ALLUXIO

Trino Optimization with Distributed Caching on Data Lake

Hope Wang
Beinan Wang

Hope Wang

Developer Advocate @ Alluxio
Trino Contributor
PrestoDB Contributor

Dr. Beinan Wang

Senior Staff Engineer @ Alluxio
Trino Contributor
PrestoDB Committer

100,000,000,000,000,000,000,000

bytes of data will be stored in the cloud by 2025

Source: Cybersecurity Ventures

10%

of your data is hot data

Source: Alluxio

THE GAP

between Trino and the Data

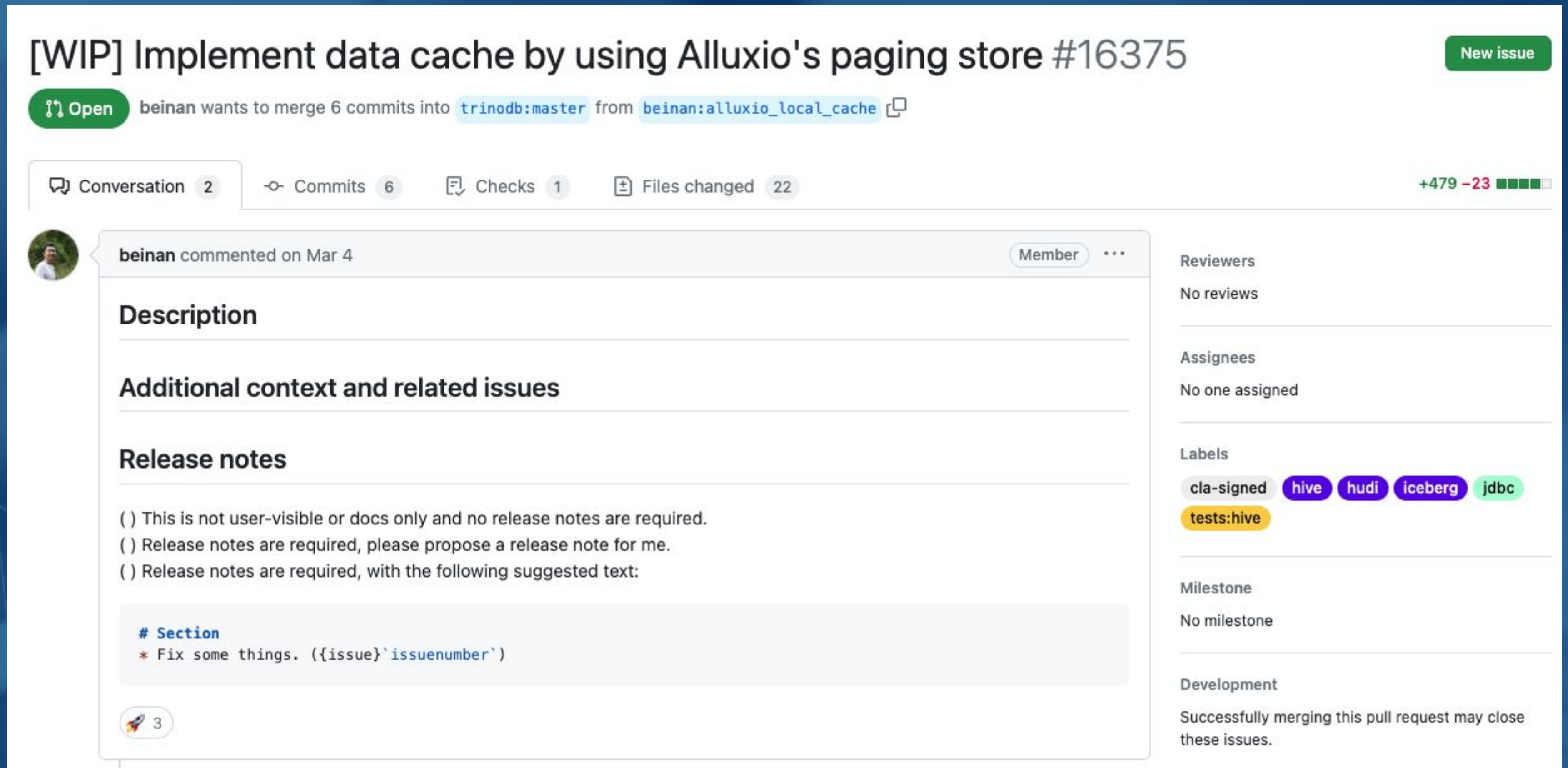
Data Caching Helps

Boost Performance

Prevent Network Congestion

Save Costs

Offload Under Storage



[WIP] Implement data cache by using Alluxio's paging store #16375

beinan wants to merge 6 commits into `trinodb:master` from `beinan:alluxio_local_cache`

