



## **Informe académico del curso Google Cloud Data Analytics**

Stefania Nieto Téllez  
Diego Alejandro Mora Suárez

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                       | (Nieto Téllez & Mora Suarez, 2025)                                                                                                                                                              |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Referencia</b>          | Nieto Téllez, S., & Mora Suarez, D. A. (2025). <i>Informe académico del curso Google Cloud Data Analytics</i> [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia. |
| <b>Estilo APA 7 (2020)</b> |                                                                                                                                                                                                 |



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](http://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.....                                                           | 5  |
| 1. Introducción.....                                                   | 7  |
| 2. Servicios en la nube para analítica de datos .....                  | 8  |
| 3. Descripción técnica del curso .....                                 | 9  |
| 3.1. Introducción a la analítica de datos en la nube .....             | 9  |
| 3.2. Administración y almacenamiento de datos en la nube .....         | 9  |
| 3.3. Transformación de datos en la nube .....                          | 10 |
| 3.4. Visualización y narración de datos .....                          | 11 |
| 3.5. Preparación para el trabajo de Analista de Datos en la Nube ..... | 12 |
| 4. Conclusiones.....                                                   | 13 |
| 5. Referencias .....                                                   | 14 |

## **Siglas, acrónimos y abreviaturas**

|                |                                             |
|----------------|---------------------------------------------|
| <b>APA</b>     | American Psychological Association          |
| <b>Cms.</b>    | Centímetros                                 |
| <b>ERIC</b>    | Education Resources Information Center      |
| <b>Esp.</b>    | Especialista                                |
| <b>MP</b>      | Magistrado Ponente                          |
| <b>MSc</b>     | Magister Scientiae                          |
| <b>Párr.</b>   | Párrafo                                     |
| <b>PhD</b>     | Philosophiae Doctor                         |
| <b>PBQ-SF</b>  | Personality Belief Questionnaire Short Form |
| <b>PostDoc</b> | PostDoctor                                  |
| <b>UdeA</b>    | Universidad de Antioquia                    |

## Resumen

En la actualidad, el análisis de datos se ha consolidado como una herramienta clave para la toma de decisiones estratégicas en las organizaciones. La creciente disponibilidad de grandes volúmenes de información y la necesidad de interpretarlos de manera eficiente ha impulsado la adopción de servicios en la nube que facilitan su almacenamiento, procesamiento y análisis. Esta monografía presenta un análisis académico del curso Google Cloud Data Analytics, ofertado a través de la plataforma Google Cloud Skills Boost, destacando su relevancia en la formación de especialistas en analítica y ciencia de datos. El estudio se centra en el impacto de las herramientas de analítica en la nube, particularmente aquellas ofrecidas por Google Cloud, como BigQuery, Dataplex y Dataproc, así como en la estructura y el contenido del curso. A lo largo de cuatro módulos—introducción a la analítica, administración y almacenamiento de datos en la nube, transformación de datos y visualización de información—se exploran conceptos técnicos, metodologías y aplicaciones prácticas que permiten a los estudiantes desarrollar competencias fundamentales para el manejo de datos en entornos empresariales.

En la era del big data, las organizaciones deben procesar grandes volúmenes de datos para extraer información útil que impulse la toma de decisiones. Según Marr (2016), el big data no solo se trata del volumen, sino también de la velocidad, variedad y veracidad de los datos. Las plataformas de computación en la nube, como Google Cloud Platform, permiten escalar el análisis de datos de forma flexible y eficiente, eliminando la necesidad de infraestructuras físicas costosas (Hashem et al., 2015). En este contexto, herramientas como BigQuery permiten realizar análisis SQL sobre conjuntos masivos de datos de manera rápida y eficiente, siendo especialmente útiles en ambientes empresariales (Owais & Benkhelifa, 2017). Además, la visualización de datos es esencial para comunicar los hallazgos de manera clara. Tableau, Power BI y Looker son herramientas ampliamente utilizadas en la industria (Few, 2009; Camm et al., 2020).

