



**Google Cloud en la Teoría y la Práctica: Análisis Comparativo desde la Perspectiva de la
Práctica Profesional**

Jonathan Carvajal Galeano

Juan Mauricio Parrado Villa

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

Asesora

Maria Bernarda Salazar Sánchez, Doctora (PhD) en Ingeniería Electrónica

Universidad de Antioquia

Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2025

Cita	(Carvajal Galeano & Parrado Villa, 2025)
Referencia	Parrado Villa, J. M., & Carvajal Galeano, J. (2025). Google Cloud en la Teoría y la Práctica: Análisis Comparativo desde la Perspectiva de la Práctica Profesional
Estilo APA 7 (2020)	[Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia.



Grupo de Investigación Intelligent Information Systems Lab In2Lab

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.

Dedicatoria

Dedicado a todos ellos que dedican su vida a la enseñanza en ciencia de datos y analítica en nuestro país.

Agradecimientos

Gracias a la Universidad de Antioquia y a su comunidad de profesores del postgrado de especialización en ciencia de datos y analítica, por hacer posible la realización de este postgrado, y enseñarnos las bases para encaminar nuestro conocimiento al uso avanzado de la información.

Gracias también a nuestras familias por el apoyo constante durante esta etapa de nuestras carreras.

Tabla de contenido

Resumen	8
1. Introducción	9
2. Justificación académica	10
3. Descripción técnica	11
3.1. Mercado Actual y Relevancia del Curso	12
3.2. Aplicabilidad Práctica del Contenido	14
3.3. Limitaciones para novatos y profesionales	15
3.4. Propuestas para cerrar brechas	15
4. Conclusiones	16
Referencias	17

Lista de Figuras

Figura 1. Relación entre los conceptos de Autoscaling y Rightsizing.	11
Figura 2. Relación entre los conceptos de <i>rehosting</i> , <i>replatforming</i> y <i>refactoring</i> .	12
Figura 3. Etapas de un proyecto de datos. Tomado de University of Wisconsin-Madison.	13

Lista de Tablas

Tabla 1. Comparativa de contenido y su uso en la industria con recursos de GCP.	14
---	----

Siglas, acrónimos y abreviaturas

APA	American Psychological Association
GCP	Google Cloud Platform
ERIC	Education Resources Information Center
Esp.	Especialización
Párr.	Párrafo
PhD	Philosophiae Doctor
UdeA	Universidad de Antioquia
SMSS	SQL Server Management Studio
ETL	Extract Transform Load
SQL	Structured Query Language
ML	Machine Learning
AWS	Amazon Web Services
Azure	Microsoft Azure

Resumen

En el ámbito del análisis de datos de la actualidad, empresas que venden servicios en la nube, han adquirido relevancia debido a su escalabilidad y facilidad de ejecución en la industria de los datos. Específicamente, GCP ha adquirido gran relevancia debido a su integración de herramientas que permiten mejorar los procesos tradicionales. Este trabajo analiza el curso “Introduction to data analytics in Google Cloud” para comparar los contenidos del curso con las prácticas de la industria, evaluando su utilidad para profesionales y estudiantes, analizando sus diferencias con escenarios reales, e identificando limitaciones y oportunidades de mejora.

Palabras clave — ciencia de datos, analítica, Google Cloud Platform, learning path.

1. Introducción

El campo del análisis de datos experimenta una evolución constante, impulsado por la innovación tecnológica y la creciente demanda de información valiosa. En este contexto, plataformas de aprendizaje como Google Cloud Skills Boost se posicionan como pilares fundamentales para la adquisición de habilidades (Kyriazatis, 2024). Sin embargo, surge una pregunta crucial: ¿cómo se traduce el conocimiento teórico impartido en estos cursos a la realidad del entorno laboral? Este escrito explora respuestas a este interrogante, tomando como punto de inicio la formación disponible en Introduction to Data Analytics in GCP (Del inglés, Google Cloud Platform) (DataScientest, 2020). A través de una lente crítica y constructiva, se analiza el contenido del curso, contrastando con la experiencia práctica obtenida en proyectos reales que involucran el uso de GCP y BigQuery.