beinan commented on Mar 4

Description

Additional context and related issues

Release notes

( ) This is not user-visible or docs only and no release notes are required.
( ) Release notes are required, please propose a release note for me.
( ) Release notes are required, with the following suggested text:

```
# Section
* Fix some things. ({issue}`issuenumber`)
```

Reviewers
No reviews

Assignees
No one assigned

Labels
cla-signed hive hudi iceberg jdbc tests:hive

Milestone
No milestone

Development
Successfully merging this pull request may close these issues.

Introducing Trino's Upcoming Data Caching

Key Features of Data Caching

Caching Data

Local SSD
Memory

Connector Support

Iceberg
Hudi
Delta Lake
Hive

Data Format Support

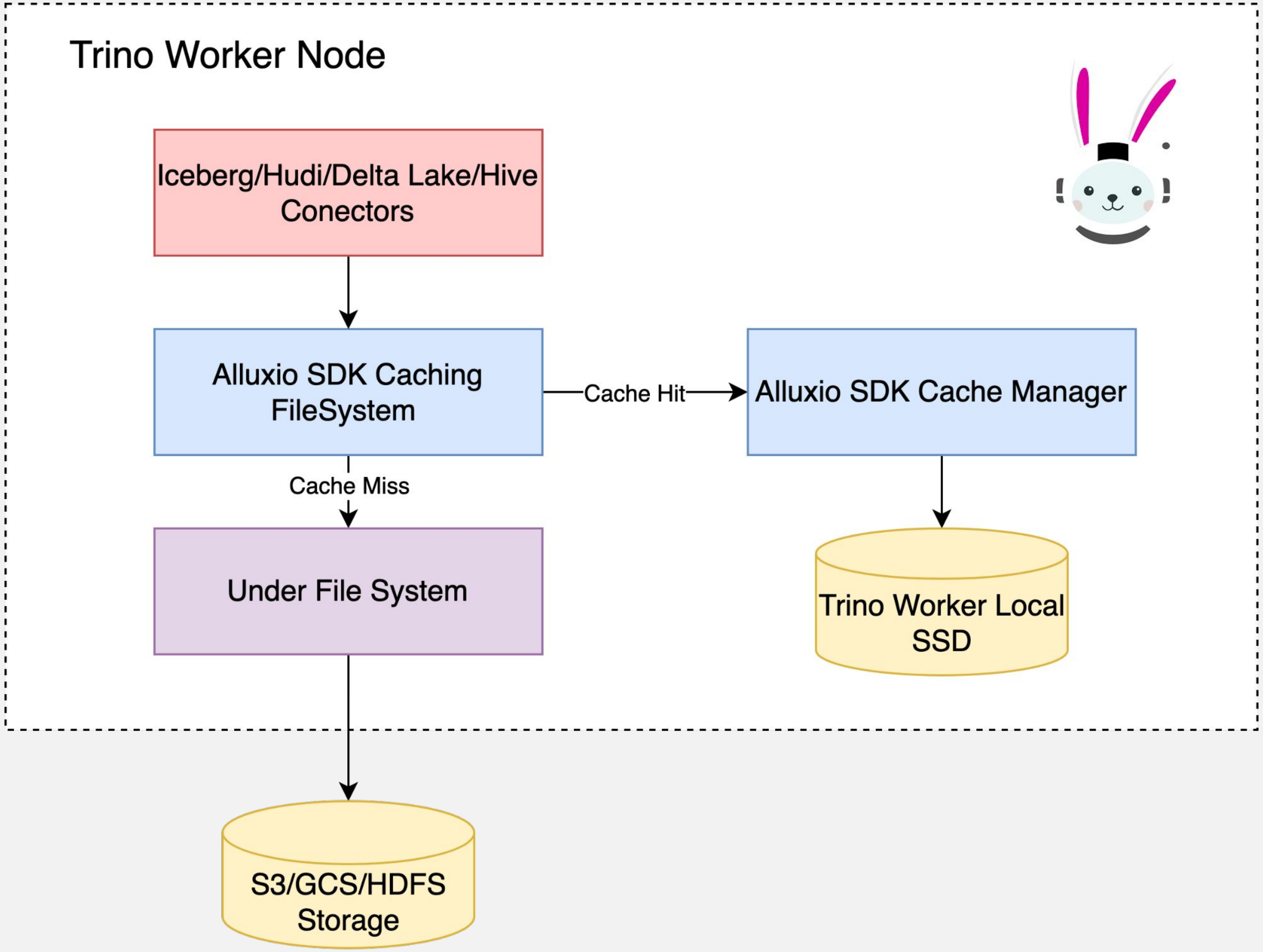
Parquet
ORC

How to Enable Data Caching?

From the view of a Trino user, nothing really changes

```
cache.enabled=true
cache.base-directory=/tmp/cache
cache.max-cache-size=10G
node-scheduler.cache-affinity-policy=SOFT
```

A Deeper Dive - How Does Data Caching Work?