El proceso de extracción, transformación y carga (ETL) también ha evolucionado hacia modelos más flexibles como ELT, en los cuales los datos se cargan antes de ser transformados directamente en plataformas como BigQuery (Wang et al., 2021). Este cambio ha permitido agilizar el flujo de datos en arquitecturas modernas como los Data Lakes y Data Warehouses (Inmon & Linstedt, 2014). Mediante escenarios de un problema común, se demuestra cómo la integración de estos servicios puede mejorar significativamente la eficiencia organizacional, optimizar recursos y fomentar una toma de decisiones más informada. En este sentido, el curso

no solo aporta conocimientos técnicos, sino que también ofrece una visión estratégica del papel que juega la analítica en la nube en el contexto actual de transformación digital.

## 1. Introducción

En un mundo donde los datos se han convertido en uno de los activos más valiosos, la capacidad de analizarlos y transformarlos en información útil es esencial para la toma de decisiones estratégicas. La analítica de datos en la nube ha revolucionado la forma en que las organizaciones almacenan, procesan y aprovechan grandes volúmenes de información, permitiendo obtener *insights* en tiempo real y optimizar procesos en diversos sectores (Hashem, Yaqoob, Anuar, Mokhtar, Gani, & Khan, 2015).

El curso de Google Cloud Data Analytics es una ruta diseñada para desarrollar habilidades en herramientas y metodologías clave para la analítica en la nube, como BigQuery (Google Cloud, 2023a), Dataplex (Google Cloud, 2023b) y Dataproc (Google Cloud, 2023c). Con la creciente demanda de profesionales capacitados en estas tecnologías, la formación en este campo se vuelve fundamental para quienes desean destacarse en el ámbito de la ciencia de datos y la inteligencia empresarial.

Este trabajo analiza el impacto de la analítica de datos en la nube, a través de ejemplos empresariales y cómo su implementación puede mejorar la eficiencia en la toma de decisiones. A través del uso de herramientas y la adquisición de conocimientos y habilidades en el manejo de grandes volúmenes de datos, tal que se haga evidente la relevancia en el panorama actual y su potencial para transformar el futuro de la formación de los especialistas en analítica y ciencia de datos.

## **2. Servicios en la nube para analítica de datos**

En la era digital, los datos se han convertido en uno de los activos más valiosos para las organizaciones. La capacidad de analizarlos y transformarlos en información útil es fundamental para la toma de decisiones estratégicas. En este contexto, la analítica de datos en la nube ha revolucionado la manera en que las empresas almacenan, procesan y aprovechan grandes volúmenes de información. A nivel global, existe una amplia variedad de servicios orientados a la analítica de datos, y empresas como Google ofrecen un extenso catálogo de soluciones a través de Google Cloud, además de programas de formación que facilitan el aprendizaje y la adopción de estas tecnologías.

El curso de Google Cloud Data Analytics introduce a los participantes en el creciente campo del análisis de datos, brindándoles conocimientos clave sobre el almacenamiento, organización y administración de información en la nube. Esta estrategia de formación hace uso de ejercicios prácticos que permiten a los estudiantes desarrollar habilidades en procesamiento, análisis y visualización de datos para la toma de decisiones. El curso está estructurado en cuatro módulos: (1) introducción a la analítica; (2) administración y almacenamiento de datos en la nube; (3) transformación de datos en la nube y (4) visualización de información; que buscan proporcionar una formación integral en el uso de estas herramientas en el manejo de grandes volúmenes de datos.

De acuerdo con Zhang et al. (2022), el uso de servicios como Google Cloud Dataflow ha permitido automatizar flujos de trabajo complejos para el procesamiento paralelo de datos, lo cual mejora considerablemente los tiempos de respuesta en entornos empresariales. Asimismo, herramientas como Looker Studio facilitan la visualización interactiva de datos, permitiendo que los analistas creen paneles personalizados que integran múltiples fuentes en tiempo real (Bihanic, 2015).

### **3. Descripción técnica del curso**

A continuación, se realiza la presentación de los módulos académicos que constituyen el curso de interés.