Este trabajo busca tender un puente entre la teoría y la práctica, evaluando la relevancia y aplicabilidad de los conceptos aprendidos, exponiendo las fortalezas y limitaciones en el contexto de desafíos laborales concretos tanto de las herramientas disponibles en el curso como de los servicios de GCP. Además, se presta especial atención a los nuevos contenidos, considerando las últimas actualizaciones y tendencias en el ecosistema de Google Cloud.

En consecuencia, este estudio no solo evalúa la eficacia del curso como medio de capacitación, sino también ofrece una reflexión que pueda servir como guía para profesionales y estudiantes interesados en fortalecer sus habilidades en análisis de datos en la nube. Compartir esta experiencia y análisis permitirá una mejor comprensión sobre la transición del aprendizaje formal al desempeño técnico en escenarios reales del mercado laboral.

2. Justificación académica

En el contexto actual de transformación digital, las plataformas de computación en la nube han adquirido un papel protagónico en la forma en que se accede, procesa y analiza la información. Dentro de este ecosistema, Google Cloud Platform (GCP) se ha consolidado como una opción destacada gracias a su integración nativa con herramientas analíticas como BigQuery, Dataflow y Looker (Praveen, 2024), las cuales permiten desarrollar soluciones escalables y eficientes para el análisis de datos. Específicamente en el curso *Introduction to Data Analytics in Google Cloud*, ofrecido a través de Google Cloud Skills Boost, constituye una puerta de entrada al dominio de estas herramientas desde una perspectiva teórico-práctica.

Este trabajo se enfoca su estudio en GCP debido a su relevancia creciente en el mercado (Tecnozero, n.d.), su facilidad de integración y su enfoque centrado en la conservación y análisis de datos. Esta decisión se fundamenta en criterios técnicos y prácticos derivados del uso continuo de estas plataformas en distintos entornos organizacionales (Bello et al., 2021).

La motivación central de este trabajo radica en analizar críticamente los contenidos del curso mencionado contrastándolos con experiencias laborales reales, sin particularizar en los sectores o tamaños de las organizaciones, sino destacando la aplicabilidad del conocimiento en diversos contextos profesionales. Esta comparación permitirá identificar el grado de alineación entre la formación académica y las competencias exigidas en la práctica, aportando así elementos valiosos para la toma de decisiones formativas.

Asimismo, este estudio resulta en una guía útil para estudiantes y profesionales que aún no se han decidido por una nube específica, ofreciendo una visión estructurada y argumentada sobre la utilidad de GCP para el análisis de datos. Al documentar nuestra experiencia y apreciaciones personales, buscamos facilitar el camino a quienes se inician en este campo, aportando a una toma de decisiones más informada y orientada a las necesidades reales del entorno laboral.

3. Descripción técnica

El curso *Introduction to Data Analytics in Google Cloud*, ofrecido a través de la plataforma Google Cloud Skills Boost, introduce de manera estructurada los fundamentos técnicos y las prácticas operativas más relevantes para el análisis de datos moderno en la nube. Con un enfoque centrado en Google Cloud Platform (GCP), el contenido abarca tanto herramientas específicas como principios fundamentales de arquitectura, aprovisionamiento y optimización de cargas de trabajo.

Entre los conceptos más destacados se encuentran los modelos de despliegue en la nube, el rol del analista de datos en proyectos de inteligencia de negocios, aspectos generales de arquitectura en la nube, aspectos básicos del lenguaje SQL, el ciclo de vida de los datos, roles en equipos de datos, conceptos generales de automatización en proyectos de datos, y conceptos importantes para la optimización de consumo y costo de recursos de la nube como *provisioning* (aprovisionamiento), el *autoscaling* (autoescalado), y el *rightsizing* (tamaño adecuado), teniendo presente en cada momento la administración eficiente de recursos en la nube (Google Cloud Blog, 2022). La Figura 1, refleja parte de los conceptos que se aprenden a través del desarrollo del curso, en los diferentes módulos del mismo:

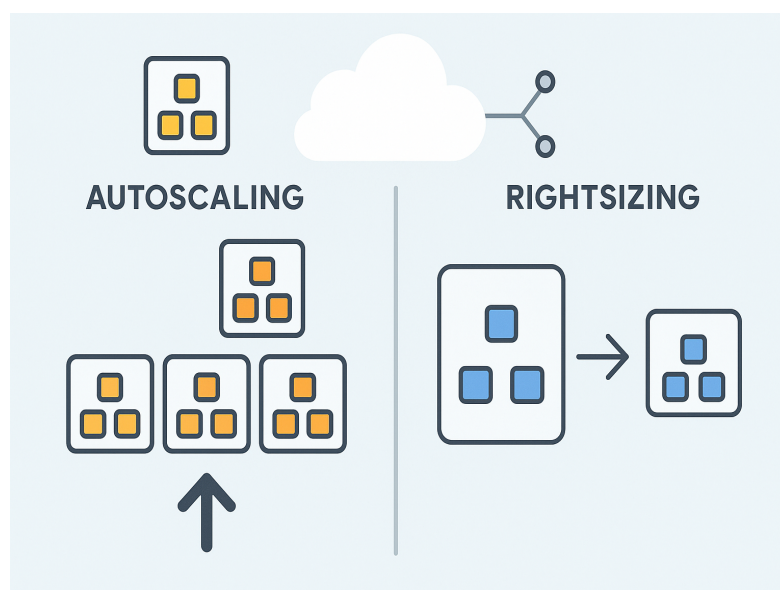


Figura 1. Relación entre los conceptos de Autoscaling y Rightsizing.

Estos conceptos permiten el entendimiento de una visión general del propósito de los proyectos de inteligencia de negocios, sus diferentes etapas, así como los roles implicados en cada una de ellas, y su vez permite al estudiante hacerse una idea de infraestructura que permite ajustarse automáticamente a las necesidades de procesamiento y almacenamiento sin intervención manual constante. Por ejemplo, el *autoscaling* en GCP puede configurarse fácilmente en servicios como App Engine, Cloud Run o Kubernetes Engine para reaccionar ante cambios en la demanda, lo cual no solo mejora el rendimiento, sino que también reduce costos operativos (Chaabnia, S., 2023).

Adicionalmente, el curso introduce estrategias clave de modernización como el *rehosting*, *replatforming* y *refactoring* (Figura 2). Estas estrategias permiten que las organizaciones, sin importar su tamaño o nivel de madurez tecnológica, puedan migrar y optimizar sus aplicaciones en la nube con base en sus capacidades y objetivos. GCP, con herramientas como *Migrate for Compute Engine* o *Cloud SQL*, habilita estos procesos de forma fluida y escalable. Muchos de estos términos no son necesariamente evidentes o intuitivos para quienes inician su carrera en la nube, viéndolo desde un punto de vista pedagógico son claves en la construcción profesional como teoría, y en la práctica profesional son claves en el entendimiento del consumo de recursos para su aplicación en codificaciones, visualizaciones o tratamiento en general.

