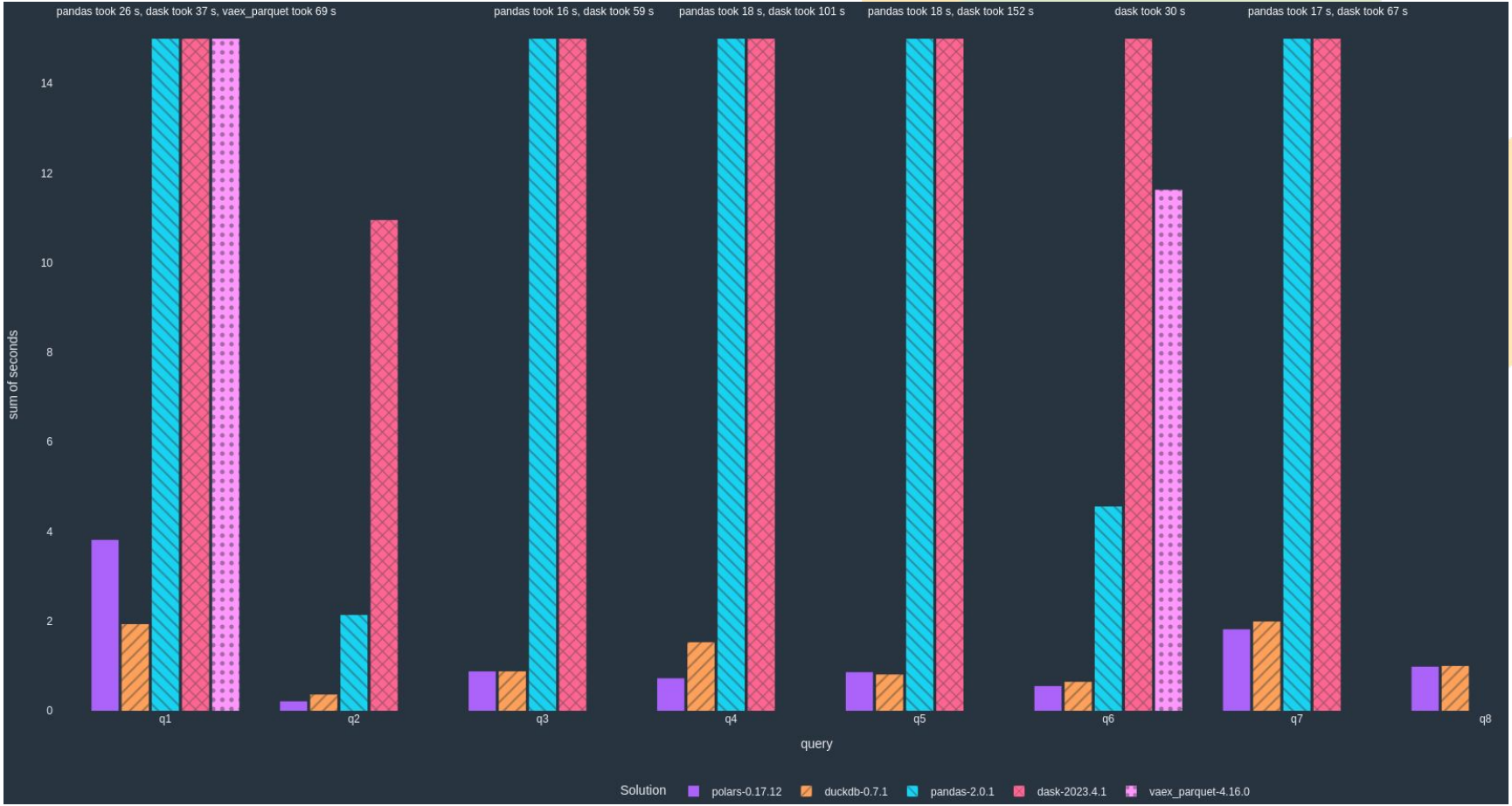
Polars vs Pandas

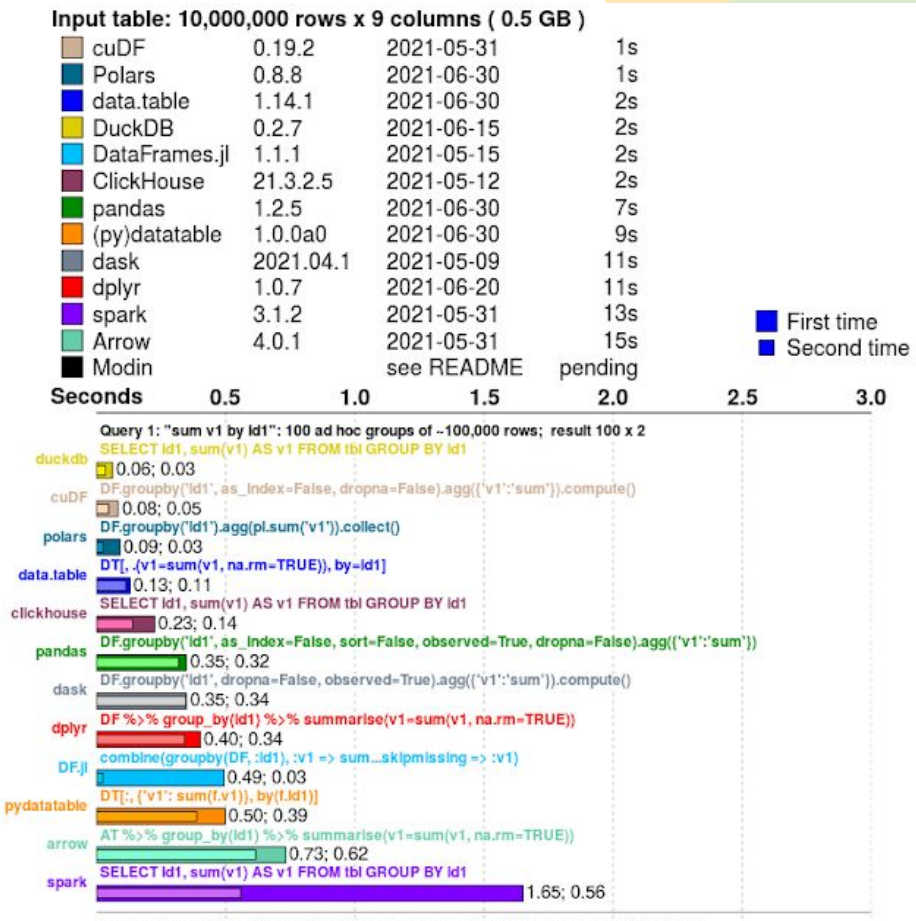
by Minh Huu Nguyen

Polars is fast

TPCH Benchmark



Database-like ops benchmark



Why Polars is fast?

Parallelization

Polars uses **parallelization** and **cache efficient algorithms** to speed up.

- Reduce redundant copies
- Traverse memory cache efficiently
- Minimize contention in parallelism

Lazy execution >< Eager execution

- It means that an expression is evaluated not immediately, but **only when needed.**
- Because of Lazy execution, Polars can perform optimization.
  - Run what is needed
  - Ignore what is not required
- Lazy execution is technique used in Big data framework such as Apache Spark.

Rust

Polars is implemented with Rust,

and Rust is much better than Python at implementing concurrency.

But, why Python is bad in concurrency?

Because of **Python Global Interpreter Lock (GIL)**

(But, this part doesn't belong to the seminar today!!)

Let’s demo...


PIXTA
VIETNAM