#### **3.1. Introducción a la analítica de datos en la nube**

El curso ofrece una introducción estructurada a la computación en la nube, proporcionando a los participantes los fundamentos esenciales para comprender su funcionamiento y aplicación en el análisis de datos (Google Cloud Skills Boost, s.f.). A lo largo de cuatro módulos enfatiza la diferencia entre el análisis de datos tradicional y en la nube, permitiendo explorar modelos de facturación en la computación en la nube y el framework de arquitectura de Google Cloud, elementos cruciales para evaluar costos y optimizar recursos dentro de una empresa que requiera el uso de esta tecnología (Google Cloud Skills Boost, s.f.). Otro aspecto central es el estudio del ciclo de vida de los datos, que abarca todas sus etapas, desde la entrada hasta la eliminación de información no relevante para el análisis de la información en el contexto del problema de interés. Se destaca la protección y la recuperación de los datos ante eventos de vulneración del protocolo de seguridad, además de su impacto en el cumplimiento normativo y la gestión de costos (Google Cloud Skills Boost, s.f.). Este conocimiento es esencial para garantizar la seguridad de la información en entornos empresariales y minimizar riesgos asociados a la administración ineficiente de los datos.

Según Tankard (2012), la adopción de tecnologías en la nube debe considerar no solo el costo-beneficio, sino también los riesgos asociados a la seguridad y la privacidad. Estas consideraciones son clave en contextos donde se gestionan datos sensibles y se requiere cumplir con marcos regulatorios internacionales.

#### **3.2. Administración y almacenamiento de datos en la nube**

El módulo de administración está diseñado para proporcionar a los participantes conocimientos esenciales sobre la extracción, almacenamiento y organización de datos en la nube. Se inicia con el módulo de extracción y almacenamiento de datos, donde se exploran métodos tales como: el uso de conectores de datos, ingestión por lotes y en tiempo real, así como el almacenamiento en servicios como Cloud Storage y BigQuer.

Estos métodos permiten trasladar información entre sistemas, incluyendo el uso de APIs (del inglés, Interfaz de Programación de Aplicaciones) y conectores. Así mismo, se aborda un componente clave relacionado con la organización de los datos, que expone la utilización de esquemas normalizados y de estrella, así como la gestión eficiente de grandes volúmenes de información (lakehouse). Esta estrategia de formación permite el desarrollo de habilidades prácticas que afiancen habilidades en buscar, administrar y optimizar datos en la nube. Generando mayor familiaridad con herramientas como BigQuery y Dataplex en la identificación y el seguimiento de información almacenada.

Además, los conceptos de almacenamiento orientado a columnas, particionamiento y clustering son relevantes para mejorar el rendimiento en las consultas. Según Abadi et al. (2008), este tipo de arquitecturas permite mejorar significativamente la eficiencia y el tiempo de respuesta en consultas analíticas.

### **3.3. Transformación de datos en la nube**

El módulo de transformación de datos está diseñado para brindar a los participantes habilidades fundamentales en el manejo, depuración y conversión de grandes volúmenes de datos dentro de entornos de nube. Este curso se enfoca especialmente en el uso de herramientas como BigQuery, Cloud Storage y Dataflow para automatizar y escalar procesos de análisis de datos. Se inicia con una introducción al recorrido de los datos, abordando desde su captura inicial hasta su transformación y análisis final. En esta fase, se profundiza en los distintos momentos que componen el “viaje del dato”, los cuales permiten entender cómo la información puede ser transformada en valor para la toma de decisiones organizacionales. Posteriormente se exploran técnicas de transformación utilizando SQL, enfocándose en cómo estructurar consultas para normalizar, limpiar y unir datos provenientes de diversas fuentes. Este apartado permite comprender cómo pasar de datos en bruto a estructuras analíticas óptimas que favorezcan la visualización y el modelado posterior.

