

**Básico - Big Data**

# **PROGRAMA DE ESTUDIOS: INTRODUCCIÓN A BIG DATA**



**KOICA**

**IGU** HANDONG GLOBAL  
UNIVERSITY



**UNA**

## Para empezar el curso:

1

Es necesario que el estudiante tenga conocimientos básicos de programación.

2

El estudiante debe contar con una computadora y conexión a internet.

3

Manejo básico de PC para configurar herramientas necesarias para el curso.

## Información Esencial:

**Inversión** Pago único de 450.000 Gs.

**Inicio** 06 de abril del 2026

**Duración** 8 semanas (40 horas totales).

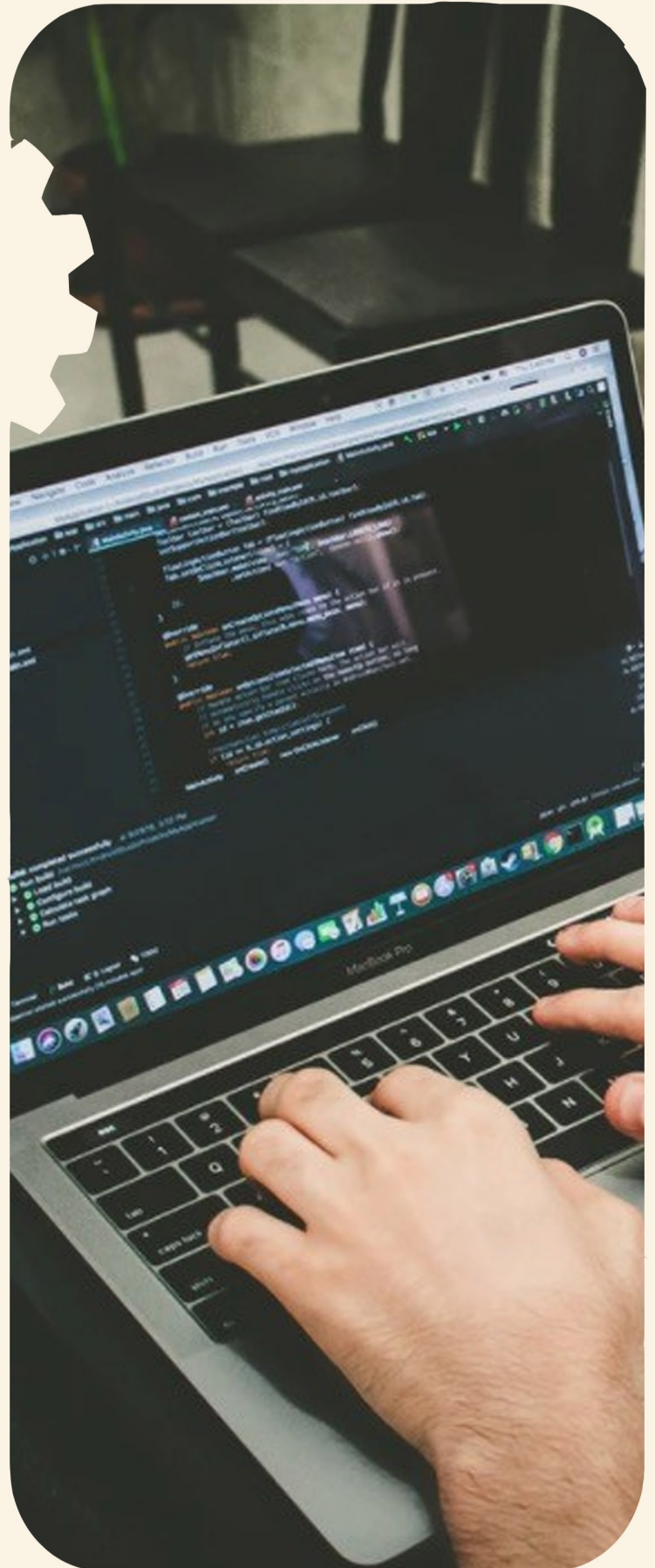
**Carga Horaria** 5 horas semanales: Clases virtuales e interactivas.

