

Shubham Diwakar - Querying Apache Parquet files with PromQL

May 25, 2023

- Thanos mentees meetup
 - Prem: will talk with Giedrius about organising
- Querier: help with some issues
- Retrospective next sync

May 11, 2023

- Inside the FrostDB repo:
 - `go run cmd/parquet-tool/main.go <parquet-file>`

Apr 28, 2023

- Updates
 - Do single append for the entire TSDB block. Since we work with a small block, this should not cause memory problems for the time being.
 - Use `frostdb/cmd/parquet-tool` to visualize the created parquet file.
 - You can try using `pip install parquet-cli`

Apr 20, 2023

- Updates
 - Slow conversion time can be solved somewhat with batching before inserts.
 - PR with batched ingestion in epimetheus:
<https://github.com/polarsignals/epimetheus/pull/86/files>
 - Reference for enabling persistent storage with FrostDB
<https://github.com/polarsignals/epimetheus/blob/main/frostdb/frostdb.go#L63>
 -
- Next
 - Persistence
 - Querying

Apr 6, 2023

- Updates
 - Time series data is now written in FrostDB:
https://github.com/Shubham4359/TSDB_Parquet
- Next steps
 - Add short readme
 - Persistence
 - Querying
 - Slow conversion time
 - Filip Petkovski Look into why appending to FrostDB is slow

Mar 30, 2023

- Updates
 - Managed to get a prototype with static schema
- Next steps
 - Investigate using FrostDB for a dynamic schema

Mar 23, 2023

- Updates
 - Converting from TSDB to Parquet is in progress
 - Serializing time as timestamp: <https://github.com/segmentio/parquet-go/pull/321>
 - Let's stick to int for now
 - Look at <https://github.com/polarsignals/frostdb> for some ideas/reference
 - Let's skip using this as we don't need dynamic columns yet
- This week:
 - Create empty repository
 - Raise a PR with the work in progress

Mar 16, 2023

- Updates
 - Shubham went through quite a bit of learning materials including the TSDB blogs from Ganesh and the PromQL talk
- Some questions, around the blogs/talks
- Plan for next week
 - Create a TSDB block. One way to do it is to start a Prometheus instance that scrapes itself and let it run for a few minutes. Alternatively use [thanos bench](#).
 - Take a look at [ploomtol analyze](#).

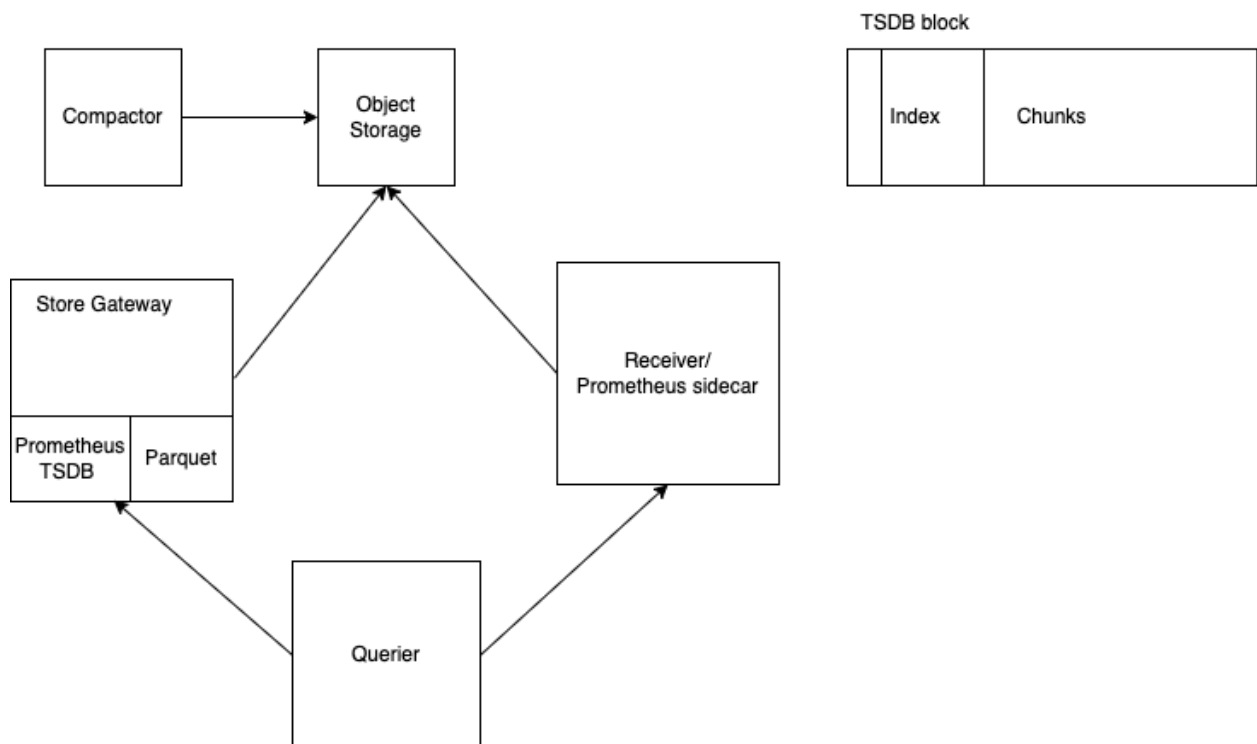
Use a Parquet library to convert TSDB to Apache Parquet. One example is:

<https://github.com/segmentio/parquet-go>

- One file per metric. Large metrics get broken down into multiple tables.
- Sorted by timestamp column.

07.03.2023

- [Mentorship checklist](#)
- [KubeCon Thanos talk](#)
- [Prometheus TSDB talk](#)
- [TSDB Block overview](#) blog
- [PromQL Engine talk](#)
- [Prometheus Tutorial](#)
- Prometheus concepts: [Official docs](#)
 - Timeseries
 - Metric
 - Sample
 - TSDB block



// Repeated Field