

Online /  Perú

Programa de Especialización Analítica

DATA ENGINEER

BROCHURE 2023 / I




#AprendeDesdeCasa
#AprendeConLosPioneros



PRESENTACIÓN

El avance tecnológico ha promovido el crecimiento exponencial de los datos hasta lo que hoy conocemos como Big Data y, consecuentemente, han surgido nuevas técnicas y herramientas de recolección, organización y almacenamiento de grandes volúmenes de datos. Hoy en día, los proyectos tecnológicos requieren de profesionales capacitados en ingeniería de datos que puedan crear, mantener y gestionar el Big Data durante todo su ciclo de vida a fin de garantizar su permanente disponibilidad.

Por ello, presentamos el Programa de Especialización **"Data Engineer"** que te permitirá comprender y dominar las tecnologías vinculadas a la Ingeniería de Datos y Big Data, dominando para ello medios de almacenamiento SQL y No-SQL e implementando sus respectivos sistemas ETL, aplicando sistemas de carga Batch y Real-Time. Adicionalmente, conocerás técnicas y buenas prácticas que podrás incluir en los flujos de datos como son el Web Scraping y la metodología DevOps.

Pre-Requisitos	 Inicio: 30/03/23 - Fin: 19/09/23
<ul style="list-style-type: none">• 6 meses de experiencia profesional en áreas relacionadas a despliegue de servidores, análisis de datos, Big Data y afines.• Conocimientos básicos en sistemas distribuidos, despliegue de servidores, programación y bases de datos.	 176 horas académicas
	 Martes y jueves de 7:30pm a 10:30pm

PRESENTACIÓN

OBJETIVOS



Objetivo General

El participante podrá comprender y dominar las tecnologías de ingeniería de datos y big data, empleando e implementando SQL, No-SQL, ETL, sistemas de carga batch y real-time para mostrar información valiosa a través de dashboards; así como dominar buenas prácticas de web scraping y DevOps.



Objetivos específicos

- Comprender la función del almacenamiento SQL y No-SQL con grandes volúmenes de datos, así como conocer los sistemas ETL en la ingeniería de datos y los fundamentos de su implementación con Python.
- Comprender la arquitectura y patrones de diseño del big data, implementar un data lake y los procesos batch para carga de datos y posterior explotación con Spark, así como trabajar con Apache Kafka.
- Conocer e implementar las herramientas y metodologías que completan el ciclo de la ingeniería de datos como son la visualización de datos, web scraping y DevOps.
- Comprender la infraestructura cloud en torno al big data, los proveedores disponibles y los servicios para su implementación.

DIRIGIDO A

- Profesionales que se desarrollen y tengan experiencia comprobable en áreas de programación, manejo y despliegue de servidores y diseño de bases de datos.
- Profesionales y estudiantes de últimos ciclos de las carreras de ingeniería de sistemas, software, informática o afines con conocimiento comprobable sobre las asignaturas relacionadas a la programación, servidores, bases de datos y sistemas distribuidos.



COMPETENCIAS

Al finalizar el programa el alumno será capaz de:

- Construir una solución de ingeniería de datos para big data sobre un caso real con flujos ETL, carga de datos batch/real-time, visualización de datos y web scraping, on-premise u on-cloud y DevOps.
- Implementar medios de almacenamiento SQL y No-SQL aplicables a entornos para grandes volúmenes de datos, y flujos ETL básicos con el lenguaje Python.
- Implementar un data Lake y los procesos batch relacionados para su carga de datos y explotación con Spark y Apache Kafka.
- Utilizar las herramientas y metodologías que completan el ciclo de la ingeniería de datos como visualización de datos, web scraping y DevOps.
- Comprender la infraestructura cloud en torno al big data, los proveedores disponibles y los servicios para su implementación.



CARACTERÍSTICAS

Clases en Vivo

El 100% de las clases que se desarrollan en el programa son en vivo.

Asesoría Académica

Resuelve tus dudas con el asesor académico en línea

Plataforma E-Learning

Accede en cualquier momento a materiales complementarios: lecturas, videos, tutoriales, clases grabadas y más.



Aprende haciendo

Desarrolla casos con datos reales, incluso puedes proponer casos de tu propio sector.

Proyecto integrador

Pondrás a prueba tus ideas para convertirlas en soluciones analíticas.

Soporte técnico

Asistencia técnica permanente y acceso a máquinas virtuales de ser necesario.

MALLA CURRICULAR

01

STORAGE & ETL
ESSENTIALS

Data Storage Essentials
(28 horas académicas)

Python Essentials for ETL
(20 horas académicas)

02

BIG DATA
INGESTION

Big Data Batch
(24 horas académicas)

Big Data Real Time
(16 horas académicas)

03

TOOLS FOR
DATA ENGINEER

Data Visualization
(16 horas académicas)

Web Scraping con Python
(12 horas académicas)

DevOps
(28 horas académicas)

04

BIG DATA ON CLOUD
FUNDAMENTALS

Big Data Infrastructures
(16 horas académicas)

PROYECTO DE
INGENIERÍA DE DATOS



1

Data Storage Essentials

Utilizar herramientas de bases de datos relacionales y no relacionales.

SQL Data:

- Archivos planos y SQL.
- Consultas generales en SQL.
- Consultas con múltiples tablas, subconsultas y vistas.
- Manejo de tablas temporales.

