Alternatives to Python Pandas

When to Use Them

Polars

DASK

Dask natively scales Python

DuckDB

Where Are We Going?

➢Why Use pandas?

➢Why Not Use pandas?

➢Alterative 1: Polars to the Rescue.

➢Alternative 2: DuckDB Makes a Splash.

➢Alternative 3: Dask – to infinity & beyond.

➢Wrap Up

Why Use Pandas?

➢ Ubiquitous. Everyone knows it. Its the standard.

➢ Deep & Wide Objects and Methods.

➢ Pandas is like a Data Hub. See my Python/SQL Series

➢ Supports Creating Visuals.

➢ Pandas 2.0 ++ – New Features.

➢ YAL (Yet Another Library) - DevOps

Why Not Use Pandas?

➢Slow performance.

➢Dataframe must fit in your computer's memory.

➢Storage is Inefficient.

➢Clumsy API.

Alternative 1: Polars

➢ It's Fast.

➢ Very Fast.

➢ Very, Very Fast. Up to 50 X faster than pandas.

https://pola.rs/

Alternative 2: DuckDB

➤ It's Fast.

➤ It Can Scale Beyond Computer's Memory.

➤ Provides a Database Service & Full SQL Interface.

➤ Basically, DuckDB is an open-source Tabular Model.

https://duckdb.org/ ⬅

Alternative 3: Dask

➤ Competes with PySpark.

➤ Use Locally Or Scale Out When Needed.

➤ Easy to Learn & Use.

➤ Use pandas under the hood and compatible with pandas API.

https://www.dask.org/

Wrapping Up

➢Why Use pandas?

➢Why Not Use pandas?

➢Alterative 1: Polars to the Rescue.

➢Alternative 2: DuckDB Makes a Splash.

➢Alternative 3: Dask – to infinity & beyond.

Thank You!