Uno de los componentes centrales del curso es el desarrollo de pipelines de transformación de datos, utilizando servicios como Dataflow, que permiten automatizar el flujo de trabajo entre distintas etapas de procesamiento. Se hace énfasis en el diseño eficiente de estos pipelines para asegurar que las transformaciones sean reproducibles, seguras y escalables. Zhang et al. (2022) destacan que la implementación de pipelines de

datos en plataformas en la nube ha mejorado la eficiencia operativa en distintos sectores, al reducir la carga manual y facilitar la gestión continua de grandes flujos de información.

Adicionalmente, se introduce el concepto de planificación estratégica en la transformación de datos, destacando la importancia de identificar los requerimientos del negocio y definir claramente los objetivos de transformación antes de ejecutar cualquier proceso técnico. Esta parte también cubre aspectos clave como la minimización de redundancias, la trazabilidad de datos y la eficiencia computacional. A través de ejercicios prácticos, se refuerzan competencias en la manipulación de datos directamente desde BigQuery, permitiendo ejecutar y probar transformaciones en tiempo real. Asimismo, se explora la integración de estos procesos con Cloud Storage y Dataflow, facilitando el diseño de flujos automatizados que pueden ser desplegados en proyectos de análisis avanzados.

Esta estrategia de formación permite desarrollar una mirada crítica sobre los procesos de transformación, alineando la capacidad técnica con el entendimiento estratégico del dato.

### **3.4. Visualización y narración de datos**

Se inicia con el módulo de narración de datos, en el cual se exploran los principios fundamentales de la comunicación visual y cómo estos permiten conectar con distintos públicos, facilitando la comprensión de los datos y promoviendo la toma de decisiones. También se abordan la planificación de visualizaciones, donde se presentan metodologías para traducir los requerimientos de las partes interesadas en wireframes y prototipos de informes, promoviendo un enfoque centrado en el usuario. Este componente también introduce criterios para seleccionar los gráficos adecuados, de acuerdo con el tipo de información y los objetivos de análisis. El curso también incluye un módulo orientado a la exploración de datos, que permite a los participantes desarrollar habilidades analíticas para identificar patrones, relaciones y valores atípicos. Este proceso es fundamental para sustentar narrativas sólidas basadas en evidencias cuantitativas. Posteriormente, se profundiza en la construcción de visualizaciones mediante herramientas nativas de Google Cloud, destacando el uso de Looker Studio, con el cual se crean informes interactivos que integran datos provenientes de fuentes como BigQuery y Google Analytics. Según Few (2009), el éxito de una visualización no radica solamente en su apariencia, sino en la claridad con la que comunica el mensaje. Además, Bihanic (2015)

señala que las herramientas de visualización modernas deben facilitar la exploración y el descubrimiento interactivo. A través de un laboratorio práctico, los participantes aprenden a representar indicadores clave, como el número de visitantes únicos o las transacciones realizadas durante campañas específicas, e identificar la procedencia geográfica de los usuarios.

Finalmente, el módulo cierra con una sección enfocada en la colaboración y el intercambio de resultados, permitiendo a los usuarios compartir visualizaciones de manera segura y efectiva con sus equipos de trabajo, promoviendo entornos de análisis colaborativos. Esta estrategia de formación permite el desarrollo de competencias clave en la visualización de datos en la nube, consolidando habilidades para diseñar paneles que respondan a necesidades empresariales y fomenten una cultura organizacional basada en datos.

### **3.5. Preparación para el trabajo de Analista de Datos en la Nube**

El primer módulo está pensado para prepararte profesionalmente en el mundo del análisis de datos en la nube. Aquí encontrar recursos útiles y consejos prácticos para mejorar un currículum y prepararte para entrevistas laborales, todo con un enfoque en las demandas del mercado actual.

En segundo y último módulo del curso, se hace énfasis en todo lo aprendido en los módulos (Introducción en la analítica de datos en la nube, Administración y almacenamientos de datos en la nube, transformación de datos en la nube y visualización y narración de datos) anteriores a través de un proyecto final. Este proyecto se basa en el ciclo completo del manejo de datos: desde la recolección y el almacenamiento, hasta el análisis y la visualización. Utilizando herramientas en la nube, trabajarás con datos reales, resolverás problemas prácticos y demostrarás tu capacidad para presentar soluciones claras y efectivas. Este trabajo final se convierte en parte de un portafolio profesional, mostrando los conocimientos adquiridos.