No-SQL Data:

- ¿Qué es No-SQL?
- Query Create, Drop Database and Collection.
- Insert, Find.
- Sort, Delete, Update.
- Limit & Join

2

Python Essentials for ETL

Sistemas ETL con lenguaje Python, tanto para con orígenes on-premise y on-cloud.

- Introducción a Python y ETLs.
- Trabajo de datos con Pandas.
- Conexión a bases de datos SQL y No-SQL.
- Extracción, Transformación y carga de datos con Python.

3

Big Data Batch

Arquitectura y patrones de diseño del Big Data, implementar un Data Lake y los procesos batch.

- Arquitectura y patrones de diseño sobre Big Data.
- Hadoop y el almacenamiento distribuido.
- Procesamiento distribuido con Hive.
- Implementación de un data lake y flujos ETL.
- Spark y procesamiento estructurado.
- Spark y procesamiento semiestructurado.

4

Big Data Real Time

Carga de datos hacia un repositorio en tiempo real, con data proveniente de IoT, redes sociales, etc.

- ¿Qué es real-time?
- ¿En qué se diferencia de una carga Batch?
- Tecnología para carga Real-Time en Big Data.
- Uso y configuración de Apache Kafka para carga de datos Real-Time.



5

Data Visualization

Tableros y dashboards a partir de un origen Big Data con visualización de datos.

- Inteligencia de negocios y herramientas de visualización.
- Power BI for Big data.
- Transformación de datos con Power Query.
- Uso de DAX.

6

Web Scraping con Python

Realizar scraping de páginas web como medio para poblar bases de datos.

- Introducción a Web Scraping.
- Debate legal-ético para el escraqueo.
- Web scraping: Conceptos prácticos.
- Web scraping: Métodos de extracción de datos.
- Técnicas de web scraping.
- Web scraping estático.
- Web scraping requests.
- Web scraping dinámico.
- API's

7

DevOps

Control de versiones de proyectos, manejo y despliegue de servidores con distribución que usa Linux.

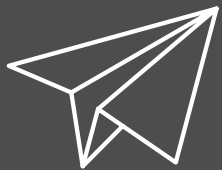
- Linux: Introducción y comandos.
- Linux: Perfil y Procesos.
- Redes, Conexiones y LOGS.
- Git.
- Jenkins.

8

Big Data Infraestructures

Seleccionar la infraestructura Big Data on-cloud más adecuada, según el mercado.

- Big Data en AWS.
- Big Data en Azure.
- Big Data en GCP.
- Big Data en Cloudera.



PROYECTO INTEGRADOR

Elaboración, seguimiento y presentación de un proyecto de ingeniería de datos.

Desarrolla una solución de ingeniería de datos para un entorno Big Data aplicable a una casuística real, que incluya flujos ETL, carga de datos batch/real-time, visualización de datos y web scraping, pudiendo ser un proyecto on-premise u on-cloud e incluir prácticas DevOps.



EVALUACIÓN

- + Evaluaciones parciales: 40%
- + Trabajo final: 60%
- + Nota mínima aprobatoria: 14
- + Asistencia mínima: 80%



CERTIFICACIÓN

CERTIFICADOS PARCIALES COMO ESPECIALISTA EN:

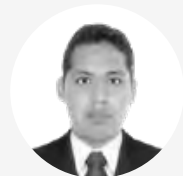
- + Storage & ETL Essentials. *48 horas acad.*
- + Big Data Ingestions. *40 horas acad.*
- + Tools for Data Engineer. *56 horas acad.*
- + Big Data On Cloud Fundamentals. *16 horas acad.*

CERTIFICADO FINAL:

- + Especialista en Data Engineer *176 horas académicas.*

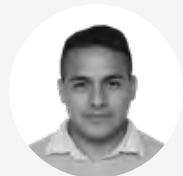
* Todos los certificados se emitirán de forma digital luego de concluido el programa.

PLANA DOCENTE



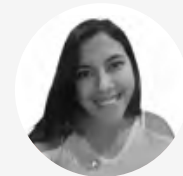
ANGEL TINTAYA

Senior Data
Engineer



JUAN SALINAS

Senior Data
Engineer



MILAGROS VILLEGAS

Business Intelligence
Consultant



FRANCIS DE LA CRUZ

Big Data Architect



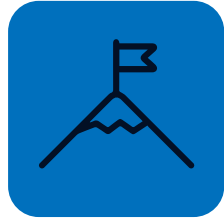
VICTOR GUTIERREZ

Data Architect



*En caso de contingencias podría cambiar alguno de los docentes por otro profesional de similar perfil.

Buscamos liderar la transformación de las empresas.



+14 Años de Experiencia

Desde el año 2009 capacitamos con técnicas de análisis de datos a profesionales de diferentes empresas y sectores.



+15K Profesionales Capacitados

Nuestros alumnos inscritos pertenecen a las mejores compañías del medio y amplifican con nosotros su red de contactos especializada.



+300 Empresas Asesoradas

Las empresas top del mercado buscan nuestra asesoría y les brindamos soluciones analíticas ad hoc.

Formando profesionales mediante la analítica de los datos.



+100 Expertos en Analítica

Nuestra plana docente ocupa los cargos más importantes en las áreas analíticas de todos los sectores.



+50 Capacitaciones Especializadas

Contamos con una variedad de líneas temáticas y niveles de especialización.



+5 Big Data Analytics Summit

Organizamos el evento más grande del Perú, con los mejores ponentes internacionales.



DiHC
Perú

