



**Síntesis académica de herramientas de Google Cloud para el manejo y análisis de
datos en la nube**

Juan Sebastián Espinosa Uribe

José Luis Mena Palomeque

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesora

María Bernarda Salazar Sánchez, Doctora (Ph.D.) en Ingeniería Electrónica

Universidad de Antioquia

Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2025

Cita	(Espinosa Uribe & Mena Palomeque, 2025)
Referencia	Espinosa Uribe, J. S., & Mena Palomeque, J. L. (2025). <i>Síntesis académica de herramientas de Google Cloud para el manejo y análisis de datos en la nube</i> [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia.
Estilo APA 7 (2020)	



Especialización en Analítica y Ciencia de Datos, Cohorte VIII.

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Danny Alejandro Múnera Ramírez

Coordinadora del Programa: Maria Bernarda Salazar Sánchez

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Tabla de contenido

Resumen	6
1. Introducción	7
2. Justificación académica.....	8
3. Descripción técnica del curso.....	9
4. Herramientas de Google Cloud para la ciencia de datos.....	11
4.1. Herramientas de Google Cloud	11
4.2. Procesamiento de datos no estructurados: Vertex AI.....	14
5. Conclusiones	21
Referencias	22

Lista de Figuras

Figura 1. Ciclo de aprendizaje del curso..	10
Figura 2. Botón “Consola” de la interfaz de Google Cloud Platform.	14
Figura 3. Barra de búsqueda de la interfaz de Google Cloud Platform.....	15
Figura 4. Panel de inicio de Vertex AI.....	15
Figura 5. Opción “Vertex AI Studio” de la plataforma Vertex AI.....	16
Figura 6. Barra de instrucciones de Vertex AI Studio.....	16
Figura 7. Botón “Ver todos” de la galería de instrucciones de Vertex AI Studio	17
Figura 8. Opciones de filtrado y procesamiento según tipo de dato.....	17
Figura 9. Opciones para la importación de archivos multimedia	18
Figura 10. Carga de video a la plataforma mediante URL	18
Figura 11. Área de prompt para instrucciones y ejecución del modelo	19

Siglas, acrónimos y abreviaturas

AI	Artificial Intelligence
SQL	Structured Query Language
BI	Business Intelligence
ACID	Atomicity, Consistency, Isolation, Durability
ML	Machine Learning
GCP	Google Cloud Platform
UdeA	Universidad de Antioquia

Resumen

Este informe es un análisis del curso Data Management and Storage in the Cloud de Google Cloud, en el que se abarcan desde arquitecturas de almacenamiento hasta técnicas avanzadas de consulta de datos.

El trabajo hace énfasis en las herramientas clave de Google Cloud para la gestión de datos como: BigQuery, útil para consultas analíticas a gran escala; Dataplex, relevante en la gobernanza de datos unificada; Catalog, facilita la administración de metadatos; y Vertex AI, herramienta que permite el desarrollo de modelos de Machine Learning (ML). Además, se realiza un ejemplo práctico con Vertex AI utilizando el modelo Gemini para generar resúmenes automáticos de contenido audiovisual y así observar el potencial de este servicio en el procesamiento de datos no estructurados.

La metodología usada integra conceptos importantes de la gestión de datos en la nube con ejercicios aplicados sobre conjuntos de datos de comercio electrónico (TheLook). La utilidad del curso se evidencia a lo largo del mismo mediante ejemplos prácticos que muestran el potencial de estas tecnologías para resolver desafíos comunes en la ciencia de datos actual, especialmente en el manejo de datos heterogéneos y la escalabilidad de procesos.

Palabras clave— Google Cloud, Data Management, BigQuery, Dataplex, Vertex AI, Machine Learning, Análisis de Datos, Gobernanza de Datos, Gestión de Datos.

1. Introducción

En la actualidad, la gestión eficiente de información se ha convertido en la base del profesional en ciencia y analítica de datos (Badshah et al., 2024). Más allá del dominio de lenguajes y modelos predictivos, la capacidad para almacenar, organizar y acceder a datos de manera óptima delimita la calidad misma del análisis y, por lo tanto, el impacto real de los informes generados. Un científico de datos que comprende profundamente los principios de arquitecturas modernas como el data Lakehouse, los flujos de gobernanza y las estrategias de consulta distribuida, no solo optimiza recursos computacionales, sino que desarrolla una visión holística para resolver problemas complejos. Esta competencia permite navegar el ciclo completo de los datos, desde su obtención hasta su implementación en modelos productivos, transformando información cruda en decisiones estratégicas.

El curso Data Management and Storage in the Cloud ofrecido por Google está centrado en resumir los conceptos clave como Datalakehouse, BigQuery, Datafabric y Vertex AI (Google LLC, 2024a). El presente informe es una síntesis académica de este curso, considerando como los contenidos potencian habilidades críticas para el análisis de datos de los profesionales de la Especialización en Analítica y Ciencia de Datos de la Universidad de Antioquia, especialmente en el manejo moderno de datos en la nube.

2. Justificación académica

La evolución de la gestión y almacenamiento de datos ha respondido a las problemáticas tecnológicas de cada época. Los almacenes de datos predominaron en las décadas de 1980 y 1990, basados en estructuras rígidas orientadas al análisis empresarial, pero presentaban dificultades para gestionar datos no relacionales. Con el auge del Big Data en la década de 2010, surgió el Data Lake como alternativa flexible para almacenar cualquier tipo de información, aunque su falta de gobernanza y los grandes volúmenes de datos sin procesar lo hacían propenso a convertirse en "fango de datos", lo que resultaba inviable en entornos productivos. La solución a estos problemas llegó con el Data Lakehouse, un modelo híbrido que combina catálogos centralizados, soporte para aprendizaje automático y procesamiento en tiempo real, posicionándose como el estándar actual en gestión de datos (Harby & Zulkernine, 2022).

