CURSO

DATA ENGINEER

Plan de estudio

educación IT

Nuestro propósito

Transformar positivamente la vida de las personas.

Queremos que seas protagonista en la transformación que estamos viviendo. Por eso, nos comprometemos a capacitarte para que estés al día con las necesidades digitales actuales.

Te invitamos a trabajar en conjunto para que descubras tu mejor versión y la potencies. Anímate, toma las riendas de tu futuro.

Code your future!

Contenido del curso

Elabora un proceso de extracción, transformación y carga en un repositorio unificado. Escala la estructura de información a partir de la ingeniería de datos.

Proyecto Integrador

Creación de un proyecto escalable a una arquitectura Big Data

Elaborarás un proceso de ETL (extracción, transformación y carga de datos) en un Datawarehouse. Luego, evolucionarás hacia un data lake en un clúster Hadoop.

- Reconocerás las distintas fuentes de datos, las dimensiones y sus métricas.
- Diseñarás un modelo de almacenamiento de datos centralizado.
- Aplicarás los procesos de transformación necesarios según los criterios de calidad del dato y los procesos de limpieza y normalización.
- Tomarás decisiones de diseño de arquitectura relacionada a la optimización de los recursos del clúster.
- Utilizarás herramientas como HDFS, Hive y Spark para las transformaciones.
- Replicarás esa estructura de ETL en un entorno Big Data.

Este trabajo formará parte de tu portfolio personal, y te servirá como experiencia profesional. Al finalizar, podrás subir tu proyecto a GitHub y compartir el diploma en Linkedin para destacar tu perfil utilizando @educacionit desde tus redes.

¿Qué aprenderás?

- Responsabilidades del data engineer.
- Confección de un pipeline de datos (ETL).
- Gobernanza de datos.
- Data Warehouse y Data Lake.
- Implementación de un proyecto Big Data.
- Hadoop y Spark.
- Componentes Core de Hadoop.
- Virtualización y Docker.
- Formatos de almacenamiento en Big Data.
- Bases de datos SQL y NoSQL (Hive).
- Analytics y Machine Learning en Big Data.
- Orquestación de flujo de datos.

Plan de estudios

1 Ingeniería de datos

- Concepto de ingeniería de datos.
- Data Governance.
- Ciclo de vida del dato.
- Preparación del dato.
- Integración de datos.
- Limpieza de datos .
- Normalización de datos.
- Imputación de valores faltantes.
- Transformación de datos.
- Reducción de datos.

2 Modelo de entidad relación

- Modelo ERD.
- Primera forma normal (1FN).
- Segunda forma normal (2FN).
- Tercera forma normal (3FN).
- Cuarta forma normal (4FN).
- Quinta forma normal (5FN).
- ETL: extracción, transformación y carga.
- Indexación de tablas.
- Optimización de consultas SQL.
- Cargas incrementales.

3 Big Data

- Casos de uso.
- Data lake.
- Estrategias de procesamiento.
- Arquitectura de un data lake.
- Implementación de Hadoop.
- Ecosistema Hadoop.
- Clúster Hadoop.
- Sistemas de archivos de cluster (HDFS).
- Escritura en HDFS.
- Monitor de recursos con YARN.
- Ejecución de Jobs en YARN.
- MapReduce.
- Vendors y frameworks.
- Virtualización y Containerización.
- Docker.
- Arquitectura Docker.
- Imágenes y contenedores.
- Docker Compose.
- Procesamiento por Lotes (Batch).

4 Hive y formatos de almacenamiento

- Hive.
- HiveQL.
- Tipos de tablas.
- Tipos de datos.
- Formatos de almacenamiento.
- Particiones.
- Hue.

- Ranger y Altas.
- Factores de elección.
- Row vs Column.
- Schema Evolution.
- Compression.
- CSV.
- JSON.
- Avro.
- Parquet.
- Hadoop Compression Types.
- Hive Ser/Des.
- Casos de uso.

5 Bases estructuradas y no estructuradas

- SQL.
- NoSQL.
- ACID.
- BASE.
- Teorema CAP.
- Tolerancia a particiones.
- Consistencia.
- Disponibilidad.
- Tipos de Motores NoSQL.
- DynamoDB.
- HBase.
- Cassandra.
- Neo4J.
- MongoDB.

- Apache Drill.

6 Apache Spark

- Arquitectura Spark.
- RDD.
- Operaciones sobre RDD's .
- Arquitectura Hadoop - Spark.
- Spark Core .
- SQL.
- Streaming.
- Machine Learning.
- GraphX.
- Fuentes de Datos.
- Lenguajes Soportados.
- DataFrame.
- DataSet.
- Catalyst Optimizer.