#### **4. Conclusiones**

Como resultado de la realización del trabajo, se demuestra que la implementación de pipelines ETL/ELT en Google Cloud —utilizando BigQuery, Dataplex, Dataproc y Dataflow— permite un flujo de datos eficiente y escalable, capaz de integrar fuentes heterogéneas, garantizar la trazabilidad y mantener la calidad de la información. La aplicación práctica de una estructura modular de aprendizaje facilitó la construcción de soluciones reproducibles, redujo los tiempos de procesamiento y optimizó los recursos computacionales, evidenciando la viabilidad de migrar procesos empresariales tradicionales a entornos en la nube.

Asimismo, la incorporación de técnicas de storytelling y visualización en Looker Studio potenció la capacidad para transformar grandes volúmenes de datos en insights accionables, mejorando la comunicación de hallazgos para distintos públicos. El trabajo final no solo valida las competencias adquiridas, sino que también sienta las bases para futuras investigaciones en optimización de procesos de datos y desarrollo de dashboards interactivos orientados a la toma de decisiones estratégicas en organizaciones.

## 5. Referencias

- Abadi, D. J., Madden, S., & Hachem, N. (2008). Column-stores vs. row-stores: How different are they really? *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, 967–980. doi:10.1145/1376616.1376712
- Ali, M. H., Hosain, M. S., & Hossain, M. A. (2021). Big data analysis using BigQuery on cloud computing platform. *Australian Journal of Engineering and Innovative Technology*, 1-9.
- Bihanic, D. (2015). *Empowering users through design of visual tools for big data exploration*. (D. Bihanic, Ed.) Springer. doi:10.1007/978-3-319-22180-9\_6
- Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., & Anderson, D. R. (2020). *Essentials of business analytics* (3 ed.). Cengage Learning.
- Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*. Analytics Press.
- Google Cloud. (16 de 5 de 2025). *Business insights visualizations*. Obtenido de Google Cloud: <https://cloud.google.com/looker-studio>
- Google Cloud. (16 de 5 de 2025). *Dataplex: Data governance*. Obtenido de Google Cloud: <https://cloud.google.com/dataplex>
- Google Cloud. (16 de 5 de 2025). *Overview of BigQuery analytics*. Obtenido de Google Cloud: <https://cloud.google.com/bigquery/docs/query-overview>
- Google Cloud Skills Boost. (2025). *Introduction to Data Analytics in Google Cloud*. Obtenido de [https://www.cloudskillsboost.google/paths/420/course\\_templates/961](https://www.cloudskillsboost.google/paths/420/course_templates/961)
- Hashem, I. A. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115. doi:10.1016/j.is.2014.07.006
- Inmon, W. H. (2014). *Data architecture: A primer for the data scientist*. Morgan Kaufmann.
- Marr, B. (2016). *Big data in practice: How 45 successful companies used big data analytics to deliver extraordinary results*. Wiley.
- Owais, M., & Benkhelifa, E. (2017). Data analytics in cloud computing: A survey. *International Journal of Cloud Applications and Computing*, 7(4), 1–17. doi:10.4018/IJCAC.2017100101
- Rivery. (16 de 5 de 2025). *ETL vs ELT: Key differences, comparisons, & use cases*. Obtenido de Rivery: <https://rivery.io/blog/etl-vs-elt/>

- Tankard, C. (2012). The security issues of cloud computing. *Network Security*(1), 5-8. doi:10.1016/S1353-4858(12)70011-2
- Wang, J. G. (2021). ELT vs. ETL in cloud computing: A case study. *Journal of Cloud Computing*, 10(1), 1-12. doi:10.1186/s13677-021-00238-z
- Zhang, Y., Zhang, H., & Ren, X. (2012). Dataflow-based processing for real-time big data pipelines in cloud computing. *Journal of Big Data*, 9(1), 22. doi: 10.1186/s40537-022-00567-3