverage Cloud Datastore's system behavior on top of Cloud Firestore's powerful storage layer.

SELECT DATASTORE MODE

	Native mode	Datastore mode
API	Firestore	Datastore
Scalability	Automatically scales to millions of concurrent clients	Automatically scales to millions of writes per second
App engine support	Not supported in the App Engine standard Python 2.7 and PHP 5.5 runtimes	All runtimes
Max writes per second	10,000	No limit
Real-time updates	✓	✗
Mobile/web client libraries with offline data persistence	✓	✗

# Crear una entidad

Google Cloud Platform prueba-datastore

**Datastore**

- Entities
- Dashboard
- Indexes
- Admin

**Entities** + CREATE ENTITY

QUERY BY KIND QUERY BY GQL

**i** Since your database is empty, you can

← Create an entity

Namespace  
[default]

Kind  
Libro

Key identifier  
Numeric ID (auto-generated)

▼ SPECIFY PARENT

Properties

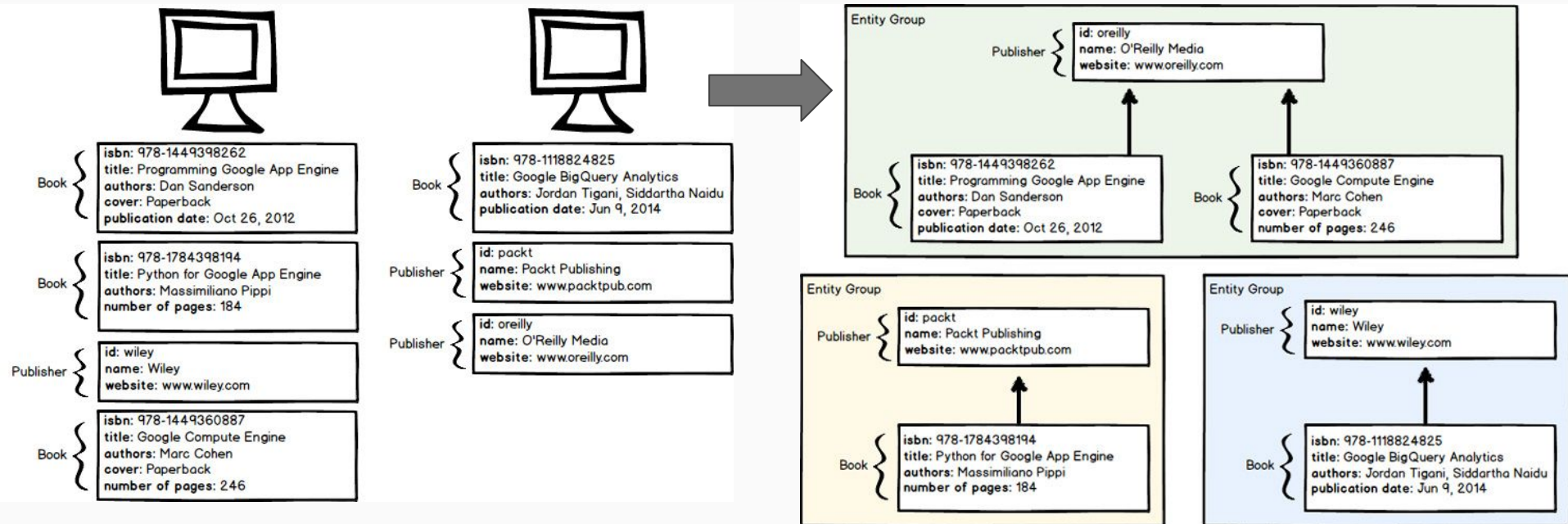
titulo: Cien años de soledad	Indexed	🗑️ ▼
idioma: español	Indexed	(Not saved) ▼
editorial: harper	Indexed	(Not saved) ▼
año: 1967	Indexed	(Not saved) ▼

ADD PROPERTY

# Ejercicio 1.

- Cree otra entidad y agregue una propiedad nueva.
- Usando [GQL](#), consulte busque registros por nombre.
- Haga una consulta por una propiedad que no está indexada. ¿Qué sucede?

# Transacciones con Entity Groups



# Limitaciones

- No incluye soporte para *join* o filtrado en los resultados de una subconsulta.
- No se permiten agrupaciones.
- Las transacciones no pueden involucrar más de 25 grupos de entidades.
- El rendimiento para un grupo de entidades es una transacción por segundo.



# Google Cloud DataStore

## Bueno para:

- Catálogos de productos que proporcionan inventario en tiempo real
- Transacciones basadas en ACID.
- Perfiles de usuario con registro de actividades y preferencias.

## No conveniente para:

- Soporte completo para alta transaccionalidad (Use MySQL)
- Almacenar objetos binarios de gran tamaño (Use Cloud Storage)
- Cuando ACID no es un requisito (Use Cloud BigTable)

# Ejercicio 2

- Cree 3 entidades para cada uno de los tipos que se describen en el modelo de datos
- Liste las respuestas de una pregunta en particular
- ¿Cuál esquema jerárquico es útil en este modelo de datos?, ¿Por qué?

Usuario	
PK	<u>id</u>
	nombre
	genero
	fecha_nac

Pregunta	
PK	<u>id</u>
	fecha_cre
	texto
	nb_respuestas
	usuario_id

Respuesta	
PK	<u>id</u>
	fecha_cre
	usuario_id
	pregunta_id
	texto