En el ámbito industrial, el Lakehouse mejora la gestión de datos, el procesamiento en tiempo real y el análisis predictivo, aprovechando servicios en la nube (Freitas et al., 2025). Google Cloud ofrece herramientas integradas para abordar estos desafíos, desde la gobernanza automatizada hasta el almacenamiento eficiente. En el sector académico, Analytics Hub permite trabajar con datos abiertos proporcionados por diversas entidades, facilitando la comprensión de servicios como BigQuery (optimización de almacenamiento) y Dataplex (gestión de metadatos) en contextos reales, lo que enriquece la formación en arquitecturas cloud. Los enfoques tradicionales Data Warehouse y Data Lake son complementarios pero disjuntos. Mientras los almacenes de datos ofrecen una gobernanza robusta, pero esquemas inflexibles, los repositorios de datos permiten almacenar datos estructurados, semiestructurados y no estructurados, aunque con un alto riesgo de degradación por la falta de organización (Harby & Zulkernine, 2022). El Data Lakehouse surge para aprovechar datos masivos y diversos en tiempo real, combinando la flexibilidad del Lake (soporte para machine learning y streaming) con la capacidad analítica del Warehouse (consultas SQL, transacciones ACID). Tecnologías como Delta Lake garantizan la calidad de datos no estructurados (Google LLC, 2024c), mientras que Google Cloud implementa este enfoque mediante Dataplex y Data Catalog, que generan metadatos activos (linaje, clasificación) para auditorías detalladas y evitar la degradación de los datos (Google LLC, 2024e).

El curso Data Management and Storage in the Cloud proporciona conocimientos necesarios para la industria actual, donde los servicios en la nube son esenciales. Google Cloud, una de las plataformas más utilizadas, facilita el acceso, la escalabilidad bajo demanda y la reducción de costos al eliminar la necesidad de infraestructura física. Sus herramientas resuelven problemas complejos: Dataplex unifica datos dispersos en una capa gobernada, Analytics Hub permite compartir conjuntos de datos sin duplicarlos, y BigQuery ejecuta análisis a gran

escala con un modelo de pago por uso. La integración con Vertex AI automatiza tareas como la clasificación de datos y la generación de metadatos, mejorando la eficiencia.

Al enseñar estos conceptos, el curso prepara a los estudiantes para implementar soluciones modernas que integran almacenamiento flexible, gobernanza dinámica (mediante Data Catalog) y procesamiento avanzado, habilidades clave en un mercado que demanda interoperabilidad y agilidad en el manejo de datos (Al Hadwer et al., 2019; Freitas et al., 2025). Este curso complementa la formación en ciencia de datos al profundizar en servicios cloud actuales. Su enfoque en herramientas como BigQuery, Dataplex y Vertex AI, junto con conceptos de arquitectura, almacenamiento y procesamiento, prepara a los estudiantes para enfrentar retos como la integración de fuentes diversas y la implementación de gobernanza en entornos complejos. En la industria, estas tecnologías son indispensables para optimizar procesos gracias a su escalabilidad sin inversión en hardware. Para la investigación, Analytics Hub facilita el acceso a datos públicos compartidos por entidades, fomentando proyectos colaborativos. El curso se distingue por equilibrar teoría y práctica, centrándose en la gestión de datos con arquitecturas tradicionales y modernas, además de desarrollar habilidades transversales como la documentación de linajes y la comunicación de resultados, competencias críticas en roles de ciencia de datos.

3. Descripción técnica del curso

El curso está estructurado en cuatro módulos secuenciales que integran aprendizaje teórico y práctico. Los temas se desarrollan mediante materiales textuales, videos explicativos, evaluaciones conceptuales con preguntas de selección múltiple y laboratorios guiados que permiten aplicar los conceptos en entornos reales utilizando tecnologías como Dataproc y BigQuery. El enfoque didáctico se complementa con casos de estudio basados en TheLook, una entidad ficticia de comercio electrónico, que simula desafíos empresariales típicos. Adicionalmente, se fomenta el aprendizaje autónomo a través de documentación técnica oficial de Google Cloud (Google LLC, 2024b).

Los datasets utilizados provienen de TheLook y presentan diversas estructuras (tablas relacionales, JSON, arrays), ejemplificando los tipos de campos disponibles en BigQuery, desde tipos simples (cadenas, números) hasta complejos (structs, arrays anidados). Estos datos sirven para prácticas de normalización, particionamiento, procesamiento por lotes y streaming, así como para consultas SQL (Google LLC, 2024d), mientras que en el módulo 4 se emplean notebooks de Jupyter en Dataproc para complementar los laboratorios.

El curso utiliza SQL estándar y dialectos específicos de BigQuery, junto con Apache Spark para procesamiento distribuido en Dataproc. Las herramientas clave incluyen Data Catalog para gestión de metadatos y Dataplex para orquestación. La infraestructura se basa

completamente en Google Cloud: BigQuery como almacén analítico principal, Dataproc para clusters efímeros, y Compute Engine para recursos bajo demanda. Dataplex unifica metadatos entre BigQuery, Cloud Storage y bases de datos externas, demostrando la gobernanza en entornos multicloud bajo el modelo de pago por uso, que destaca por su escalabilidad y reducción de costos frente a infraestructura física.

Los conceptos y herramientas esenciales explicados en el curso para la gestión de información en la nube de Google Cloud se mencionan a continuación y son esquematizados en la Figura 1.