**Horario** Lunes y Miércoles, 19:00 a 21:00 h. Viernes (asíncrona)

**Modalidad** 100% Online / Sincrónico a través de la plataforma EDUCA.

Este curso introduce los fundamentos de Big Data desde una perspectiva aplicada, centrada en los problemas que resuelve y en la estructura de un ecosistema moderno de datos. Durante 8 semanas, el estudiante construirá de forma incremental un pipeline completo que incluye ingesta, validación, transformación, carga y consulta. Se trabajará con Python, PostgreSQL y Apache Airflow como herramientas principales.

Se enfatizan buenas prácticas de Arquitectura e Ingeniería de Datos como diseño por capas, cargas incrementales, control de calidad y trazabilidad. El enfoque es práctico, con laboratorios guiados y entregables progresivos que culminan en un proyecto final integrador.



# Objetivos del curso

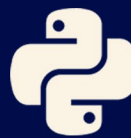
- Explicar conceptos clave (4Vs, batch vs incremental, arquitectura por capas).
- Configurar entorno local (Python, PostgreSQL, Git, Airflow).
- Modelar datos en PostgreSQL (tablas, claves, diseño básico analítico).
- Implementar ingestas en Python (CSV/JSON/API) con validaciones.
- Transformar y cargar datos por capas (staging → core → datamart).
- Documentar y presentar un proyecto reproducible.

## Objetivos específicos



### FUNDAMENTOS

Comprender las bases conceptuales de Big Data y sus arquitecturas principales.



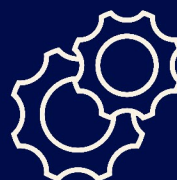
### INGESTA

Implementar procesos de extracción y transformación inicial de datos con Python.



### MODELADO

Diseñar y gestionar estructuras de datos en PostgreSQL con criterios básicos de organización.



### ORQUESTACIÓN

Automatizar y documentar un pipeline de datos utilizando Apache Airflow.



## Perfil del egresado

El egresado será capaz de comprender y aplicar los fundamentos de Big Data en entornos prácticos, participando en la construcción de pipelines de datos básicos de principio a fin. Podrá realizar procesos de ingesta, transformación y carga de datos utilizando Python y PostgreSQL, aplicando buenas prácticas de organización, calidad y trazabilidad.

# Plantel Docente

El curso es dictado por el Prof. Ing. Richard Daniel Jiménez Riveros Ingeniero en Informática, con sólida formación en desarrollo de software, bases de datos y arquitectura de soluciones tecnológicas. Se desempeña como docente en el área de Tecnologías de la Información, impartiendo asignaturas vinculadas a programación, sistemas de información y gestión de datos. Cuenta con experiencia en el diseño e implementación de soluciones orientadas al procesamiento, integración y análisis de datos.



**Prof. Ing. Richard Jiménez**

# Cronograma

Semana	Módulo	Enfoque Principal
Semana 1	Fundamentos	Conceptos básicos de Big Data y presentación del proyecto.
Semana 2	Arquitectura	Diseño por capas y estructura del pipeline.
Semana 3	Entorno	Configuración de herramientas y modelado inicial.
Semana 4	Ingesta	Extracción y validación básica de datos con Python.
Semana 5	Transformación	Procesamiento y carga en PostgreSQL.
Semana 6	Orquestación	Automatización del flujo con Airflow.
Semana 7	Optimización	Mejoras, documentación y preparación de demo.
Semana 8	Presentación	Exposición y evaluación del proyecto final.

# Contenido del curso

## Introducción a Big Data

Conceptos generales, 4Vs y aplicaciones en entornos reales.

## Arquitectura de Datos

Ecosistema de datos, diseño por capas y flujo de un pipeline.

## Procesamiento de Datos

Batch, incremental y nociones básicas de almacenamiento y transformación.

## Orquestación y Buenas Prácticas

Automatización básica, organización del proyecto y documentación técnica.

*Fundamentos: Principios esenciales de programación y manejo estructurado de datos.*

*Colecciones: Uso de listas, tuplas, diccionarios y conjuntos en Python aplicados al procesamiento de datos.*

# Distribución de la Calificación y Condiciones para Aprobar

**La evaluación del curso se basa en un enfoque práctico y progresivo:**

**Actividades y Tareas (Plataforma):**

Desarrollo de ejercicios técnicos y avances del pipeline en la plataforma EDUCA, incluyendo scripts en Python, consultas en PostgreSQL y participación en foros para reforzar los conceptos trabajados en clase.