7 Clustering de datos

- Streaming Processing.
- Arquitectura Lambda.
- Tecnologías.
- Kafka.
- Spark Streaming.
- Apache Flink.
- Ejemplo MapR.
- Ejemplo Cloudera.
- Ejemplo Airbnb.

- Analytics.
- Machine Learning.
- Casos de uso.
- Aprendizaje Supervisado.
- Regresión.
- Clasificación.
- Aprendizaje No Supervisado.
- Clustering.
- Filtro Colaborativo.

8 Streaming de Datos

- Spark ML.
- Sparklyr.
- Spark GraphX.
- Inteligencia de Negocios.
- Ciclo de Vida del Dato.
- Calidad del Dato.
- Visualización del Dato.
- Workflows.
- DAG.
- Notación Cron.
- Orquestadores.
- Oozie.
- Nifi.
- Airflow.
- Computadora cuántica.
- Principios de la computadora cuántica.
- Qbits.

- Programación.
- Tipos de problemas que resuelve .
- Futuro de la computación cuántica.
- Conclusiones

Modalidad del Curso

Duración
8 semanas / 30 h

Frecuencia semanal
2 encuentros de 2 h

Modalidad
Online en vivo

Grupos reducidos
Promedio 15 personas

Nivel: Experto

- [ ] Principiante
- [ ] Intermedio
- [ ] Avanzado
- [x] Experto

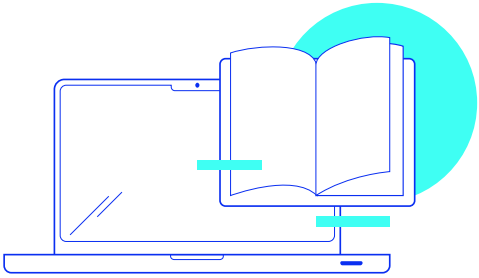
Requisitos

Es recomendable tener una base sobre:

- **Introducción a Python**
- **Introducción a Base de Datos**
- **MongoDB Fundamentos**

Dedicación fuera de clase

Además de las horas de clase, recomendamos que inviertas 4 h semanales extras para realizar los desafíos complementarios, estudiar el material de lectura y completar los exámenes del Alumni.



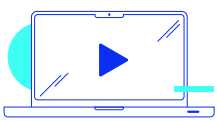
¿Cómo será tu experiencia?

Aprender haciendo

Ejercita y pon en práctica lo estudiado.

Trabajo en equipo

Une esfuerzos y potencia los resultados.



Clases grabadas

Consúltalas las veces que quieras.

Profesores expertos

Aprende de gigantes de la industria.

Asistente académico

Recibe soporte dentro y fuera de clase.

Plataforma Alumni

Encuentra recursos, materiales y clases.

¿Por qué Educación IT?

IT Créditos

Gana puntos al aprobar los exámenes de los cursos. Luego, podrás canjearlos por nuevos cursos sin costo alguno. Los IT Créditos que acumules no vencen ni se devalúan.

Garantía de aprendizaje

Si necesitas reforzar conceptos, recuperar clases o no estás satisfecho, ¡vuelve a tomar el curso sin ningún costo! Puede ser de forma total o parcial.

Comunidad en Discord

Mantente en contacto con la comunidad de EducaciónIT a través de nuestro servidor de Discord. Podrás hablar con tus compañeros, profesores, asistentes académicos y soporte.

Career Advisor

Ingresa al mundo laboral junto a nuestros asesores de carrera: crea un CV que impacte, arma y comparte tu portfolio en LinkedIn y Behance y ten simulacros de entrevistas.

Preguntas frecuentes

Si me pierdo una o más clases, ¿puedo recuperarlas?

Todas las clases quedan grabadas de por vida en tu plataforma Alumni. ¡Siempre podrás volver a verlas cada vez que lo necesites!

¿Cómo voy a aprender?

Te enfrentarás a situaciones de trabajo reales, en donde tendrás que aplicar lo aprendido de forma individual y en equipo. Por medio de la prueba y el error, irás superando desafíos y obteniendo nuevas habilidades que luego podrás aplicar en el ámbito laboral.

¿Cómo son las clases online en vivo?

Las clases duran entre 2 y 3 horas de lunes a viernes (sábados 3 o 4 hs) y se desarrollan de forma online en vivo en aulas virtuales, donde vas a poder interactuar con el instructor y tus compañeros. Manejamos cupos reducidos para que puedas tener un seguimiento más personalizado durante tu aprendizaje.

educación IT

Plan de estudio

Code your future

www.educacionit.com

@educacionit