Ciencia de Datos
Ecuador

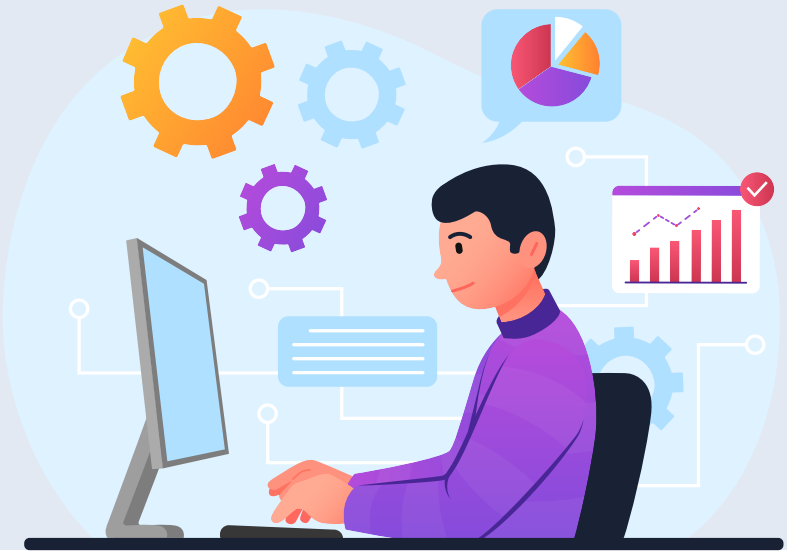
INGENIERO DE DATOS

Programa Online en Vivo

- **Certificación: 147 horas**
- **Proyectos reales**
- **Duración: 14 semanas**
- **Demo day**
- **Preparación certificación internacional**



Actualmente, las empresas generan grandes volúmenes de datos a partir de múltiples fuentes como sistemas, aplicaciones y plataformas digitales. Sin embargo, transformar estos datos en información útil para análisis o inteligencia artificial requiere procesos estructurados que aseguren su calidad, disponibilidad y escalabilidad.



¿El rol del Ingeniero de Datos se ha vuelto fundamental. Un estudio de MIT Technology Review en conjunto con Snowflake revela que el 75% de los líderes empresariales considera a estos profesionales críticos para el éxito del negocio, cifra que asciende al 86% en grandes organizaciones [1], consolidándolos como un perfil no solo demandado, sino estratégico.

1. Redefining data engineering in the age of AI
https://www.technologyreview.com/2025/10/23/1125651/redefining-data-engineering-in-the-age-of-ai/

OBJETIVO DEL PROGRAMA:

Formar profesionales capaces de diseñar, construir y optimizar sistemas de datos para recopilar, transformar y organizar información a gran escala, garantizando su calidad, disponibilidad y escalabilidad.

¿EN QUÉ ROLES PODRÁS TRABAJAR?

Ingeniero de Datos

Diseña y automatiza los flujos que transforman millones de datos en información lista para impulsar decisiones.

Ingeniero de Datos en la Nube

Construye soluciones de datos escalables utilizando tecnologías y plataformas cloud.

Ingeniero de Plataformas de Datos

Desarrolla la infraestructura tecnológica que permite gestionar y procesar grandes volúmenes de información.

Desarrollador de Integración de Datos

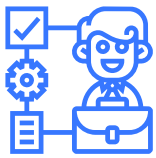
Conecta sistemas, transforma información y automatiza procesos esenciales para las empresas.

INSIGHT DEL MERCADO LABORAL

Según la guía salarial de LinkedIn, la ingeniería de datos es uno de los campos con mayor proyección, con un crecimiento cercano al 50% anual y salarios altamente competitivos. [2]

En este contexto, el diferencial ya no es solo técnico, quienes logran destacar son los que comunican su impacto en negocio y construyen una marca personal sólida.

Estadística clave:



ResumeGo: Los candidatos que incluyen un perfil completo de LinkedIn en su currículum obtienen una tasa de respuesta del 13,5 %, frente al 7,9 % de quienes no lo tienen, lo que representa un incremento del 71 % en oportunidades laborales. [3]



The personal Brand Studio (2025): Los candidatos con una marca personal bien definida obtienen un 40% más de respuesta por parte de los reclutadores y es menos probable que pasen desapercibidos. [4]



LinkedIn-Hiring Lab (2025): El 68 % de los reclutadores prioriza candidatos con experiencia práctica demostrable (proyectos o portafolios) por encima de credenciales académicas. [5]

Este programa integra formación en habilidades blandas, marca personal y comunicación profesional, asegurando que tu conocimiento técnico se traduzca en visibilidad, confianza y mayores oportunidades laborales.