FLUJO DE APRENDIZAJE EN GESTIÓN DE DATOS EN GOOGLE CLOUD

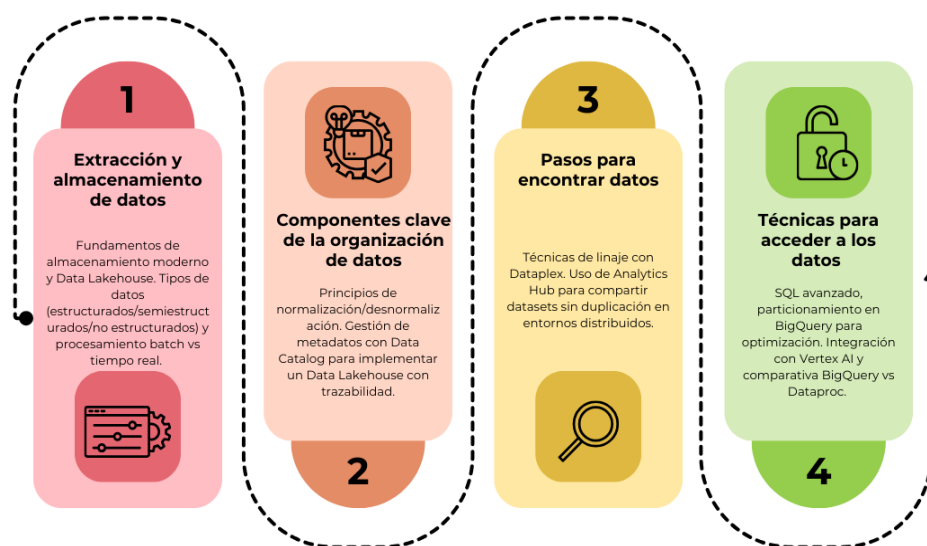


Figura 1. Ciclo de aprendizaje del curso. Elaboración propia.

- **Módulo 1: Extracción y almacenamiento de datos**

Define las bases conceptuales del almacenamiento moderno, profundizando en la arquitectura Data Lakehouse y los tipos de datos (estructurados, semiestructurados y no estructurados). Explica los campos disponibles en BigQuery y diferencia entre procesamiento por lotes y en tiempo real, habilidades fundamentales para diseñar sistemas escalables.

- **Módulo 2: Componentes clave de la organización de datos**

Enseña a estructurar información eficientemente mediante principios de normalización/desnormalización y su impacto en el rendimiento de consultas. Cubre los fundamentos de gobernanza con Data Catalog para administración de metadatos y trazabilidad, permitiendo implementar un Data Lakehouse funcional con mejores prácticas de documentación.

- **Módulo 3: Pasos para encontrar datos**

Desarrolla habilidades para localizar y utilizar datos distribuidos, enseñando técnicas de linaje con Dataplex para rastrear origen y transformaciones. Presenta Analytics Hub como solución para compartir conjuntos de datos entre equipos sin duplicación, preparando para escenarios reales con fuentes dispersas.

- **Módulo 4: Técnicas para acceder a los datos**

Consolida técnicas profesionales de consulta y optimización, incluyendo particionamiento de tablas en BigQuery para reducción de costos, SQL avanzado e integración con Vertex AI para crear modelos predictivos directamente sobre los datos almacenados. Compara BigQuery y Dataproc para seleccionar la herramienta adecuada según cada caso de uso.

4. Herramientas de Google Cloud para la ciencia de datos

La práctica cotidiana de la ciencia de datos implica la gestión de múltiples fuentes de información en diferentes formatos y estructuras. El dominio en la selección de herramientas apropiadas para el procesamiento de cada tipo de dato constituye una competencia esencial en este campo en constante evolución.

4.1. Herramientas de Google Cloud

A continuación, se presenta un resumen de algunas herramientas que Google Cloud pone a disposición de todo aquel interesado en el mundo de la analítica y la ciencia de datos.

- **Data Lakehouse**

Un Data Lakehouse es una arquitectura moderna de gestión de datos que busca combinar el almacenamiento flexible de datos no estructurados de un data lake y la capacidad de gestión de un data warehouse, permitiendo ejecutar procesos de BI y ML sobre todo tipo de datos

(Oracle Corporation, 2025). Se diferencia de servicios de almacenamiento como Amazon S3, en que no es sólo para el almacenamiento de objetos, sino que es una arquitectura de almacenamiento, gestión y análisis unificado. La flexibilidad de un data lakehouse beneficia al científico de datos moderno ya que le facilita el desarrollo de modelos avanzados de predicción, clasificación y recomendación a partir de datos estructurados, semiestructurados y no estructurados (Krasovytskyi, 2025).

- **BigQuery**

BigQuery es una solución totalmente administrada para la analítica empresarial. Es una plataforma pensada para el Big Data en la cual Google integra almacenamiento de datos en la nube, optimización de consultas SQL sobre grandes volúmenes de datos, herramientas de visualización como Looker y conexión con herramientas de inteligencia artificial (IA). Mediante esta herramienta los científicos de datos tienen a su alcance frameworks de BigData como Apache Hadoop y Apache Spark, Google Cloud Storage para el almacenamiento, Pub/Sub para la ingesta de datos en streaming, entre otras herramientas; todo ello unificado en una sola plataforma (Google LLC, 2024b; uCloud, 2023).

- **Dataplex**

Dataplex es un tejido de datos inteligente que ayuda a unificar datos distribuidos y automatizar el control y la administración de esos datos para potenciar las estadísticas a gran escala (Google LLC, 2025c). En otras palabras, es una herramienta potente que simplifica la administración de datos y permite a las empresas el manejo centralizado, monitoreo y control de sus datos en Data Lakes, Data Warehouses y Data Marts; facilitando su seguridad, gobernanza y unificación. Mediante Dataplex, las empresas pueden gestionar sus soluciones de almacenamiento en la nube para evitar una mala organización de los datos, inadecuada administración de metadatos, endeble gobernanza, entre otros problemas que podrían derivar en la pérdida del contenido informativo de la data almacenada, ya que a pesar de que se añaden constantemente nuevos datos, la información relevante es cada vez más difícil de encontrar (Lauer, 2022).

- **Analytics Hub**

Analytics Hube ayuda a superar la dificultad de acceso a datos de calidad, ya que se trata de un servicio de intercambio de datos que permite compartir recursos de forma eficiente y segura entre organizaciones para abordar los desafíos relacionados con la fiabilidad y el coste de los datos (Google LLC, 2025a). Esta biblioteca de datos internos y externos permite a los productores de datos o “publicadores” poner sus conjuntos de datos a disposición de los consumidores de datos o “suscriptores”. De esta manera, los publicadores pueden monetizar

sin poner en riesgo la integridad de sus datos mientras que los suscriptores pueden analizar una copia de datos a los que, probablemente, no tendrían acceso de otra manera.

