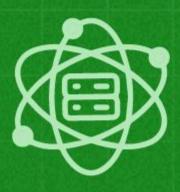


Datacademy

Semana 2



Data science en las empresas y sus equipos





La jerarquía de necesidades de data science

IA, Deep learning

Pruebas A/B, experimentación, algoritmos simples ML

Analytics, métricas, segmentación, agregaciones, características, entrenamiento de datos

Limpieza, detección de anomalías, preparación

Datos confiables, flujo, infraestructura, pipelines, ETL, datos estructurados y no estructurados

Instrumentación, logging, sensores, datos externos, contenido generado por el usuario

Aprendizaje/ Optimización

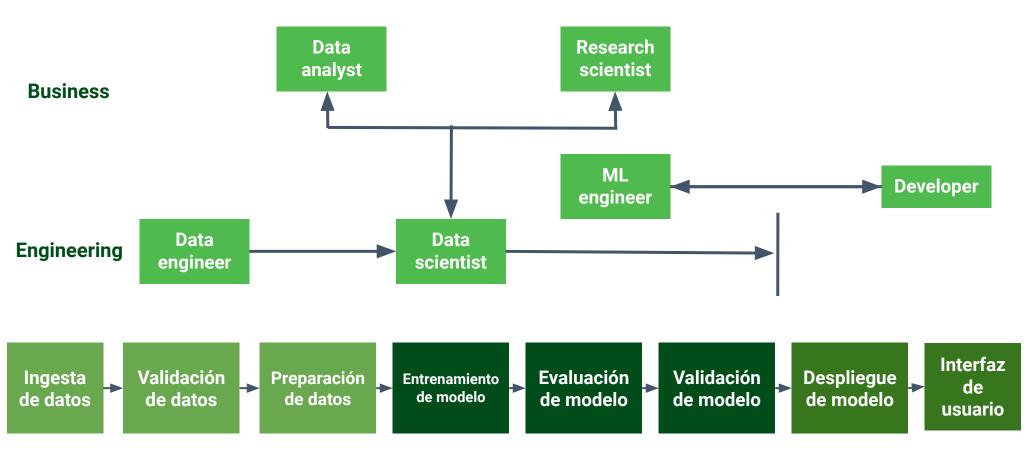
> Agregaciones/ Etiquetado

> > **Exploración/ Transformación**

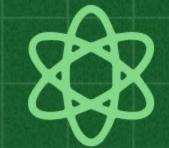
Movimiento/ Almacenamiento

Recolección

Referencia: 2. Data Science Hierarchy of needs (Monica Rogati — Hackernoon)



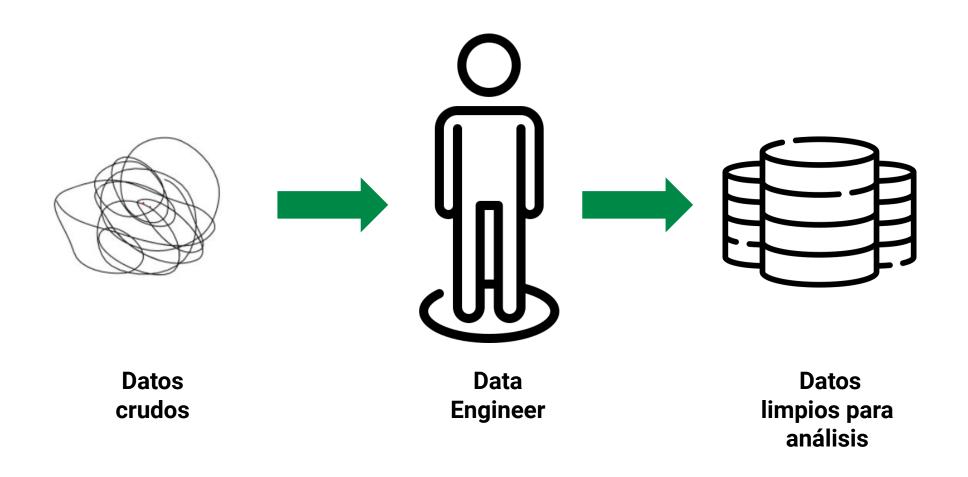
Pasos y roles en el flujo de trabajo de data science (Design Patterns in Machine Learning).



¿Qué hace una Data Engineer?



¿Qué hace una Data Engineer?

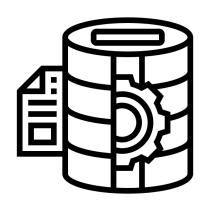


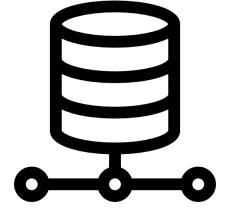


¿Qué hace un Data Engineer?

Trabaja para que el equipo tenga datos para análisis.

Crea pipelines ETL.

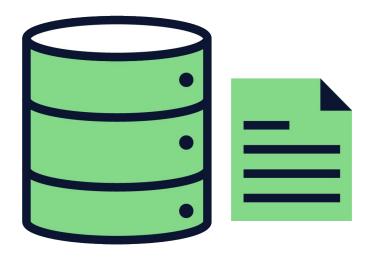






Día a día de una Data Engineer

- Data pipelines de ETL y bases de datos.
- Extraer datos de diferentes fuentes.