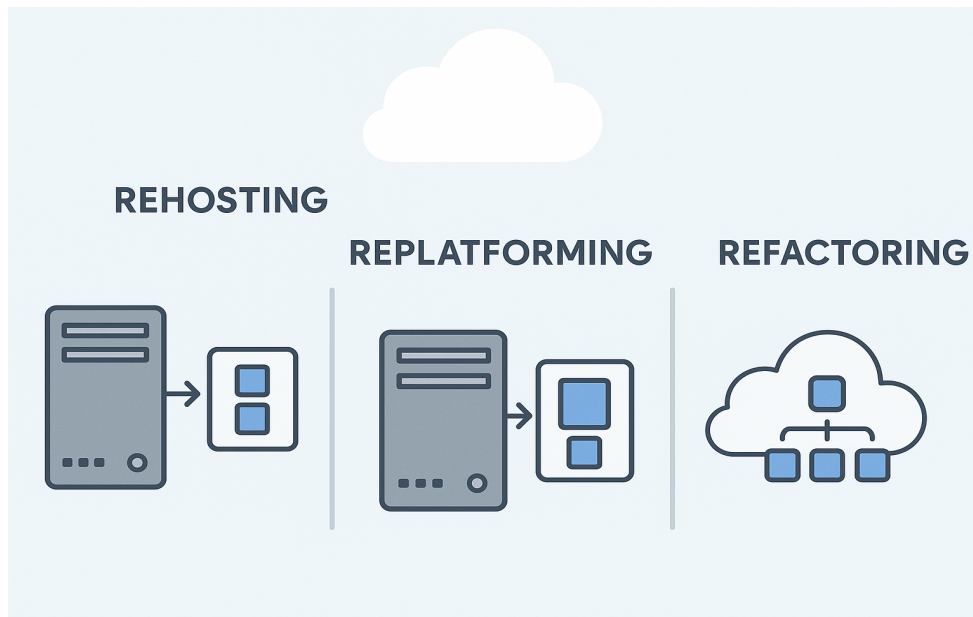


Figura 2. Relación entre los conceptos de *rehosting*, *replatforming* y *refactoring*.

3.1. Mercado Actual y Relevancia del Curso

Teniendo en cuenta el contenido del curso y cómo aborda conceptos de ingeniería de datos, desde nuestro punto de vista, brinda claridad terminológica y conceptual sobre procesos que, aunque pudiesen ser ejecutados de manera empírica, requieren de una base teórica sólida para su comprensión integral y su correcta aplicación. La forma en la que Google estructura sus materiales facilita que tanto novatos como profesionales en transición técnica puedan comprender no sólo el “cómo”, sino también el “por qué” detrás de cada práctica recomendada.

Luego, también se enfatiza el valor de servicios como BigQuery, el motor de consultas analíticas *serverless* de Google, que supera ampliamente los *frameworks* SQL tradicionales en términos de rendimiento, escalabilidad y facilidad de uso (Airbyte, 2024). BigQuery elimina la necesidad de administrar infraestructura, permitiendo ejecutar consultas complejas sobre volúmenes masivos de datos en segundos, y apoyándose en técnicas como particionamiento y *clustering* para optimizar el rendimiento y reducir costos (Google Cloud Blog, 2022).

En suma, la formación impartida no sólo aporta herramientas técnicas de alto nivel, sino también un marco conceptual que resulta vital para una práctica profesional más informada,

estratégica y eficiente. GCP, en conjunto con esta propuesta educativa, habilita una transición más fluida desde el aprendizaje empírico hacia una ingeniería de datos fundamentada y alineada con estándares globales. Sin embargo, en un contexto laboral real muchas empresas no emplean al 100% un plan fiel al mostrado pedagógicamente en el curso. El enfoque conceptual es especialmente importante a nivel de aprendizaje, sin embargo, por ejemplo, los conceptos de migración de datos no son necesariamente usados en cualquier compañía de manera estándar y controlada como se enseña el curso (Ruiz Bermejo et al., 2019), puede llegar a ser más bien una visión poco realista de lo que el analista se enfrenta en un proceso de migración.

Uno de los primeros puntos prácticos del curso, corresponde al uso de BigQuery, especialmente a través del primer laboratorio. Este ejercicio guía al estudiante en la ejecución de consultas básicas dentro de la consola de BigQuery, estableciendo una introducción a la interfaz, estructura de datos y uso de SQL en el contexto de la nube. Sin embargo, desde una perspectiva profesional, este laboratorio resulta limitado para perfiles que ya cuentan con experiencia previa en entornos de bases de datos relacionales y herramientas gráficas como *SQL Server Management Studio (SSMS)* o *MySQL Workbench*.

Adicional a esto, durante las etapas intermedias del curso, se hace un análisis de las diferentes etapas del ciclo de vida de los datos en la nube, que van desde la planeación, implementación y uso, hasta las etapas para compartir, archivar y destruir datos. Estas etapas se pueden observar en la *Figura 3*.

