# DuckDB

## What it is:

It is an Embedded, OLAP database.

Why use it:

1. Like mentioned before it is an embedded OLAP. No extra head for set up like in the case of SQL, Postgre etc.
2. It can handle very large files amounting to 6GB or more in fraction of seconds.
3. While Polars too perform at similar speed, its performance diminishes as the file size increases beyond 1GB.
4. When loaded a file size of 60% of RAM, both pandas and polars crash.
5. DuckDB performs well, performing operations such as groupby, filter at a very good speed.
6. One outstanding feature of DuckDB is concurrency. Multiple reads and writes are possible in duckdb. The same is not possible in case of pandas.
7. Can handle parquet files too, so can be employed in Apache Spark.
8. Integrates well with DBT.

# Conclusion:

DuckDB cannot be missed out in the Data Engineering tech stack.