

DASK



¿QUÉ ES DASK?



"Dask es una biblioteca de código abierto diseñada para proporcionar paralelismo a la pila de Python existente. Dask proporciona integraciones con bibliotecas de Python como NumPy Arrays, Pandas DataFrames y scikit-learn para permitir la ejecución paralela en varios núcleos, procesadores y computadoras, sin tener que aprender nuevas bibliotecas o lenguajes."

¿POR QUÉ DASK?



""Python ha crecido hasta convertirse en el lenguaje dominante tanto en el análisis de datos como en la programación general. Este crecimiento ha sido impulsado por bibliotecas computacionales como NumPy, pandas y scikit-learn. Sin embargo, estos paquetes no fueron diseñados para escalar más allá de una sola máquina. Dask se desarrolló para escalar de forma nativa estos paquetes y el ecosistema circundante a máquinas de varios núcleos y clústeres distribuidos cuando los conjuntos de datos exceden la memoria."

Características



Se reduce a computadoras individuales

Permite a los usuarios manipular conjuntos de datos de más de 100 GB en una computadora portátil o conjuntos de datos de más de 1 TB en una estación de trabajo

Tiene una API de Python

Se integra de forma nativa con el código de Python para garantizar la coherencia y minimizar la fricción con el programador

Soporta aceleración de GPU

Se integra con otros marcos para análisis de datos y aprendizaje de máquina acelerados por GPU

Admite aplicaciones complejas

Habilita algoritmos avanzados para estadística o aprendizaje de máquina, series temporales u operaciones locales, aplicaciones de paralelismo

Ventajas de usar dask



Fácil de usar

No requiere configuración o
instalación

Es seguro

Admite cifrado y autenticación
mediante certificados TLS/SSL

Es eficiente

Opera con baja sobrecarga y baja
latencia para un alto rendimiento

Es personalizable

Se puede utilizar para crear su propio
sistema informático paralelo con
lógica empresarial personalizada

Integrados con Dask

Scikit Learn



NumPy



PyTorch