- **Dataproc**

Google Cloud nos ofrece una opción para automatizar procesos de Big Data mediante Dataproc, un servicio administrado de Spark y Hadoop que permite el uso de herramientas de código abierto para tareas tales como el procesar, consultar y transmitir datos por lote, incluyendo también herramientas de machine learning (Google LLC, 2025b). Esta solución facilita la creación de clusters, la gestión de los mismos y permite la disminución del coste asociado a la infraestructura puesto que sólo se paga por los recursos que se consumen, además de la posibilidad de apagar los clusters cuando no estén en uso o retirar instancias de los mismos si es que están infrautilizadas. Gracias a esta flexibilidad al redimensionar los clusters, se pueden realizar operaciones de creación y escalado en un tiempo estimado de 90 segundos.

Adicionalmente, las últimas versiones de Spark, Hadoop, Pig y Hive están siempre disponibles, eliminando la necesidad del mantenimiento de las versiones de las herramientas open source (Oscar Fernández, 2025).

- **Vertex AI**

Para facilitar la construcción de modelos de machine learning, Google creó Vertex AI, una plataforma que combina la ingeniería de datos, la ciencia de datos y los flujos de trabajo de ingeniería de ML. Vertex AI permite a los equipos de ML colaborar con un conjunto de herramientas común para entrenar modelos y aplicaciones de inteligencia artificial, personalizar grandes modelos de lenguaje (LLMs) y desplegarlos a gran escala. No es necesaria mucha experiencia para desarrollar e implementar modelos personalizados de ML en Vertex AI, pues la funcionalidad AutoML permite entrenar modelos a partir de todo tipo de datos sin escribir código ni preparar divisiones de datos. Además, en su “Galería de instrucciones”, Vertex AI ofrece un abanico de modelos pre entrenados para una gran variedad de tareas específicas.

Por otra parte, si ya se tiene experiencia, también es posible implementar modelos con código propio al cargar notebooks de Jupyter en Vertex AI Workbench e inclusive colaborar con otros desarrolladores utilizando Colab Enterprise, una versión de Colaboratory que está integrada en la plataforma. (Google LLC, 2025d).

4.2. Procesamiento de datos no estructurados: Vertex AI

Durante nuestro tiempo de formación académica hemos adquirido un espectro de herramientas, modelos, plataformas y metodologías rigurosas para abordar desafíos de analítica de datos. En su mayoría, la aplicación de dichos conocimientos se validó mediante conjuntos de datos estructurados provenientes de organizaciones públicas (e.g., UNdata, Datos.gov.co, Medata), así como de organizaciones privadas (e.g., Kaggle, GitHub, Google). No obstante, un científico de datos integral también debe estar en capacidad de explorar el vasto potencial de las fuentes de datos semiestructurados y no estructurados que tenemos disponibles en la actualidad. En este contexto, presentamos un ejemplo de cómo podemos utilizar Vertex AI para la generación de resúmenes de contenido audiovisual aprovechando las capacidades de Gemini, un modelo de inteligencia artificial creado por Google, aplicado sobre un video de YouTube; demostrando así la utilidad de la herramienta para el procesamiento de datos no estructurados.

Los pasos a seguir son los siguientes:

- 1) Crear una cuenta de Google Cloud.
- 2) Tras la creación y configuración de una cuenta en Google Cloud Platform (GCP), el usuario deberá dirigirse a la interfaz de la "Consola", ubicada generalmente en la esquina superior derecha de la pantalla (ver Figura 2).

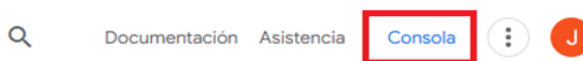


Figura 2. Botón “Consola” de la interfaz de Google Cloud Platform. Imagen tomada de Google Cloud Platform

- 3) Una vez dentro de la consola de GCP, se procederá a utilizar la barra de búsqueda (ver Figura 3) para localizar y seleccionar el servicio "Vertex AI".