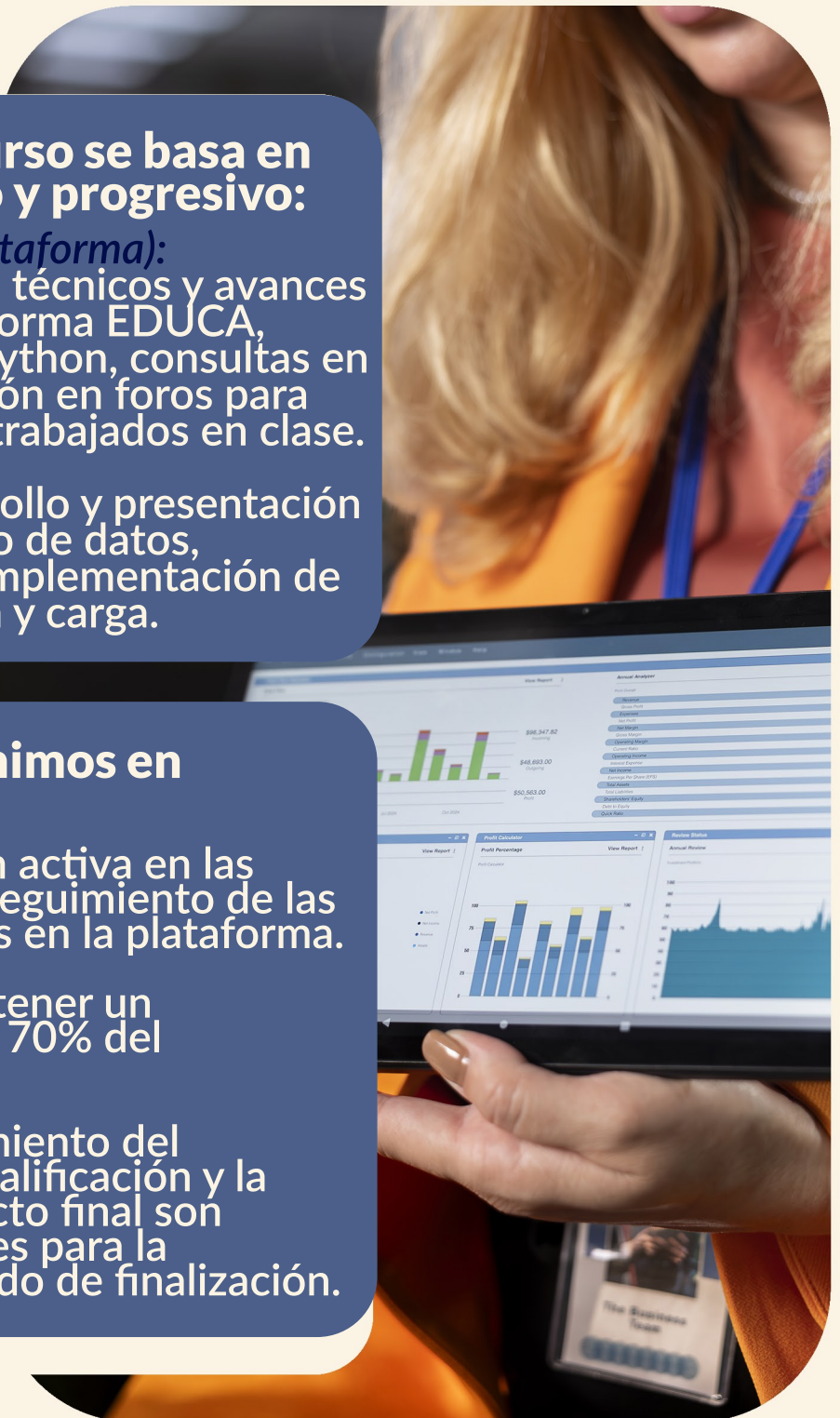
**Evaluación Final:** Desarrollo y presentación de un pipeline completo de datos, evaluando la correcta implementación de ingesta, transformación y carga.

**Requerimientos mínimos en cada módulo:**

**Asistencia:** Participación activa en las sesiones sincrónicas y seguimiento de las actividades asincrónicas en la plataforma.

**Calificación Mínima:** Obtener un rendimiento superior al 70% del puntaje total del curso.

**Certificación:** El cumplimiento del porcentaje mínimo de calificación y la presentación del proyecto final son requisitos indispensables para la expedición del certificado de finalización.





[cit.pol.una.py](http://cit.pol.una.py)

