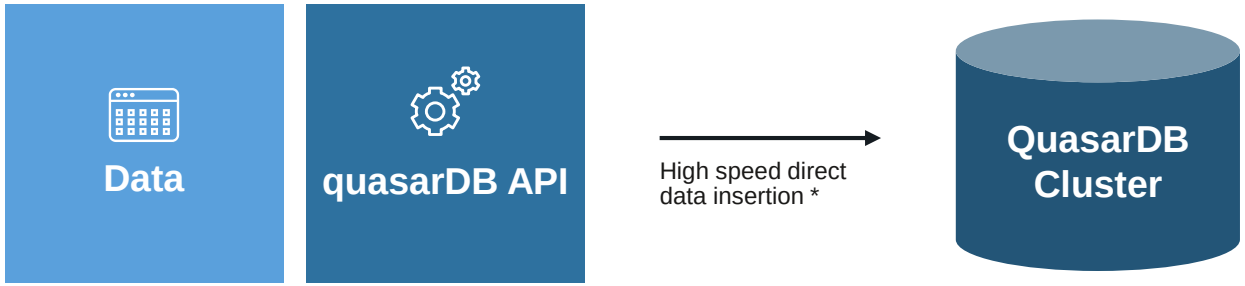
QuasarDB

Run your code in the database

User Defined Functions in Python

The Problem

QuasarDB data insertion as of 2.x



Data is:

- Sliced
- Distributed
- Compressed
- Indexed
- Persisted to disk

… before insertion acknowledgement

*** e.g., tick-by-tick market data stream captured at up to 100M+ updates/second**

Data access: QuasarDB query

- Queries are translated to low level API calls
- Users are unconcerned with data location, or if it's on disk or in memory

e.g. select min(bid) from my_stocks in range(2001, +1d)

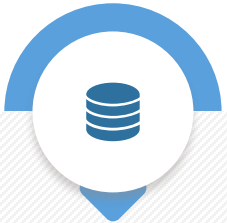
QuasarDB computes the minimum and returns only the result.

Queries run on the nodes containing the requested data, and QuasarDB transparently loads data from disk as needed.

Benefit: slow data exchanges are minimized

Benefit: analysts can work naturally on the whole history

Advanced queries



Use the built-in functions from QuasarDB?

There will always be a mathematical function you need we didn’t implement

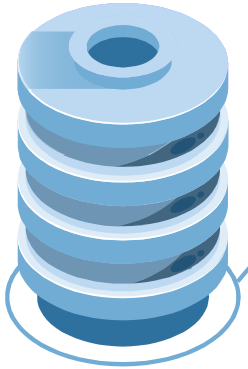
Get the data from QuasarDB and run it in your Python?

Not sensible if the data transfer time is much greater than the computation time

Run your own function inside QuasarDB?

But how?

UDF in QuasarDB 3.1



Python embedded in the database

Write your code in Python

Push it as an UDF

Call the UDF from your query

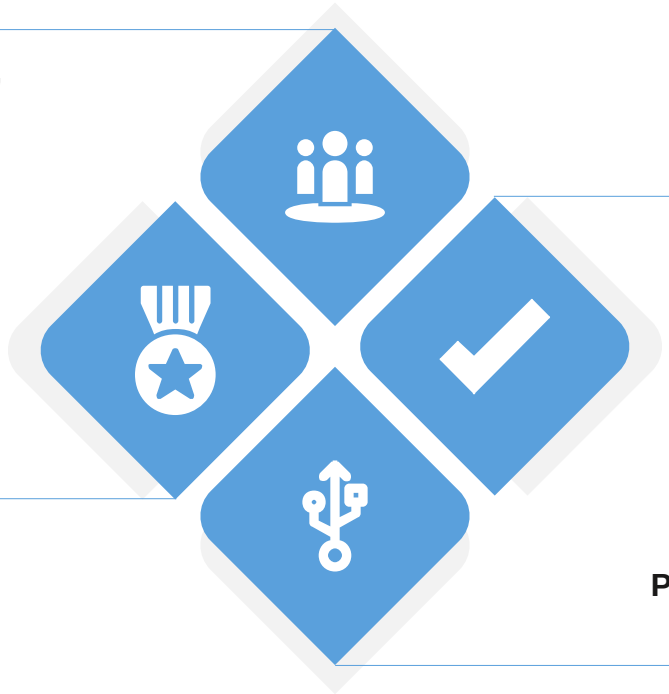
The server will run your function on the data

Example

Select udf_my_sum(bid) from my_stocks in range(2001, +1d)

Why Python?

Universally used: transferable skill, no advanced training required

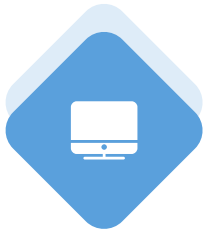


The license allows it ☺

Satisfactory performance

Python can be embedded and called from C++

Benefits



Any kind of computation is now possible

Optimize data transfer

Greatly simplifies workflows

In thefuture:

Allow numpy and Pandas

ENROLL IN THE BETA PROGRAM!

HTTPS://INFO.QUASARDB.NET/BETA

QuasarDB

THANK YOU

www.quasardb.net

USA

222 BROADWAY – 19TH FLOOR
NEW YORK
NY 10038

UK

40 BANK STREET
CANARY WHARF
LONDON E14 5NR

FRANCE

24, RUE FEYDEAU
75002 PARIS