¿SERÉ CAPAZ DE CONSTRUIR PIPELINES REALES?

Sí. Durante este programa aprenderás a diseñar flujos de datos end-to-end, desde la ingesta hasta el consumo analítico. Trabajarás con tecnologías modernas como SQL, Spark, Airflow y arquitecturas cloud, construyendo soluciones reproducibles, escalables y listas para producción.



2. Guía salarial para ingenieros de datos: qué esperar como nuevo profesional.
https://www.linkedin.com/pulse/data-engineering-salary-guide-what-expect-new-professional-czrlf/
3. Cómo influye LinkedIn en las posibilidades de entrevista de los solicitantes de empleo.
https://www.resumego.net/research/linkedin-interview-chances/
4. Cómo impulsar el crecimiento profesional a través de la marca personal.
https://www.thepersonalbrandstudio.com/post/the-impact-of-personal-branding-on-professional-success-what-recruiters-data-reveals?utm_source
5. How to Build a Data Science Portfolio That Gets You Hired
https://www.linkedin.com/pulse/how-build-data-science-portfolio-gets-you-hired-newtonschool-g8hpc/

MODALIDAD Y DURACIÓN Y HORARIO

Semanas:	14 semanas
Duración:	147 horas académicas
Modalidad:	Online en vivo
Horario:	19:00 a 21:00
Clases:	Lunes a jueves

¿QUIÉN PUEDE INSCRIBIRSE EN ESTE CURSO?

Analistas o perfiles BI

Profesionales que trabajan con datos y desean evolucionar hacia roles más técnicos, comprendiendo cómo se construyen y gestionan los pipelines de datos.

Ingenieros o desarrolladores

Perfiles técnicos que buscan especializarse en arquitectura de datos, procesamiento y construcción de sistemas escalables.



Profesionales en transición hacia data

Personas con conocimientos básicos en SQL o programación que desean iniciar una carrera en el mundo de datos desde el lado de la ingeniería..



Data Analysts o científicos de datos junior

Profesionales que ya trabajan con análisis o modelos y buscan fortalecer la base técnica para trabajar con datos a mayor escala y en entornos productivos.



Profesionales de áreas estratégicas orientadas a datos

Perfiles de negocio (finanzas, operaciones, logística, retail) que necesitan comprender cómo se estructuran y gestionan los datos para mejorar la toma de decisiones.

¿POR QUÉ ELEGIR ESTE CURSO?

Proyectos aplicados a la industria

Desarrollarás 9 proyectos basados en casos reales, trabajando con datos y escenarios de sectores como retail, banca, telecomunicaciones y energía. Además, crearás un repositorio en GitHub donde documentarás tu progreso y resultados.

Demo Day

Presentarás tu proyecto final a un equipo de expertos (simulación stakeholders), donde podrás aplicar tu conocimiento técnico en la resolución de problemas de negocio y tus habilidades para contar historias.

Certificación Internacional

Te prepararás para rendir la certificación Data Engineer Professional de Databricks, que valida y respalda tu capacidad para trabajar con tecnologías como Spark, Delta Lake y arquitecturas Lakehouse, en entornos reales.

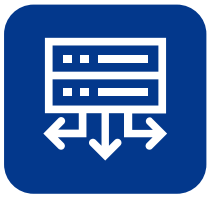
Enfoque en empleabilidad

Construirás un portafolio técnico con proyectos reales que podrás presentar en procesos de selección y fortalecerás tu marca profesional.



Mentorías y acompañamiento personalizado

Recibe soporte continuo, resolución de dudas y feedback experto para avanzar con claridad durante todo el programa.



Herramientas y tecnologías de la industria

Trabajarás con tecnologías utilizadas en entornos profesionales como SQL, Spark, Airflow y arquitecturas en la nube (AWS, Azure, Databricks, GCP).

¡TODO PERFECTO!

AQUÍ INICIA TU VIAJE COMO INGENIERO DE DATOS



CONTENIDO	MODALIDAD
Fundamentos de bases de datos	En vivo
Ingeniería de datos Multi-Cloud	En vivo
Big Data con Spark	En vivo
DataOps + Producción	En vivo
Boost Profesional (CV, marca personal, portafolio de proyectos, entrenamiento entrevistas de trabajo)	En vivo
Mentorías	En vivo

Para aprovechar al máximo esta experiencia de aprendizaje, te sugerimos contar con un nivel básico de SQL y Python, una computadora con al menos 8 GB de RAM y 100 GB de espacio disponible en tu disco.