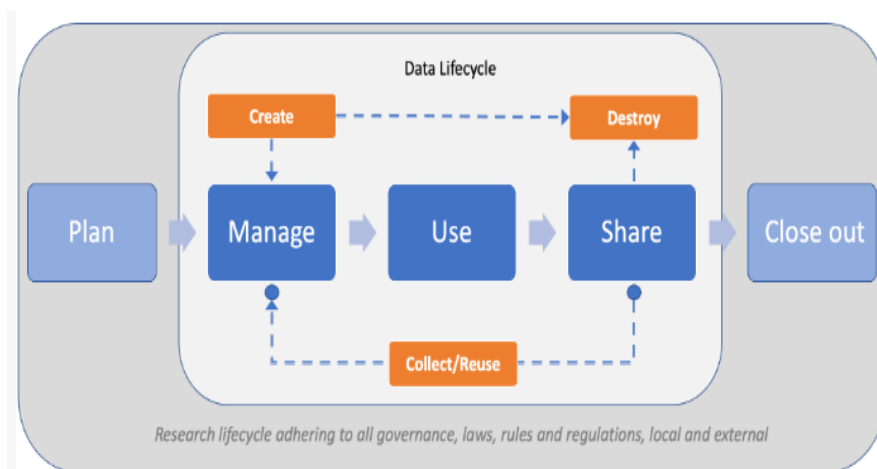


Figura 3. Etapas de un proyecto de datos. Tomado de University of Wisconsin-Madison.

Este modelo de desarrollo de proyecto de datos por etapas es sin dudas, uno de los temas más importantes del curso, y debido a que no guarda relación directa con los productos de GCP, esta estrategia se puede aplicar a cualquier otro ambiente de desarrollo de procesos de analítica y de datos en general. Este módulo incluye información clave en cualquier proyecto de datos como lo son los diferentes roles dentro de este tipo de proyectos, políticas de gobierno y acceso a los datos, políticas de eliminación y archivado de datos, políticas gubernamentales y no gubernamentales de privacidad de los datos (University of Wisconsin-Madison, s.f), entre otros conceptos no técnicos que, en nuestra experiencia, son de importancia vital en el desarrollo exitoso de estos proyectos.

3.2. Aplicabilidad Práctica del Contenido

Para abordar la aplicabilidad del contenido del curso en la industria, vale la pena preguntarnos: ¿Puede un estudiante implementar un *pipeline* de datos real solo con lo aprendido en el curso?. Con la experiencia obtenida en los laboratorios del curso, el estudiante puede hacer uso de las funciones básicas de *Big Query*, que le permiten una ingesta sencilla de datos con conexiones nativas de la herramienta, así como hacer queries dentro del explorador SQL, permitiendo una exploración inicial de un *pipeline* de datos a un conjunto de datos limitado por las conexiones nativas.

Según Chaabnia (2023), un ingeniero de datos debe dominar una serie de habilidades que le permiten no sólo la construcción de pipelines de datos, sino también el manejo adecuado de acceso a la información, *metadata* y protección de datos, elementos que se mencionan constantemente en el curso, incluyendo políticas de protección de datos personales, como las necesarias en pacientes de la industria de la salud; esto permite un acercamiento bastante interesante en términos de gobierno y seguridad de datos en casos reales, sin embargo el curso aún carece de ejemplos concretos de implementación de este tipo de políticas y las herramientas de Google Cloud que permiten este tipo de administración. Para ilustrar de manera más adecuada las diferencias entre las habilidades mencionadas en el curso y su uso en la industria, la Tabla 1 muestra dicho comparativo a continuación:

Tabla 1. Comparativa de contenido y su uso en la industria con recursos de GCP.

Habilidad	Cubre el Curso	Se Usa en la Industria
Ciclo de vida de datos	Si	Si (puede ser diferente en cada empresa)
Roles en ciencia de datos	Si	Si
Gobierno de datos	Si, básico	Si
SQL		
BigQuery	Sí (básico)	Si
Cloud Storage	Si (básico)	Si
Dataform	No	Si
Dataflow	No	Sí (prioritario)

Con esto en mente y en miras de responder la viabilidad de la construcción de un *pipeline* de datos con el conocimiento adquirido en este curso, se puede concluir que el curso tiene las bases que le permitirían al aprendiz construir un *pipeline* de datos básico con herramientas nativas, pero que se alejan de los problemas reales que enfrentan las organizaciones de hoy, siendo necesaria una exploración mucho más profunda de las herramientas mencionadas, así como de

soluciones de la vida real que pueden dar ideas más claras de una arquitectura aplicable a problemas verídicos.

