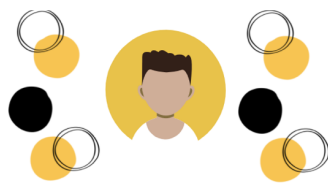


# Data Engineering



## ANALISTA DE DATOS

Explica los datos de una forma que sea fácil de entender y resumida. Usa los datos para tomar mejores decisiones

### Habilidades:

- Estadística 
- Comunicación 
- Conocimiento del negocio

### Herramientas:

Excel

Power BI




SQL



## INGENIERO DE DATOS

preparar los datos para el análisis, su objetivo es desarrollar, construís, probar, mantener las arquitecturas de datos necesarios para obtener la información.

### Habilidades:

- Matemáticas 
- Programación 
- Big Data 

### Herramientas:

SQL




Python



## CIENTÍFICO DE DATOS

Utiliza varios algoritmos avanzados de aprendizaje automático para identificar la ocurrencia de un evento particular en el futuro.

### Habilidades:

- Matemáticas 
- Programación 
- Big Data 

### Herramientas:

SQL

Python



La profesión de ingeniero de datos es una de las especializaciones que van ganando enteros en el ecosistema del Big Data. El Emerging Jobs Report de 2020 de LinkedIn la puso entre las 15 ocupaciones emergentes más destacadas de los últimos cinco años, compartiendo protagonismo con puestos tan sugerentes como experto en Inteligencia Artificial, científico de datos o Site Reliability Engineer. En este listado, los ingenieros de datos se sitúan en la octava posición, con la constatación por parte de LinkedIn de que su contratación ha crecido un 35% entre 2015 y 2019.

## ¿Qué es un ingeniero de datos?

El ingeniero de datos es el profesional encargado de sentar las bases para la adquisición, el almacenamiento, la transformación y la gestión de los datos en una organización. Este especialista asume la configuración de la infraestructura tecnológica necesaria para que el gran volumen de datos no estructurados recogidos se convierta en materia prima accesible para otros especialistas en Big Data, como los data analysts y los científicos de datos.

Los data engineers trabajan diseñando, creando y manteniendo la arquitectura de las bases de datos y de los sistemas de procesamiento, de manera que la posterior labor de explotación, análisis e interpretación de la información pueda llevarse a cabo sin incidencias, de manera ininterrumpida, segura y eficaz.

## ¿Qué hace un ingeniero de datos en su día a día?

El día a día del ingeniero de datos transcurre, fundamentalmente, entre procesos ETL (Extract, Transform, Load), es decir, desarrollando tareas de extracción, transformación y carga de datos, moviéndose entre diferentes entornos y depurándolos para que lleguen normalizados y estructurados a las manos de analistas y data scientists. El papel del data engineer es, en este caso, comparable al de un fontanero, ya que se centra en implementar y mantener en buen estado la red de pipelines (tuberías) por la que los datos (a semejanza del agua) correrán para alimentar el funcionamiento de toda la organización.

### 1. Extracción

En la primera etapa del proceso ETL, el ingeniero de datos se encarga de sacar los registros de distintas localizaciones, así como de estudiar la incorporación de nuevas fuentes al flujo Big Data de la compañía. Estos datos se presentan en diferentes formatos, integrando variables muy diversas, y pasarán a un data lake, u otro tipo de repositorio donde esta información quedará almacenada en bruto, disponible para cualquier uso futuro.

### 2. Transformación

En un segundo paso, el data engineer coordina la limpieza de los datos, eliminando duplicados, corrigiendo errores y desechando el material inservible; y los elabora y clasifica para convertirlos en un conjunto homogéneo.

### 3. Carga

Finalmente, el ingeniero de datos lidera la carga de estos en su destino, ya sea este una base de datos ubicada en un servidor propio de la compañía o un data warehouse en la nube. Además de la correcta exportación, una de las preocupaciones recurrentes en esta etapa final es la vigilancia de la seguridad, puesto que el data engineer ha de garantizar que la información se guarda a salvo de ciberataques y de accesos no autorizados.