CONTENIDOS DEL PROGRAMA:

1. FUNDAMENTOS DE BASES DE DATOS

Aprende a estructurar, transformar y modelar datos para su uso analítico, comprendiendo cómo se construyen los flujos de datos desde su origen hasta su consumo.

PROYECTOS:

Retail: Construye un consolidador de ventas por tienda y categoría, aplicando SQL para obtener métricas clave como ventas, ticket promedio y productos líderes.



Banca: Desarrolla un modelo analítico con esquema dimensional y SQL avanzado para calcular KPIs como clientes activos, volumen transaccional y ranking de sucursales.



HABILIDADES QUE VAS A APRENDER:

1. **Comprender el flujo completo de datos**
   Identificar fuentes, ingesta, almacenamiento, transformación y consumo en un pipeline end-to-end.

2. **Dominar consultas SQL fundamentales**
   Construir queries con SELECT, WHERE, ORDER BY, GROUP BY y HAVING para analizar datos.

3. **Generar datos listos para análisis**
   Crear tablas agregadas con resúmenes clave que faciliten la toma de decisiones.

4. **Diseñar modelos de datos eficientes**
   Construir esquemas dimensionales (hechos y dimensiones) y elegir entre bases relacionales y no relacionales.

5. **Integrar datos con JOINs estratégicos**
   Aplicar INNER, LEFT, SEMI y ANTI JOIN según el problema analítico.

6. **Analizar datos con SQL avanzado**
   Usar CTEs y funciones de ventana para calcular rankings, cohortes y top-N por grupo.

7. **Definir y medir KPIs de negocio**
   Establecer indicadores con fórmula, fuente y periodicidad, trazando el linaje de datos.

2. INGENIERÍA DE DATOS MULTI-CLOUD

Aprende a diseñar y construir pipelines de datos en la nube utilizando servicios modernos en AWS, GCP y Azure, integrando almacenamiento, procesamiento y consumo de datos en arquitecturas escalables.

PROYECTOS: DATA PRODUCT POR NUBE

AWS (Agroexportación):
Construye un data product que monitorea despachos y tiempos de entrega, consolidando datos en un data lake. Habilita consultas rápidas para detectar retrasos y publica datasets curados en un warehouse para análisis histórico y reportes operativos.



GCP (Movilidad urbana):
Desarrolla una tabla analítica para una operadora de buses que mide puntualidad y demanda por ruta. Procesa eventos de validaciones por parada para generar indicadores por franja horaria y detectar rutas críticas para optimizar frecuencias.



Azure (Energía):
Implementa un data product que integra lecturas de consumo energético por zona. Orquesta la preparación de datos y habilita análisis en SQL para identificar picos, anomalías y patrones por día y hora, facilitando decisiones operativas.



HABILIDADES QUE VAS A APRENDER:

1. **Gestionar acceso y seguridad de datos**
   Configurar permisos mínimos y validar accesos sobre datasets y tablas.

2. **Consultar datos en entornos cloud**
   Ejecutar consultas SQL sobre data lakes y obtener resultados reproducibles.

3. **Diseñar arquitecturas de datos eficientes**
   Publicar datasets y comparar cuándo usar data lake vs data warehouse.

4. **Comprender ecosistemas de almacenamiento**
   Diferenciar entre Data Lake y Data Warehouse según el caso de uso.

5. **Construir y consultar data warehouses**
   Cargar datos, aplicar particionado y realizar análisis con SQL serverless.

6. **Procesar datos para análisis**
   Transformar flujos de datos en batch y generar salidas listas para consumo.

7 **Orquestar pipelines de datos**
Automatizar ingesta y transformación mediante flujos visuales con trazabilidad.

8 **Publicar y explotar datasets curados**
Habilitar tablas analíticas y ejecutar consultas para generar insights.

9 **Trabajar con herramientas modernas en la nube**
Explorar entornos como Databricks para análisis y procesamiento de datos.

3. BIG DATA CON SPARK

Aprende a procesar grandes volúmenes de datos utilizando tecnologías de Big Data, construyendo pipelines escalables que permiten transformar y analizar información de forma eficiente.

PROYECTO:

MOTOR DE INGESTA METADATA-DRIVEN (FINTECH)

Un proyecto, desarrollado en 3 entregables:

Preparación con Python y Spark:
Prepara un dataset histórico de transacciones con Python, pandas y PySpark, definiendo un esquema explícito y métricas básicas, con apoyo de IA para generar y validar transformaciones.