3.3. Limitaciones para novatos y profesionales

Continuando con el análisis del curso, respecto a los diferentes niveles de experiencia a los que aplica el conocimiento adquirido en el mismo, es importante mencionar que el curso es útil para entender conceptos básicos como los mencionados en las secciones anteriores, pero insuficiente para empleos *entry-level* que requieren niveles de Python y SQL superiores al que provee el curso (Brown, 2024). Esto es clave ya que implica que una persona que busca comenzar su carrera en ciencia e ingeniería de datos, debe buscar más herramientas de aprendizaje para emplearse y hacer parte de organizaciones reales.

A su vez, vale la pena mencionar que el contenido recibido en el curso, para aquellos que ya son profesionales en el tema, no profundiza en optimización de costos, y otros temas de la nube que son clave en la industria. Esto implica de igual forma que se necesite más material de estudio para hacer uso de herramientas de datos dentro de GCP y por supuesto, horas de práctica para llegar a un aprendizaje adecuado de dichos productos de datos.

Para lograr una exploración un poco más profunda de las herramientas mencionadas en el curso, y otras adicionales que no se tienen dentro del marco teórico del mismo, como las mostradas en la Tabla 1, se recomienda complementar esta información con *bootcamps*, que permitan al estudiante el uso de datos reales y simplificados, así como la continuación de la ruta de aprendizaje de GCP, enfocada a productos de datos.

3.4. Propuestas para cerrar brechas

Desde nuestro punto de vista como profesionales con experiencia en casos reales de proyectos de datos basados en productos de GCP, vemos necesario la inclusión de laboratorios con casos reales que no sólo permitan una creación básica de módulos en Big Query, sino también la exploración de datos estructurados que puedan permitir al estudiante hacerse preguntas sobre la información mostrada, a través del lenguaje SQL, cómo por ejemplo análisis de logs de aplicaciones (ej: Firebase + BigQuery). Esto permitirá al estudiante conocer un poco más de uno de los enfoques que se le puede dar a la herramienta, y motivar aún más el uso de esta.

También merece la pena la actualización del curso con ejemplos de uso de herramientas como DataFlow y Dataform, esto permitirá al estudiante una visión más global de los diferentes productos de datos que existen en GCP y cuándo hacer uso de los mismos, ya que cada uno tiene un uso específico que tiende a confundirse, llevando al estudiante a entender sus diferencias más profundas y tener más claros los criterios de decisión que puedan llevar a hacer uso de una herramienta u otra (Chaabnia, 2023).

4. Conclusiones

Los resultados de este estudio demuestran que el curso Introduction to data analytics in Google Cloud ofrece una base teórica sólida para comprender conceptos clave en la carrera del análisis de datos. Sin embargo, su enfoque pedagógico contrasta con los escenarios laborales reales, donde ejemplos como migraciones y optimizaciones rara vez siguen un modelo idealizado como el que se muestra en el curso.

Estos hallazgos resaltan la utilidad para orientar a profesionales y estudiantes sobre las fortalezas y limitaciones de la formación en herramientas en la nube, especialmente en GCP. Aunque el curso es valioso para adquirir terminología y principios técnicos fundamentales, carece de profundidad en las habilidades demandadas por la industria.

Estos resultados abren las posibilidades a futuras investigaciones donde se evalúa la efectividad de programas de formación complementarios o el impacto de certificaciones especializadas en la empleabilidad de los contenidos. Además, se sugiere agregar estudios de casos reales tanto de éxito como de fracaso en los cursos para reducir la brecha entre teoría y práctica.