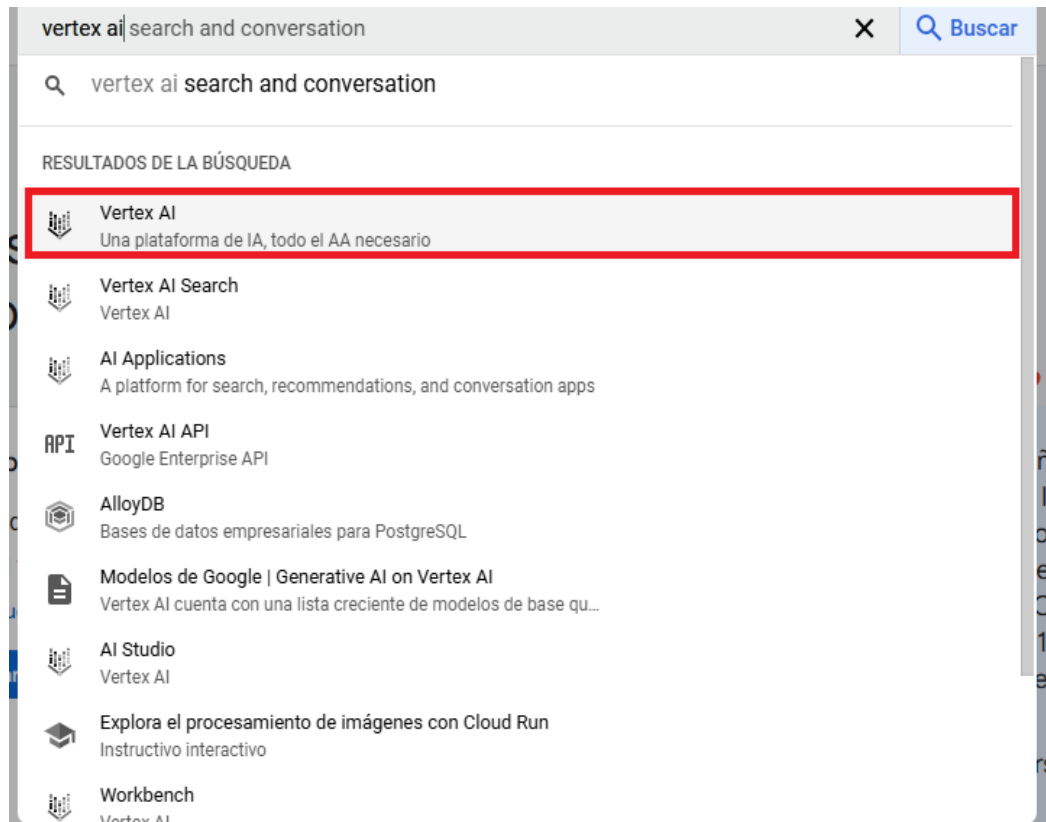


Figura 3. Barra de búsqueda de la interfaz de Google Cloud Platform. Imagen tomada de Google Cloud Platform

Esta acción permitirá el acceso al panel principal de la plataforma de inteligencia artificial de Google (ver Figura 4).

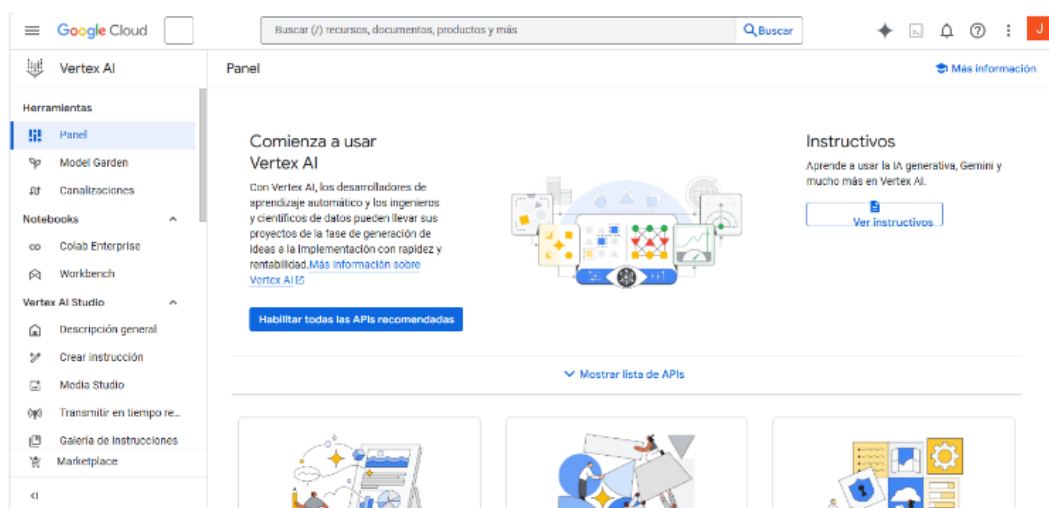


Figura 4. Panel de inicio de Vertex AI. Imagen tomada de Vertex AI

- 4) Dentro del panel de Vertex AI, se deberá elegir la opción "Vertex AI Studio" (ver Figura 5). Esta herramienta proporciona una interfaz gráfica para la experimentación y desarrollo de modelos de aprendizaje automático.

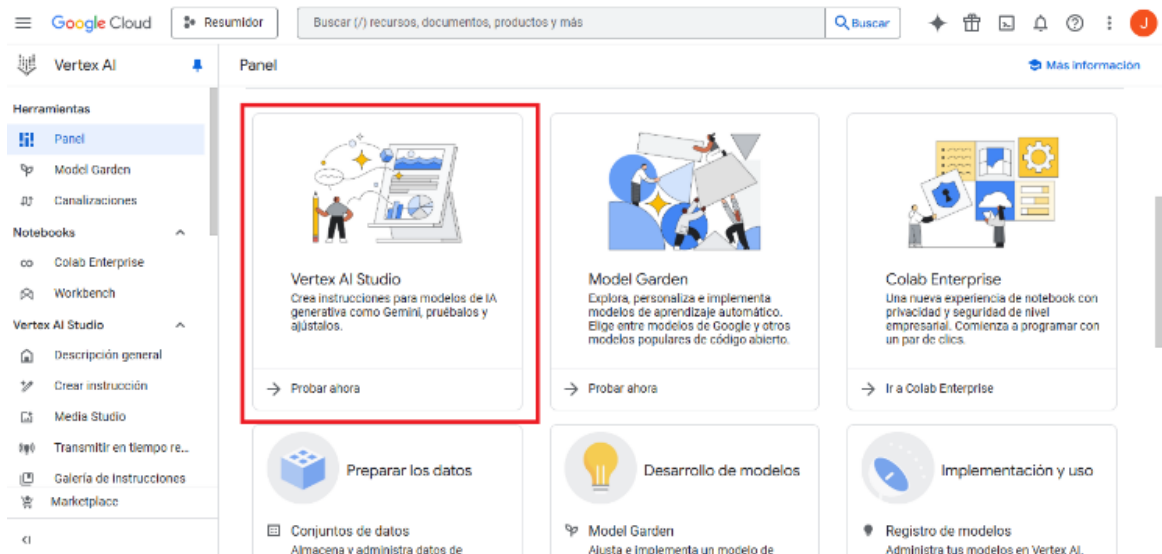


Figura 5. Opción “Vertex AI Studio” de la plataforma Vertex AI. Imagen tomada de Vertex AI

- 5) Vertex AI, al estar integrado con el modelo Gemini, ofrece la capacidad de proporcionar instrucciones directas a través de la consola para la creación de aplicaciones específicas (ver Figura 6).