Motor metadata-driven en PySpark:
Construye un motor configurable que ingiere al menos dos fuentes de datos, conserva los datos raw, ejecuta transformaciones analíticas reutilizables y publica datasets curados, con una mejora de performance demostrable.



Motor robusto batch y streaming
Extiende el motor para soportar incrementalidad, idempotencia y backfills, además de procesar eventos de Kafka en tiempo real. Publica resultados en Delta Lake, con checkpoints, recuperación ante fallos y capacidades de auditoría.



HABILIDADES QUE VAS A APRENDER:

1. **Construir datasets curados con Spark**
   Leer datos desde distintas fuentes, organizarlos correctamente y prepararlos como datasets confiables y listos para análisis.

2. **Comprender la ejecución en Spark**
   Entender cómo Spark procesa la información paso a paso, incluso cuando no ejecuta todo de inmediato (optimización automática del procesamiento).

3. **Consultar datos con Spark SQL**
   Realizar consultas similares a SQL para analizar datos y validar resultados con conteos y revisiones básicas.

4. **Garantizar calidad y consistencia de datos**
   Asegurar que los datos tengan la estructura correcta, con tipos bien definidos (números, texto, fechas) y sin errores.

5. **Aplicar transformaciones analíticas**
   Combinar datos (joins), resumir información (agregaciones) y calcular métricas avanzadas como rankings o top de resultados.

6. **Optimizar procesamiento de datos**
   Detectar procesos lentos o ineficientes y aplicar mejoras para que el manejo de datos sea más rápido y eficiente.

7 **Procesar datos en tiempo real**
Trabajar con datos que llegan continuamente (como eventos o transacciones) y procesarlos al momento para generar resultados actualizados.

8 **Gestionar evolución y trazabilidad de datos**
Manejar cambios en la estructura de los datos a lo largo del tiempo y poder revisar versiones anteriores para auditoría o control.

9 **Diseñar ingestas configurables**
Procesar distintas fuentes (archivos, APIs, eventos) con una misma lógica basada en configuración, sin duplicar código por cada fuente nueva.

10 **Generar y validar código con apoyo de IA**
Usar IA para generar y traducir transformaciones entre SQL y PySpark, y verificar que el código conserva la lógica esperada.

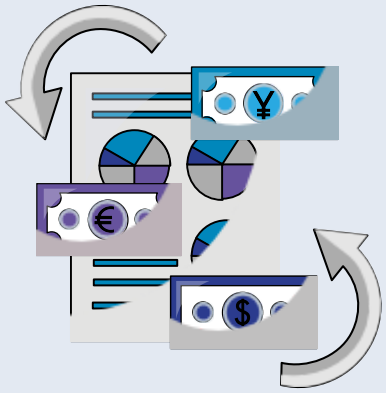
4. DATAOPS Y PRODUCCIÓN

Aprende a llevar tus pipelines de datos a entornos productivos, asegurando su calidad, automatización y funcionamiento continuo en escenarios reales.

PROYECTO:

Pipeline local financiero listo para operación:

Convierte una transformación en Python en un pipeline local que valida la calidad de los datos, protege la información sensible mediante pseudonimización y aplica migraciones de esquema versionadas, ejecutándose de forma automatizada y repetible con apoyo de IA.



HABILIDADES QUE VAS A APRENDER:

1. **Asegurar calidad de datos en pipelines**
   Implementar controles en cada etapa del proceso (valores nulos, duplicados, rangos válidos) y comparar resultados para asegurar que los datos sean confiables.

2. **Proteger datos sensibles**
   Definir accesos solo para quienes lo necesitan y aplicar técnicas para ocultar o proteger información sensible (anonimización o enmascaramiento).

3 **Gestionar cambios en estructuras de datos**
Administrar modificaciones en la forma en que se organizan los datos, asegurando que los sistemas sigan funcionando sin interrupciones (migraciones versionadas con Alembic).

4 **Automatizar control de calidad de código**
Incorporar herramientas que revisan automáticamente el código y validan su correcto funcionamiento dentro del proceso de datos.

5 **Orquestar y recuperar flujos de datos con Airflow**
Diseñar DAGs con dependencias, reintentos y seguimiento de resultados, y demostrar que el pipeline puede identificar un fallo y recuperarse sin intervención manual.

6 **Ejecutar pipelines de forma automatizada**
Programar la ejecución de procesos de datos para que se realicen de forma automática y consistente, sin necesidad de intervención manual.