Esto nos deja claro que en el mundo de los datos hay que estar al día en conocimientos técnicos y prácticos. La formación técnica debe ir a la par con lo que necesita el mercado, la competencia actual de las capacidades de los profesionales y agregando la capacidad de las inteligencias artificiales, obligan a cada nuevo profesional a estar cada vez mejor formados.

Referencias

- Google Cloud Blog. (2022, Marzo 15). Top data analytics skills to learn in 2022, recommended by Google. <https://cloud.google.com/blog/products/data-analytics/top-data-analytics-skills-to-learn-in-2022-recommended-by-google>.
Cloud.
- Chaabnia, S. (2023, mayo 10). Checklist of skills to master in data engineering on Google Cloud Platform (GCP). <https://medium.com/@saoussen.chaabnia/checklist-of-skills-to-master-in-data-engineering-on-g-o-o-g-l-e-cloud-platform-gcp-39a36807d7f2>
Medium.
- Brown, I. (2024, Enero12). 2024 data science toolkit: Top skills you need to master. LinkedIn. <https://www.linkedin.com/pulse/2024-data-science-toolkit-top-skills-you-need-master-iain-brown-ph-d--qf7ke>.
- University of Wisconsin-Madison. (s.f.). Data Literacy – Data Lifecycle. Data Science at UW. Recuperado el 16 de mayo de 2025, de <https://data.wisc.edu/data-literacy/lifecycle/>
- Airbyte. (2024, Febrero 13). BigQuery Legacy SQL vs Standard SQL. [airbyte.com. https://airbyte.com/data-engineering-resources/bigquery-legacy-sql-vs-standard-sql](https://airbyte.com/data-engineering-resources/bigquery-legacy-sql-vs-standard-sql)
- Bello, S. A., Oyedele, L. O., Akinade, O. O., Bilal, M., Dávila Delgado, J. M., Akanbi, L. A., & Ajayi, A. O. (2021, Febrero). Computación en la nube en la industria de la construcción: casos de uso, beneficios y desafíos. Automatización en la construcción, 122. ScienceDirect. [https://pdf.sciencedirectassets.com/271427/1-s2.0-S0926580520X00124/1-s2.0-S0926580520310219/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEJb%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJIMEYCIQCTJ%2FFis7UrJgSaK0coDWrsZHGNocyAXf867%2Bbv70%2BPxAIhAPisqwV](https://pdf.sciencedirectassets.com/271427/1-s2.0-S0926580520X00124/1-s2.0-S0926580520310219/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEJb%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJIMEYCIQCTJ%2FFis7UrJgSaK0coDWrsZHGNocyAXf867%2Bbv70%2BPxAIhAPisqwV)
- DataScientest. (2020, Octubre 2). Formación en Google Cloud Platform: ¿Cómo convertirse en un experto? atascientest.com. <https://datascientest.com/es/formacion-en-google-cloud-platform>
- Kyriazatis, S. (2024, Septiembre 13). How to Grow Your Skills in Google Cloud Using Google Cloud Skills Boost. <https://medium.com/google-cloud/how-to-grow-your-skills-in-google-cloud-using-google-cloud-skills-boost-01ff30d534c5>
Medium.
- Praveen, B. (2024, Septiembre 6). Comparison and Analysis of Leading Cloud Service Providers (AWS, Azure and GCP). International Journal of Advanced Research in Engineering and Technology (IJARET), 15(3), 266-278. [papers ssrn. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4914145](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4914145)
- Ruiz Bermejo, A., Torralba Musteller, J. M., & Rosa Garcia, L. (2019, Diciembre). PROPUESTA DE PROCEDIMIENTO PARA SELECCIONAR PLATAFORMAS DE GESTIÓN DE NUBE PARA PYME PROPOSAL PROCEDURE TO SELECT CLOUD MANAGEMENT PLATFORMS FOR PYME. Telemática, 18(3), 21-31.
- Tecnozero. (n.d.). AWS vs Azure vs Google Cloud: tendencias de gasto en la nube pública 2022. 2022. <https://www.tecnozero.com/azure/aws-vs-azure-vs-google-cloud-tendencias-de-gasto-en-la-nube-publica-2022/>