Adopted By Leading Organizations

ByteDance

Shopee

Uber

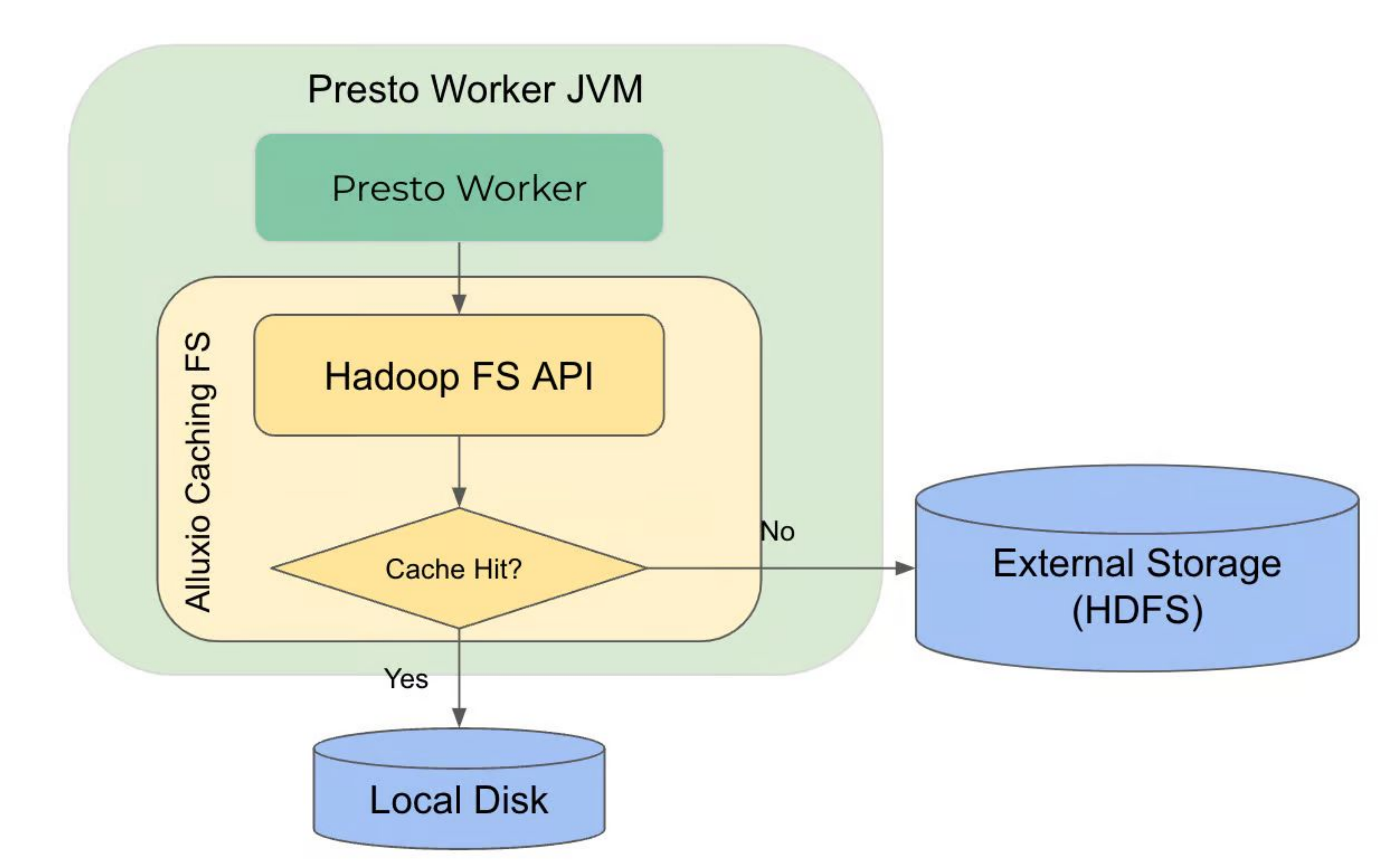
知乎
www.zhihu.com

NEWS BREAK

Tencent

Data Caching at Uber Scale

3 Clusters, 1500 Nodes



