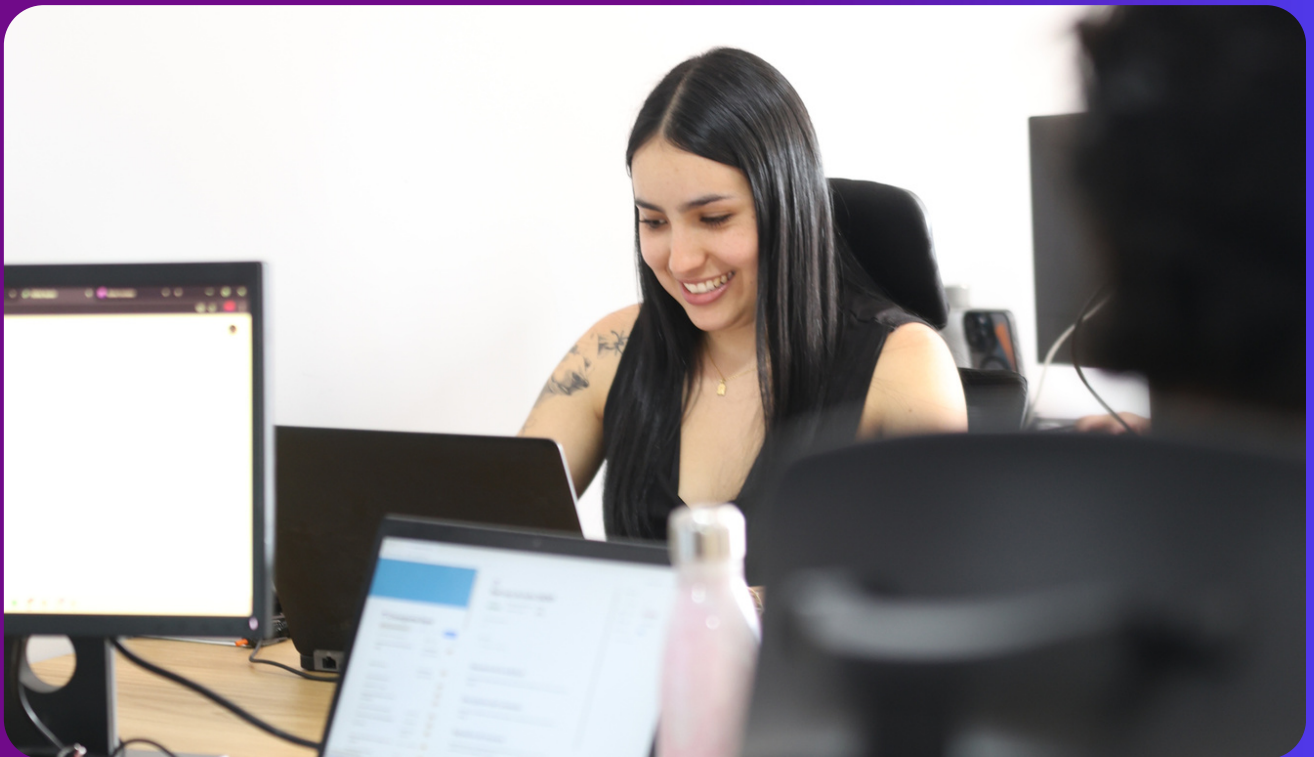


Formación en ingeniería de Datos para IA y Analítica en la Nube

PROGRAMA 2026



Este programa tiene como objetivo formar ingenieros de datos capaces de diseñar, automatizar, gobernar y escalar plataformas modernas de datos en la nube, preparadas para alimentar soluciones de analítica avanzada, inteligencia artificial y modelos generativos.

1

Duración

160h con una intensidad horaria de 9 horas semanales durante 4.5 meses.

2

Formato

100 % virtual y en vivo

3

Metodología

Aprendizaje teórico-práctico, con un proyecto práctico transversal y progresivo.

4

Admisión

Mínimo de 290 / 480 puntos en la prueba de admisión.

6

Mentoría y feedback

Revisión individual y grupal de proyectos y código con expertos de la industria. Corrección de malas prácticas en ingeniería de datos.