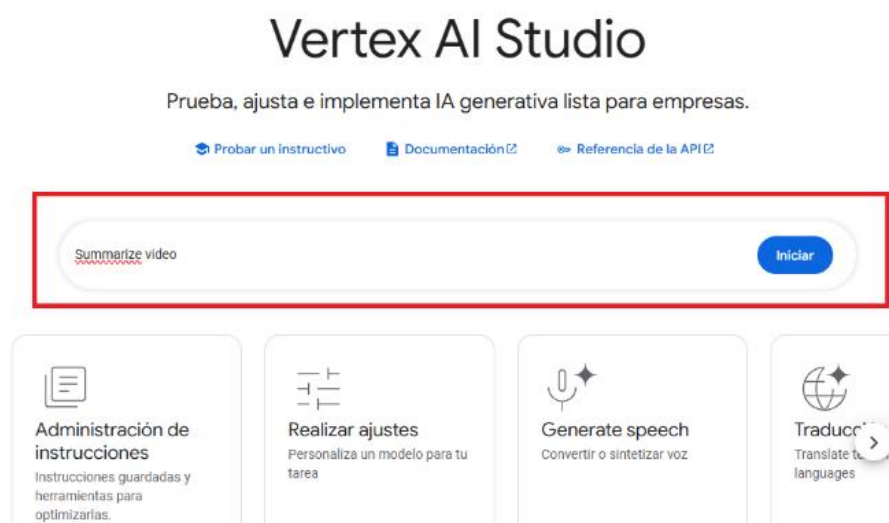


Figura 6. Barra de instrucciones de Vertex AI Studio. Imagen tomada de Vertex AI

Adicionalmente, la plataforma dispone de una "Galería de Instrucciones" que presenta un conjunto diverso de tareas preconfiguradas y listas para su implementación. Para explorar la totalidad de las funcionalidades disponibles, se recomienda seleccionar la opción "Ver todos" (ver Figura 7).

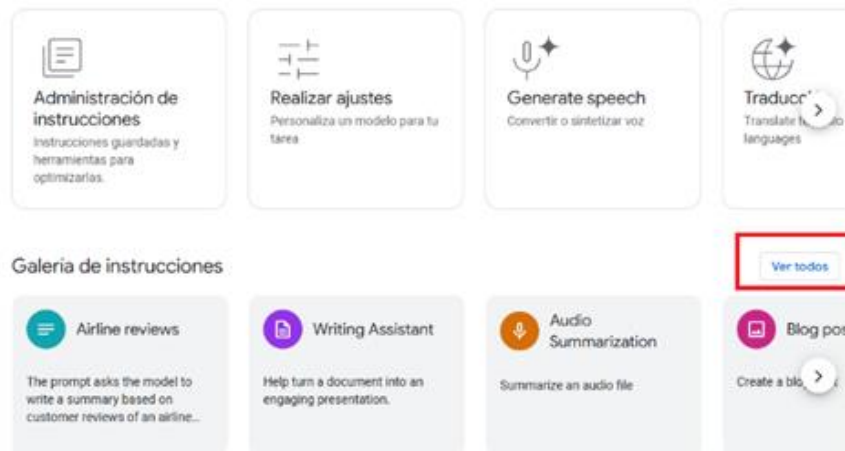


Figura 7. Botón “Ver todos” de la galería de instrucciones de Vertex AI Studio. Imagen tomada de Vertex AI

- 6) Dentro de la galería de instrucciones, se aplicará un filtro mediante la opción "Video". Posteriormente, se identificará y seleccionará la tarea específica denominada "Summarize video" (ver Figura 8).

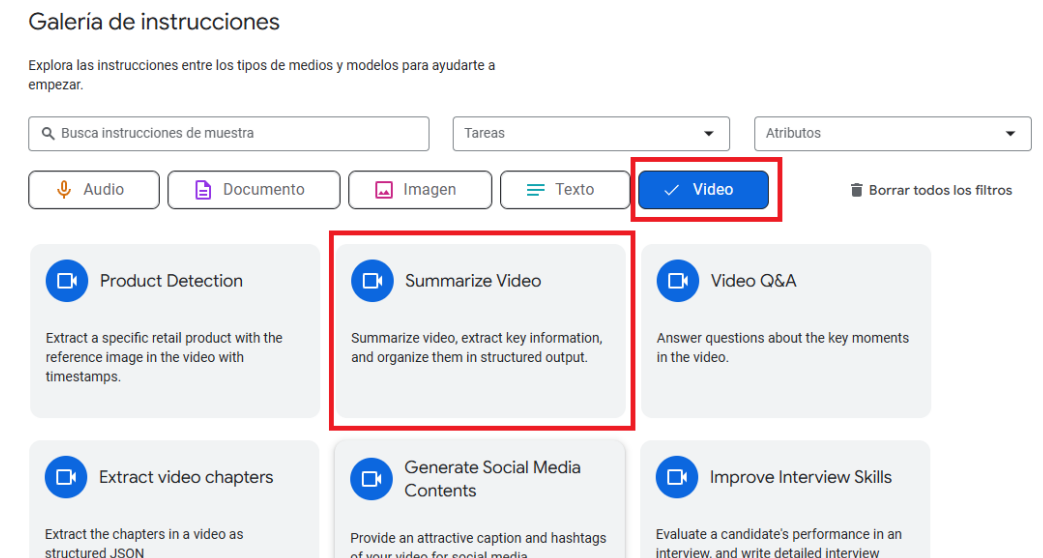


Figura 8. Opciones de filtrado y procesamiento según tipo de dato. Imagen tomada de Vertex AI

- 7) El siguiente estadio implica la elaboración de instrucciones precisas y contextualizadas dirigidas al modelo, con el objetivo de obtener resultados personalizados. Paralelamente, se deberá importar el archivo de video a resumir mediante la función "Agregar archivos multimedia" (ver Figuras 9 y 10). En el presente caso ilustrativo, el video de referencia es "Los superpoderes de los datos", una presentación TedTalk de Manuel Enciso alojada en la plataforma YouTube.