 Transformar los datos para análisis.

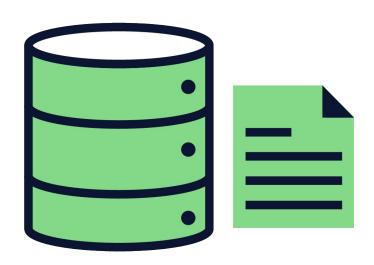
 Bases de datos especializadas para análisis.



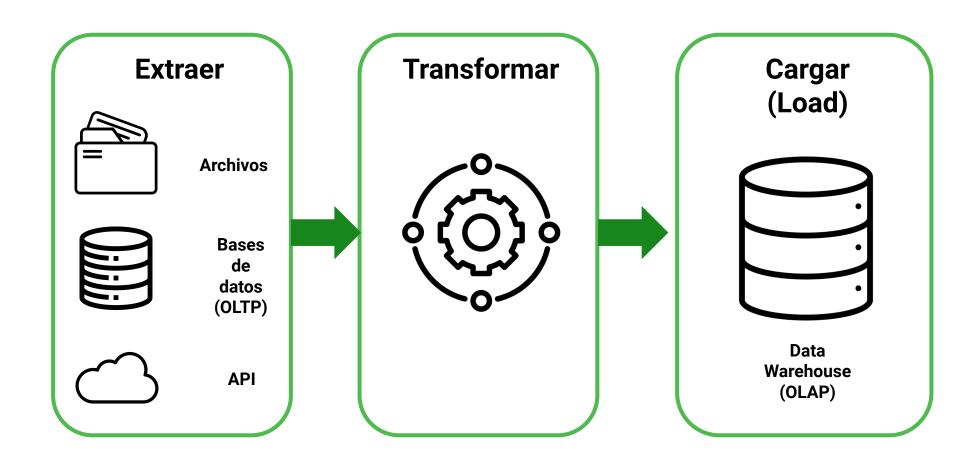


Día a día de una Data Engineer

• Crear automatizaciones para ETL.



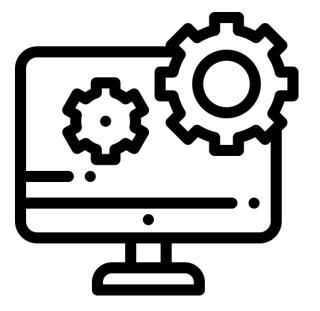






Roles relacionados

- Data Architect
- Big Data Architect





Herramientas y tecnologías para Data Engineers





Herramientas y tecnologías

 Programación con Python y bases de ingeniería de software.



- Linux.
- Automatización y scripting.



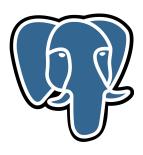


Herramientas y tecnologías

- Jupyter Notebooks y editores de código
- Manejo avanzado de bases de datos SQL y NoSQL.
- Pandas, Dask y Apache Spark.











Herramientas y tecnologías

- Airflow.
- Tecnologías cloud.
- Contenedores Docker.
- Orquestadores
 Kubernetes.







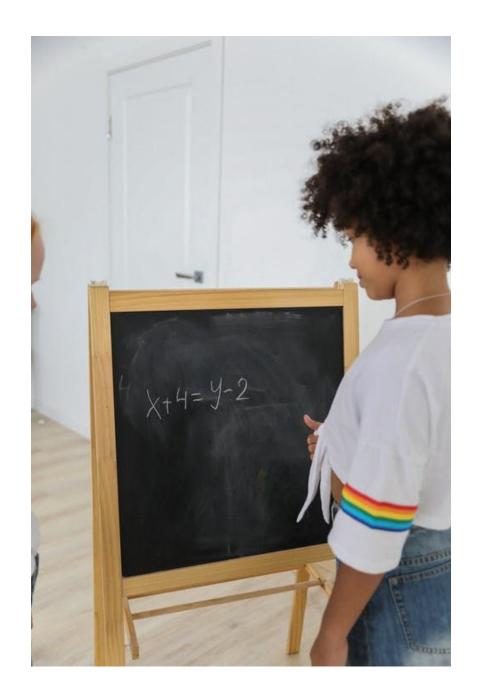






Matemáticas para Data Engineer

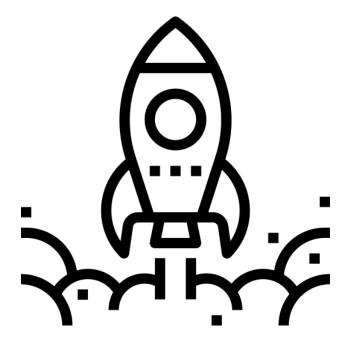
 Estadística descriptiva





¿Cómo empezar?

- Python y bases sólidas de ingeniería de software.
- Automatización y scripting.
- Pandas y Apache Spark.
- Bases de datos SQL y NoSQL.

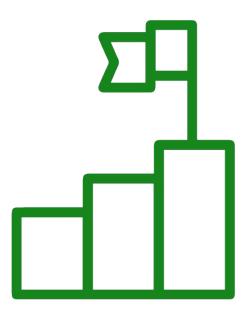




¿Cómo empezar?







Busca en sitios de trabajo vacantes de Data Engineers