¿Qué habilidades debe tener un Data Engineer?

Para dedicarte a la **Data Engineering** necesitas tener una visión práctica y especializada sobre el ámbito de los datos y las nuevas necesidades de las empresas. Por ejemplo, deberás saber cómo se modelan los datos y cómo funcionan las **BBDD SQL**.

Por otro lado, los data engineer también realizan y programan las ingestas de datos y hacen **procesos de limpieza, validación, Data Quality y agregación para que la información le llegue correctamente al científico de datos**. También configuran el clúster en Spark para que los modelos estadísticos se ejecuten de forma efectiva.

Entre las tecnologías que usan, deberás conocer **Linux y Git** si quieres trabajar en proyectos de software, así como **Hadoop y Spark** a nivel de entornos, **HDFS, MongoDB y Cassandra** a nivel de tecnologías **NoSQL y Map Reduce** a nivel de modelos computacionales, por ejemplo. Conocer **SQL** como lenguaje de programación principal es muy importante.

ambién hay que tener conocimientos de Machine Learning y de herramientas de Big Data y DataWarehouse, como Hive o Kafka. Por otro lado, dependiendo de la empresa, también suele ser imprescindible dominar alguno de los siguientes lenguajes:

* **Python**: para el procesado de datos

**Big Data**

Los Ingenieros de datos son los profesionales de datos que **preparan la infraestructura de "big data"** para ser analizados por los científicos de datos

Los Data Engineers son los que diseñan, desarrollan, construyen, prueban y mantienen los sistemas de procesamiento de datos en el proyecto de Big Data.

Debe ser capaz de **diseñar y construir los datawarehouses** (repositorios de datos en bruto) y definir cómo se recogen y transforman los datos para conseguir datos listos para analizar

La meta de un ingeniero de datos es realizar tareas relacionadas con el almacenamiento de datos, ETL (extraer, transformar y cargar.

Habilidades: Hadoop, MapReduce, Hive, Pig, Data streaming, **NoSQL, SQL**, programación como java**, JavaScript y python**(para el procesado de datos).

Herramientas: DashDB**, MySQL, MongoDB,** Cassandra, **hadoop** y spark

**diseñar y construir los datawarehouses** y definir cómo se recogen y transforman los datos para conseguir datos listos para analizar

modelar los datos y cómo funcionan las **BBDD SQL**

**procesos de limpieza, validación de datos.**

**Tester de datos**

* **Dala layer**
* **Sitios nuevos**
* **Apps nuevas**
* **Migraciones**

**Introducción a scrum**