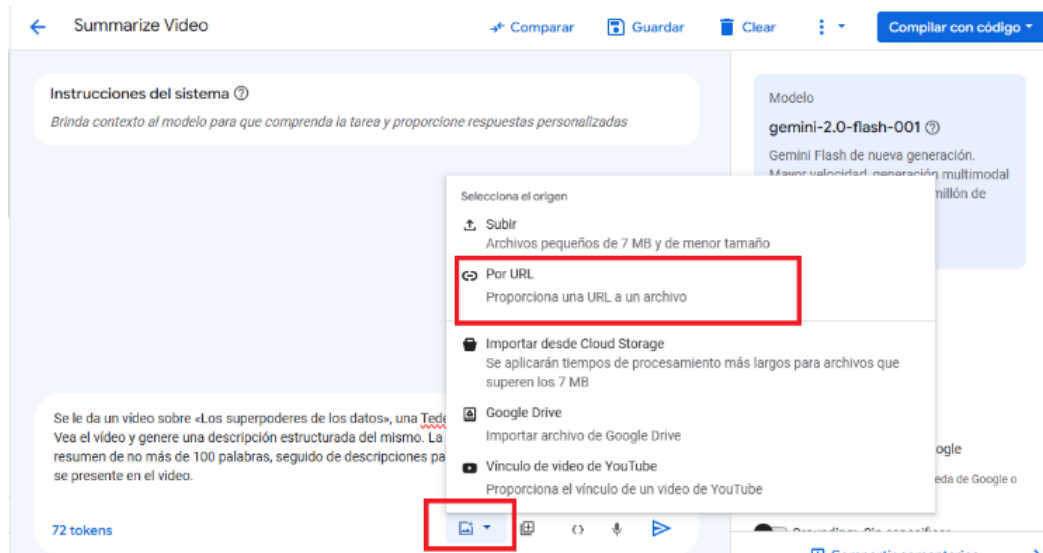


Figura 9. Opciones para la importación de archivos multimedia. Imagen tomada de Vertex AI

Insertar por URL

Agrega un archivo de una URL a tu instrucción. Tipos de archivo admitidos: imagen, video, audio, PDF y TXT. Los archivos están sujetos al [límite de tokens de entrada del modelo](#)

URL del archivo

https://www.youtube.com/watch?v=VHw7rEtawsM

Tipo de archivo

Video

Insertar

Cancelar

Figura 10. Carga de video a la plataforma mediante URL. Imagen tomada de Vertex AI

- 8) Una vez definidos los parámetros y cargado el material audiovisual, se procederá a la ejecución del modelo de síntesis (ver Figura 11). Se deberá aguardar la generación de los resultados por parte del sistema.

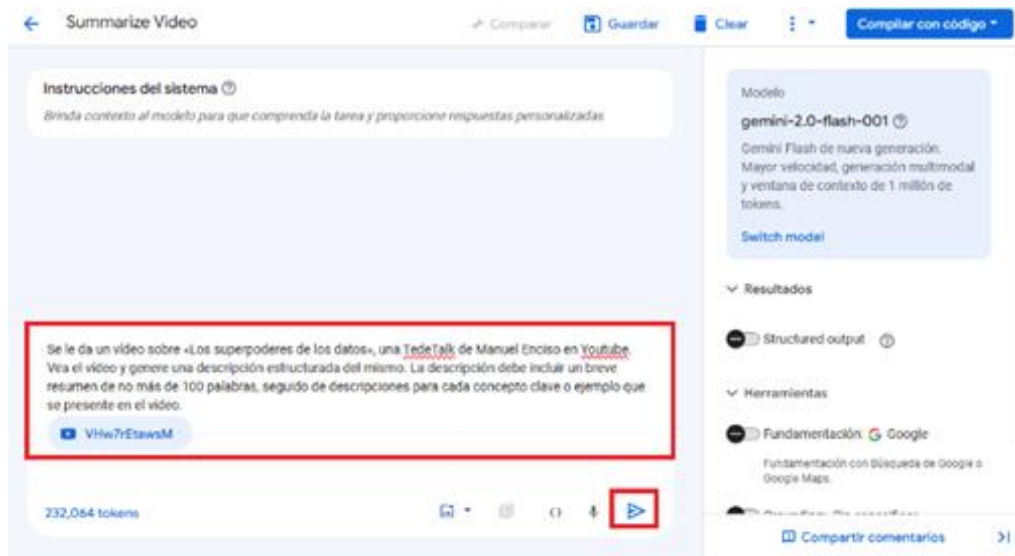


Figura 11. Área de prompt para instrucciones y ejecución del modelo. Imagen tomada de Vertex AI

- 9) Finalmente, se llevará a cabo un análisis exhaustivo de la respuesta proporcionada por el modelo. Este proceso tiene como finalidad validar la pertinencia y calidad del resumen generado, determinando la necesidad de realizar ajustes en las instrucciones (prompt), considerar la utilización de un modelo de IA generativa diferente o efectuar modificaciones adicionales.

En nuestro caso, el resultado fue el siguiente:

Resumen:

En esta charla TEDx Málaga, Manuel Enciso, un "adicto a los datos", explora el poder de los datos para ver la realidad tal como es y predecir el futuro. A través del método científico y ejemplos del mundo real, Enciso demuestra cómo los datos pueden ser utilizados para el bien o para el mal, y cómo la visualización de datos puede revelar patrones ocultos y desafiar nuestras percepciones. El orador concluye enfatizando la importancia de la responsabilidad al utilizar los datos y la necesidad de protegerse contra la manipulación.

