



POLARS VS. PANDAS

Presentación final

Alejandra Arredondo
Eduardo Santoyo
Sebastián Córdoba

Polars es una librería de Python que se utiliza para el análisis de datos, está optimizada para trabajar con datos de gran tamaño y puede ser muy rápida ya que utiliza todos los núcleos disponibles en la máquina.

Polars está diseñado para paralelizar tareas, en contraste con Pandas, una librería que utiliza otras herramientas que tratan de replicar el proceso de paralelización, como Dask.

Utiliza Rust en su núcleo, lo que permite una mayor eficiencia en términos de velocidad y uso de memoria.

PRINCIPALES DIFERENCIAS ENTRE POLARS Y PANDAS

Rendimiento: Polars es considerablemente más rápido que Pandas para operaciones de datos a gran escala debido a que Polars está diseñado para funcionar de manera óptima en sistemas distribuidos, utilizando múltiples núcleos de CPU y aceleración de hardware, como el uso de GPU. En contraste, Pandas funciona bien con pequeñas y medianas cantidades de datos, pero puede ser menos eficiente en grandes conjuntos de datos.

Funciones: Pandas tiene una gran cantidad de funciones y métodos incorporados que son útiles para una variedad de tareas de análisis de datos, como la limpieza y el filtrado de datos, la transformación de datos y el análisis estadístico. Polars, por otro lado, tiene un conjunto más limitado de funciones, pero se centra en operaciones de datos de alto rendimiento, como el procesamiento en paralelo y la computación distribuida.

Tipos de datos: Polars tiene una estructura de datos más limitada en comparación con Pandas, ya que solo admite tipos de datos numéricos y de cadena. Por otro lado, Pandas admite una amplia variedad de tipos de datos, incluidos objetos de fecha y hora, categorías y datos de tipo booleano.

Compatibilidad: Pandas es compatible con una variedad de formatos de archivo de datos, incluidos CSV, Excel y SQL, lo que lo hace ideal para trabajos de análisis de datos que requieren integración con diferentes fuentes de datos.

Polars, por otro lado, solo es compatible con formatos de archivo de datos específicos, como Arrow y CSV.

Polars utiliza estructuras de datos de alto rendimiento como los DataFrames y las Series, que son similares a los que se encuentran en Pandas.

Los DataFrames en Polars son objetos bidimensionales que contienen datos en filas y columnas, y se pueden pensar en ellos como una tabla.

Una de las principales diferencias entre Polars y Pandas es que los DataFrames en Polars son inmutables, lo que significa que no se pueden modificar después de su creación. Se pueden crear nuevos DataFrames a partir de los originales con transformaciones específicas, como filtrado, agregación y fusiones.

Las Series en Polars son objetos unidimensionales que representan una sola columna de un DataFrame. Al igual que los DataFrames, las Series son inmutables y se pueden crear a partir de diversas fuentes de datos.

Polars	Pandas
<p>Más adecuado para el procesamiento y análisis de grandes conjuntos de datos y aplicaciones que requieren una alta eficiencia de procesamiento.</p>	<p>Más adecuado para trabajos de análisis de datos que requieren una amplia variedad de tipos de datos y compatibilidad con diferentes fuentes de datos.</p>

VENTAJAS

- Tiempo de ejecución, en la mayoría de los casos la velocidad de procesamiento de Polars es más rápida que la de Pandas por mucho.

DESVENTAJAS

- Una sintaxis más larga que la de Pandas.
- Polars no tiene tantas funcionalidades para la exploración/análisis de datos como las tiene Pandas.