7

Panel frente a expertos

Los participantes desarrollarán un proyecto real de ingeniería de datos, integrando los conocimientos adquiridos en Big Data, automatización, procesamiento distribuido y despliegue en la nube.

PERFIL DE INGRESO

02/12

El programa está dirigido a desarrolladores(as), analistas técnicos o profesionales de datos con al menos 2 años de experiencia en desarrollo de software o manejo de datos, que deseen especializarse en ingeniería de datos moderna, escalable, automatizada y orientada a IA.

CONOCIMIENTOS OBLIGATORIOS

- ↘ Programación en Python nivel intermedio: funciones, estructuras de datos, manejo de archivos, POO básica y consumo de APIs.
- ↘ Bases de datos relacionales y SQL: consultas, joins, agrupamientos, creación de tablas y comprensión de llaves primarias/foráneas.
- ↘ Uso básico de Git y GitHub: clonar repositorios, commits, push, pull, ramas e integración de cambios.
- ↘ Familiaridad con APIs REST y herramientas como Postman o consumo desde Python.
- ↘ Lectura técnica en inglés para interpretar documentación oficial, errores y buenas prácticas.

CONOCIMIENTOS DESEABLES

- ↘ Automatización y scripting con Bash/Shell.
- ↘ Nociones básicas de cloud computing, especialmente AWS S3, EC2, RDS o Lambda.
- ↘ Experiencia en backend, integraciones, ETL, reportes o proyectos con datos.

Al finalizar, los participantes estarán en capacidad de desempeñarse como Ingenieros/as de Datos con visión moderna, preparados para operar en equipos de datos, analítica, IA y tecnología. Serán capaces de construir plataformas que conectan ingesta, almacenamiento, procesamiento, automatización, gobierno, visualización y consumo por modelos de inteligencia artificial.

- ↘ Diseñar pipelines escalables batch y real-time orientados a analítica, BI, machine learning y GenAI.
- ↘ Modelar datos relacionales, NoSQL, multidimensionales y estructuras orientadas a consumo por modelos.
- ↘ Construir data lakes y arquitecturas lakehouse en AWS con gobierno, linaje, particionamiento y control de acceso.
- ↘ Transformar datos con PySpark y optimizar jobs distribuidos para escenarios de producción.
- ↘ Automatizar workflows con Airflow, CI/CD y Terraform, incorporando monitoreo y buenas prácticas DevOps.
- ↘ Implementar validaciones de calidad con Great Expectations y criterios específicos para datos usados por IA.
- ↘ Diseñar features reutilizables para modelos predictivos, segmentación o detección de anomalías.
- ↘ Implementar una arquitectura RAG básica o asistente de datos conectado a información empresarial preparada.
- ↘ Crear dashboards ejecutivos en QuickSight y explicar decisiones técnicas ante un panel experto.

STACK TECNOLÓGICO

CATEGORÍA	HERRAMIENTAS / ALCANCE
Core obligatorio	Python, SQL, Bash, Git/GitHub, Docker básico.
Bases de datos	PostgreSQL, Redshift, MongoDB o DynamoDB como comparación NoSQL.
Procesamiento	Apache Spark / PySpark, jobs batch y fundamentos de optimización.
Streaming	Apache Spark / PySpark, jobs batch y fundamentos de optimización.
Orquestación y DataOps	Apache Airflow, GitHub Actions, Great Expectations.
Cloud AWS	S3, Glue, Glue Catalog, Redshift, Lambda, Lake Formation, DataZone, IAM, CloudWatch.
Infraestructura como código	Terraform con módulos reutilizables y despliegue documentado.
Visualización	Amazon QuickSight conectado a modelos analíticos.
IA aplicada	Feature engineering, SageMaker Feature Store conceptual, embeddings, vector search, RAG.

MÓDULOS DEL PROGRAMA

MÓDULO 1

Fundamentos de Ingeniería de Datos, Cloud e IA



Duración
20 horas

Este módulo introduce los conceptos esenciales de la ingeniería de datos moderna, con énfasis en su relación con la inteligencia artificial, la analítica y las arquitecturas cloud.

ALCANCE

- Rol del ingeniero de datos en proyectos de IA.
- Metodologías CRISP-DM y ASUM-DM.
- Gobierno de datos, calidad, linaje y privacidad.
- Arquitecturas modernas de datos.
- Concepto de data lakehouse.
- Patrones de ingesta de datos.