Conceptos clave y ejemplos:

- * **Adicción a los datos:** Enciso se describe a sí mismo como un "adicto a los datos", lo que significa que está fascinado por el poder de los datos para revelar información y patrones ocultos.
- * **Superpoderes de los datos:** El orador afirma que los datos le dan "superpoderes" para ver la realidad tal como es y predecir el futuro.
- * **Método científico:** Enciso utiliza el método científico para demostrar sus superpoderes, describiendo sus observaciones, formulando hipótesis y realizando experimentos.
- * **Realidad diáfana:** El orador afirma que los datos le permiten ver la realidad de forma "diáfana", es decir, clara y transparente.
- * **Interpretación y estimación:** Enciso describe sus superpoderes como la capacidad de interpretar los datos y estimar o pronosticar el futuro.
- * **Experimento en vivo:** El orador realiza un experimento en vivo con datos de una sesión TEDx anterior para demostrar sus superpoderes.
- * **Igualdad de género:** Enciso utiliza el filtro de la igualdad de género para analizar los datos de la sesión TEDx anterior, revelando que los hombres son más coquetos que las mujeres.
- * **Distribución normal:** El orador explica que la distribución de las edades de los asistentes a la sesión TEDx anterior se ajusta a una distribución normal.
- * **Derecho a la coquetería:** Enciso menciona que algunas personas no declaran su edad en las fichas de inscripción, lo que él llama "ejercer su derecho a la coquetería".
- * **Diagrama de tartas:** El orador utiliza un diagrama de tartas para mostrar la proporción de hombres y mujeres que no declaran su edad en las fichas de inscripción.
- * **Visualización de datos:** Enciso explica que la visualización de datos puede revelar patrones ocultos y desafiar nuestras percepciones.
- * **Correlación:** El orador introduce el concepto de correlación, que es la relación entre dos o más variables.
- * **Calentamiento global:** Enciso utiliza el ejemplo del calentamiento global para demostrar cómo las correlaciones pueden ser utilizadas para manipular la información.
- * **Responsabilidad:** El orador concluye enfatizando la importancia de la responsabilidad al utilizar los datos y la necesidad de protegerse contra la manipulación.

5. Conclusiones

- El curso de gestión de datos en la nube de Google Cloud es un complemento atractivo en la formación de especialistas en analítica y ciencia de datos, ya que brinda conocimientos técnicos importantes y desarrolla habilidades de adaptación para un entorno cambiante. La gestión de datos ha demostrado ser un campo dinámico que se transforma continuamente para responder a las necesidades actuales (Memarian & Doleck, 2024); la información adquiere relevancia por su potencial transformador, los desafíos para encontrar valor a la información son más exigentes en una época de generación masiva de datos diversos, a gran velocidad y volumen.
- La actualidad exige que el profesional en datos supere el rol de simple usuario de herramientas tecnológicas. Debe desarrollar la capacidad de apropiarse del conocimiento disponible para generar innovación, sea mediante la adaptación creativa de soluciones existentes o el desarrollo de nuevas propuestas. Las plataformas de Google Cloud, con su flexibilidad y versatilidad, son estratégicas para implementar arquitecturas de gestión adaptadas a cada necesidad específica.
- Estas tecnologías no solo optimizan los procesos, sino que permiten al científico y analista de datos potenciar su capacidad creativa para extraer valor de la información de manera óptima. La correlación entre conocimiento técnico y pensamiento innovador ayuda a liderar verdaderas transformaciones en la gestión de datos mediante el manejo estratégico de la información, marcando así la diferencia en un campo profesional cada vez más competitivo y demandante.

Referencias

- Al Hadwer, A., Gillis, D., & Rezania, D. (2019). *Big Data Analytics for Higher Education in The Cloud Era*.
- Badshah, A., Daud, A., Alharbey, R., Banjar, A., Bukhari, A., & Alshemaimri, B. (2024). Big data applications: overview, challenges and future. *Artificial Intelligence Review*, 57(11). <https://doi.org/10.1007/s10462-024-10938-5>
- Freitas, N., Rocha, A. D., & Barata, J. (2025). Data management in industry: concepts, systematic review and future directions. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/s10845-025-02570-z>
- Google LLC. (2024a). *Data Management and Storage in the Cloud - Español*. Google Cloud Skills Boost. <https://www.cloudskillsboost.google>
- Google LLC. (2024b). *Descripción general de BigQuery*. <https://cloud.google.com/bigquery/docs/introduction?hl=es-419>
- Google LLC. (2024c). *Descripción general de la arquitectura de data lakehouse*. https://www.cloudskillsboost.google/paths/420/course_templates/962/documents/472002
- Google LLC. (2024d). *Explora la situación del curso 2: Comercio electrónico TheLook*. https://www.cloudskillsboost.google/paths/420/course_templates/962/documents/471994
- Google LLC. (2024e). *Función de linaje de datos de Dataplex*. https://www.cloudskillsboost.google/paths/420/course_templates/962/documents/472046
- Google LLC. (2025a). *Compartir datos en BigQuery*. <https://cloud.google.com/analytics-hub?hl=es>
- Google LLC. (2025b, abril 14). *Descripción general de Dataproc*. <https://cloud.google.com/dataproc/docs/concepts/overview?hl=es-419>
- Google LLC. (2025c, abril 17). *Documentación de Dataplex*. <https://cloud.google.com/dataplex/docs?hl=es-419>
- Google LLC. (2025d, abril 17). *Introducción a Vertex AI*. <https://cloud.google.com/vertex-ai/docs/start/introduction-unified-platform?hl=es-419>

- Harby, A. A., & Zulkernine, F. (2022). From Data Warehouse to Lakehouse: A Comparative Review. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 389-395. <https://doi.org/10.1109/BigData55660.2022.10020719>
- Krasovytskyi, L. (2025, enero 13). *Data Lakehouse: Real-Life Use Cases and Examples*. <https://www.owox.com/blog/articles/data-lakehouse-use-cases>
- Lauer, C. (2022, marzo 7). *What is Google Dataplex? — The Data Fabric*. <https://medium.com/codex/what-is-google-dataplex-the-data-fabric-a6faf37e746f>
- Memarian, B., & Doleck, T. (2024). Data science pedagogical tools and practices: A systematic literature review. *Education and Information Technologies*, 29(7), 8179-8201. <https://doi.org/10.1007/s10639-023-12102-y>
- Oracle Corporation. (2025). *What Is a Data Lakehouse?* <https://www.oracle.com/big-data/what-is-data-lakehouse/>
- Oscar Fernández. (2025, enero 14). *Datapro: Spark y Hadoop en Google Cloud*. <https://aprenderbigdata.com/datapro/>
- uCloud. (2023, diciembre 13). *Google BigQuery: Qué es y para qué sirve*. <https://ucloudglobal.com/blog/google-bigquery-que-es-y-para-que-sirve/>