



Leveraging TTL in Presto Local Cache

Chunxu Tang, Jianjian Xie@ Alluxio



Chunxu Tang

Staff Research Scientist @ Alluxio
PrestoDB Committer



Jianjian Xie

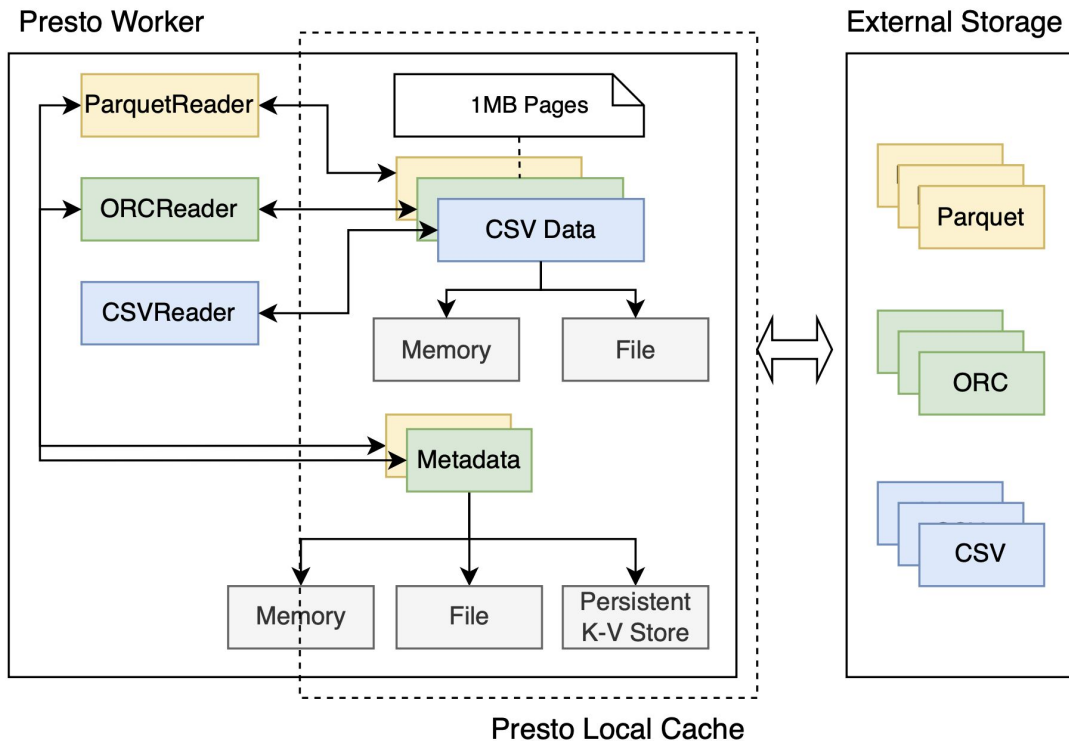
Staff Software Engineer @ Alluxio
PrestoDB Contributor

Presto Local Cache

- Reduce query latency
- Reduce the data transferred between Presto and remote storage

**Data Caching for Enterprise-Grade
Petabyte-Scale OLAP**, USENIX ATC 2024

<https://arxiv.org/pdf/2406.05962>



Motivation for TTL

- Meet data privacy requirements



Configurations & Metrics

- Background thread scanning local cached files

Properties

- `USER_CLIENT_CACHE_TTL_ENABLED = true`
- `USER_CLIENT_CACHE_TTL_CHECK_INTERVAL_SECONDS = 3600`
- `USER_CLIENT_CACHE_TTL_THRESHOLD_SECONDS`

Metrics

- `CLIENT_CACHE_PAGES_AGES`
- `CLIENT_CACHE_PAGES_INVALIDATED`

Use Case

Data Privacy

- GDPR regulation is 90 hours, set 60 hours as threshold and 1 hour as interval

Cache Freshness

- In dashboard scenario, set ttl for Cache Quick Removal