7 **Documentar y comunicar el resultado con apoyo de IA**
Generar reportes de calidad, un runbook operativo y documentación técnica del pipeline con apoyo de IA, listos para que otra persona pueda operarlo.

5. BOOST PROFESIONAL: EMPLEABILIDAD Y MARCA PERSONAL

Aprende a construir y comunicar tu perfil profesional de forma estratégica, fortaleciendo tu CV, portafolio y presencia en LinkedIn, mientras desarrollas habilidades de comunicación, autogestión y presentación profesional orientadas al mundo laboral real.

PROYECTO:

Desarrollo y optimización de tu perfil profesional, que incluye CV, perfil de LinkedIn y portafolio en GitHub, alineados a roles de Ingeniero de Datos.



HABILIDADES QUE VAS A APRENDER:

1. Construir un CV orientado a roles de Ingeniero en Datos.
2. Mejorar tu perfil de LinkedIn para aumentar visibilidad ante reclutadores.
3. Organizar y presentar tu portafolio de proyectos en GitHub.
4. Comunicar tu experiencia y habilidades de forma clara y profesional.
5. Prepararte para entrevistas y procesos de selección técnicos.

¿POR QUÉ ESTUDIAR EN CIENCIA DE DATOS?

Nuestro método de enseñanza se enfoca en dos grandes ejes: Aprendizaje basado en proyectos (PBL) y sesiones en vivo de aprendizaje con mentoría.



¿QUIÉNES SERÁN MIS PROFESORES?



Danny Díaz

Sr.Full Stack AI Engineer en IOET

Ingeniero en Sistemas con maestría en Inteligencia Artificial y más de 5 años de experiencia en proyectos de datos e IA. Especializado en el diseño y optimización de pipelines ETL, procesamiento de Big Data y arquitecturas de datos en la nube. Cuenta con experiencia en la implementación de soluciones de ingeniería de datos para el sector financiero, utilizando herramientas como Python, SQL, PySpark, Azure Databricks, Data Factory y AWS.



Byron Córdova

Analista de Datos Senior en EPAM NEORIS

Ingeniero Estadístico con experiencia en ingeniería de datos, automatización y analítica avanzada. Actualmente desarrolla procesos ETL y pipelines de datos, integrando información desde bases de datos, APIs y múltiples fuentes, con énfasis en calidad, consistencia y trazabilidad. Cuenta con experiencia en Python, SQL, R, bases de datos relacionales y NoSQL, y en el desarrollo de soluciones automatizadas y escalables para el procesamiento de datos.



Álvaro Polo

Senior Data Engineer Analyst en Farmaenlace

Ingeniero con Maestría en Sistemas de Información, mención en Inteligencia de Negocios y Analítica de Datos Masivos. Especialista en ingeniería de datos, DataOps y automatización de procesos. Cuenta con experiencia en el desarrollo de pipelines, procesos ETL y transformación de grandes volúmenes de datos utilizando Python, SQL Server, Spark y Airflow. Además, implementa soluciones de inteligencia artificial y Machine Learning para optimizar procesos empresariales.

¿ENTREGAN ALGÚN CERTIFICADO?

Al finalizar el programa, recibirás un certificado de aprobación por 147 horas con una insignia digital verificada con credencial, la cual puede ser vinculada en LinkedIn.







Adicional, te entregaremos un certificado por cada curso aprobado y también tendrás entrenamiento para tomar el examen de certificación internacional, Data Engineer Professional de Databricks.

*La certificación internacional no esta incluida en el programa

¿CUÁL ES EL VALOR DEL CURSO?

El programa INGENIERO DE DATOS es una inversión en tu desarrollo profesional y empleabilidad. A lo largo de la formación adquirirás habilidades técnicas, desarrollarás un portafolio de proyectos, optimizarás tu perfil profesional y te prepararás para obtener una certificación internacional.

La inversión total del programa es de USD 1699

*Consulta con nuestro equipo sobre becas, formas de pago, financiamiento, descuentos especiales y promociones vigentes, disponibles por tiempo limitado.

FORMAS DE PAGO

Ciencia de Datos Ecuador CDEC S.A.S.
RUC: 1793166237001
info@ cienciadedatosec.com

- **Transferencia bancaria**
- **Depósito**

Banco Produbanco:
Cuenta de ahorros: 12040582502

Banco del Pichincha
Cuenta Corriente: 2100336717

- **Tarjetas de crédito**

PayPhone PayPal

Ciencia de Datos
Ecuador

@CienciaDeDatosEcuador

+593 97 863 5237
+593 98 414 3301

info@cienciadedatosec.com
www.cienciadedatosec.com