⬆ **50%** Input Read Performance

⬇ **10%** Data Read Traffic to HDFS

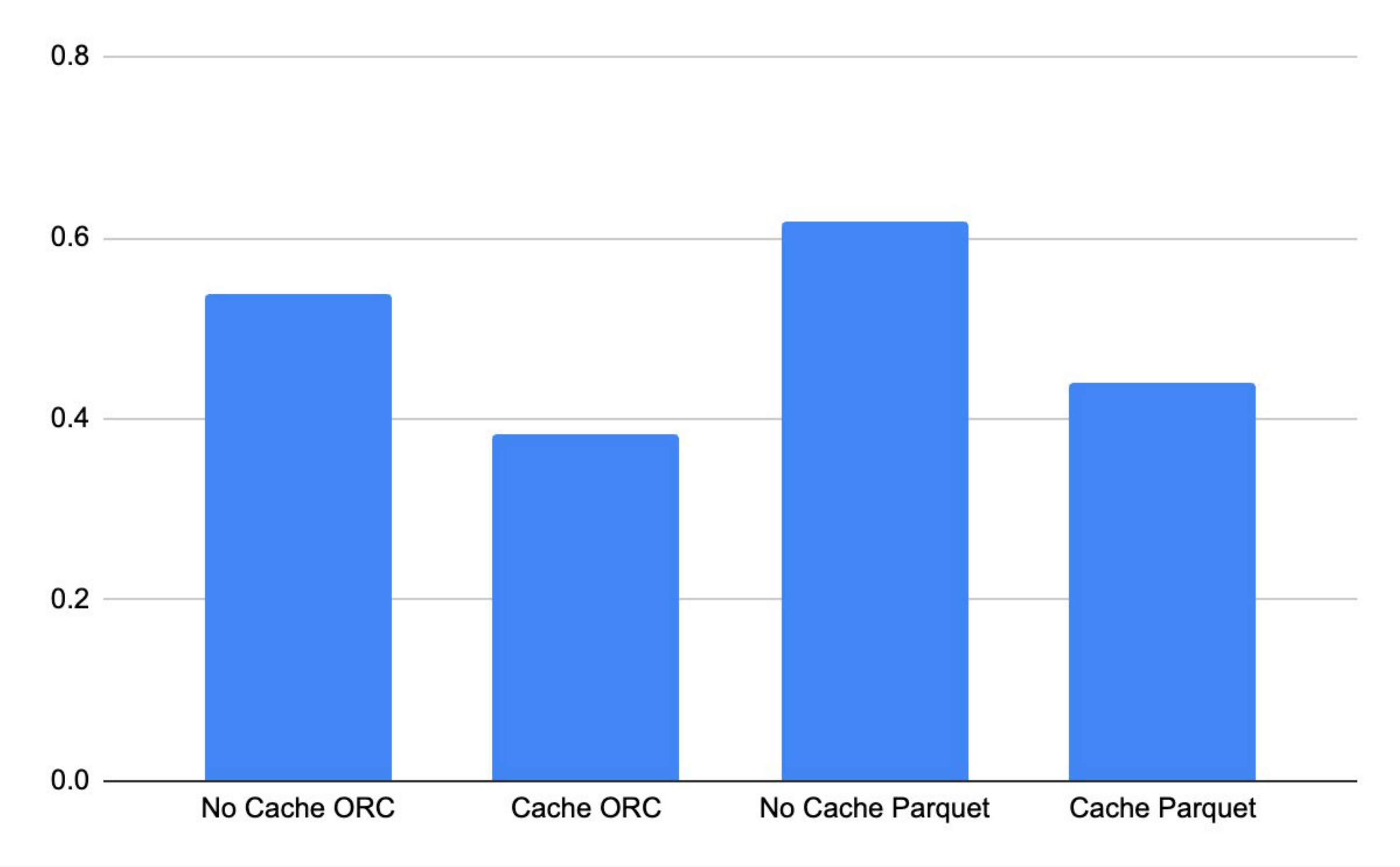
Source: https://www.uber.com/blog/speed-up-presto-with-alluxio-local-cache/