MÓDULO 2

Bases de Datos Relacionales y NoSQL para Analítica e IA



Duración
18 horas

Este módulo aborda el diseño, consulta e integración de bases de datos relacionales y NoSQL para soportar procesos analíticos y modelos de inteligencia artificial.

ALCANCE

- SQL avanzado.
- Normalización de bases de datos.
- Modelado relacional.
- Uso de MongoDB o DynamoDB.
- Particionamiento de datos.
- Clustering y sharding.
- Integración de fuentes de datos para consumo analítico.
- Preparación de datos para modelos de IA.

MÓDULO 3

Arquitectura Escalable, Lakehouse y Datos para IA



Duración
22 horas

Este módulo profundiza en arquitecturas escalables para la gestión de grandes volúmenes de datos, orientadas a BI, machine learning e inteligencia artificial generativa.

ALCANCE

- Hadoop y HDFS como base conceptual.
- Uso de Hive e Impala.
- AWS Glue y Glue Catalog.
- Formatos y tecnologías como Iceberg y Hudi.
- Lake Formation y DataZone.
- Gestión de metadatos.
- Linaje de datos.
- Estructuras de datos listas para BI, ML y GenAI.

MÓDULO 4

Procesamiento Distribuido con Spark



Duración
20 horas

Este módulo desarrolla las capacidades necesarias para procesar, transformar y optimizar grandes volúmenes de datos mediante Apache Spark y PySpark.

ALCANCE

- Fundamentos de Apache Spark.
- Uso de PySpark.
- DataFrames y RDDs.
- Transformación y limpieza de datos.
- Persistencia de datos.
- Optimización con Catalyst y Tungsten.
- Ejecución de jobs Spark en ambientes de producción.

MÓDULO 5

Streaming y Eventos para IA Operacional



Duración
16 horas

Este módulo introduce el procesamiento de datos en tiempo real y su aplicación en casos de uso operacionales apoyados por inteligencia artificial.

ALCANCE

- Fundamentos de streaming.
- Comparación entre Kafka y Kinesis.
- Productores y consumidores de eventos.
- Procesamiento de eventos de sensores.
- Generación de alertas.
- Identificación de anomalías.
- Casos de uso operacionales en tiempo real.

MÓDULO 6

DataOps, Calidad, Observabilidad y Automatización**Duración**
20 horas

Este módulo se enfoca en la automatización, monitoreo y aseguramiento de calidad de pipelines de datos mediante prácticas DataOps.

ALCANCE

- Uso de Airflow.
- Creación y gestión de DAGs.
- Scheduling de procesos.
- Logging y trazabilidad.
- Automatización con GitHub Actions.
- Validación de datos con Great Expectations.
- Monitoreo de datos.
- Pruebas automatizadas.
- Automatización de pipelines.

MÓDULO 7

Modelado Analítico, Feature Engineering y Feature Store**Duración**
20 horas

Este módulo aborda el diseño de modelos analíticos y la creación de variables reutilizables para proyectos de analítica, machine learning e inteligencia artificial.

ALCANCE

- Modelado conceptual, lógico y físico.
- Modelo multidimensional.
- Diseño de bodegas de datos.
- Diseño de métricas.
- Creación de features.
- Reutilización de variables para modelos.
- Feature store como patrón de arquitectura.

MÓDULO 8

Cloud, DevOps y Arquitecturas GenAI en AWS



Duración
24 horas

Este módulo integra servicios cloud, prácticas DevOps y componentes de inteligencia artificial generativa en AWS para el despliegue de soluciones modernas de datos e IA.

ALCANCE

- Uso de S3, Glue, Redshift y Lambda.
- Infraestructura como código con Terraform.
- Gestión de permisos con IAM.
- Monitoreo con CloudWatch.
- Encriptación y anonimización de datos.
- Fundamentos de RAG.
- Embeddings y búsqueda vectorial.
- Despliegue de componentes de IA.

enyo*i* by AceleraTI



TELÉFONO
(57) 313 710 7072



PÁGINA WEB
acelerati_latam



INSTAGRAM
acelerati_latam



LINKEDIN
AceleraTI_Latam