Shopee

Beta-tested Trino's Data Caching Functionality & Performance

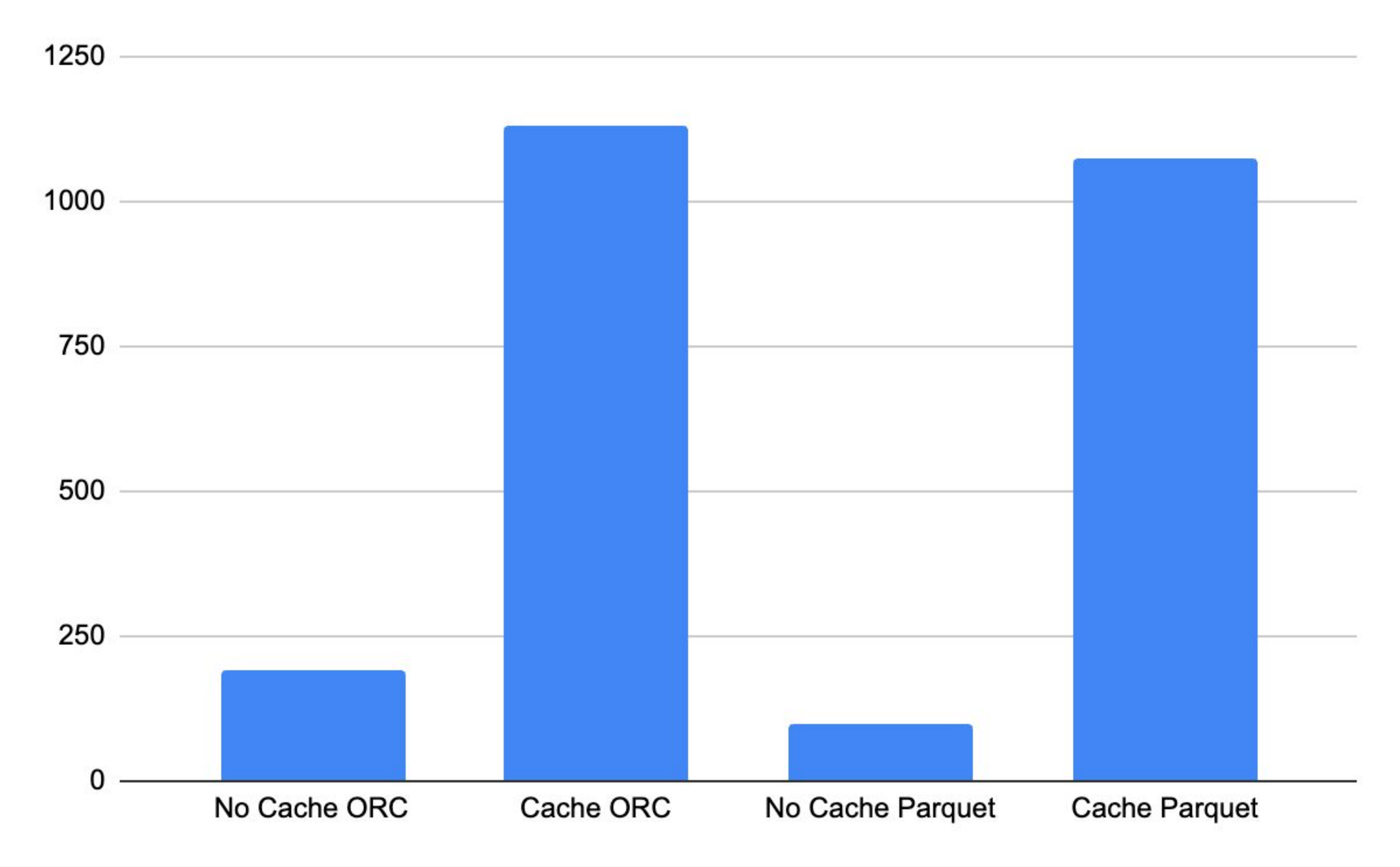
↓ 40%

Query Latency (Second)

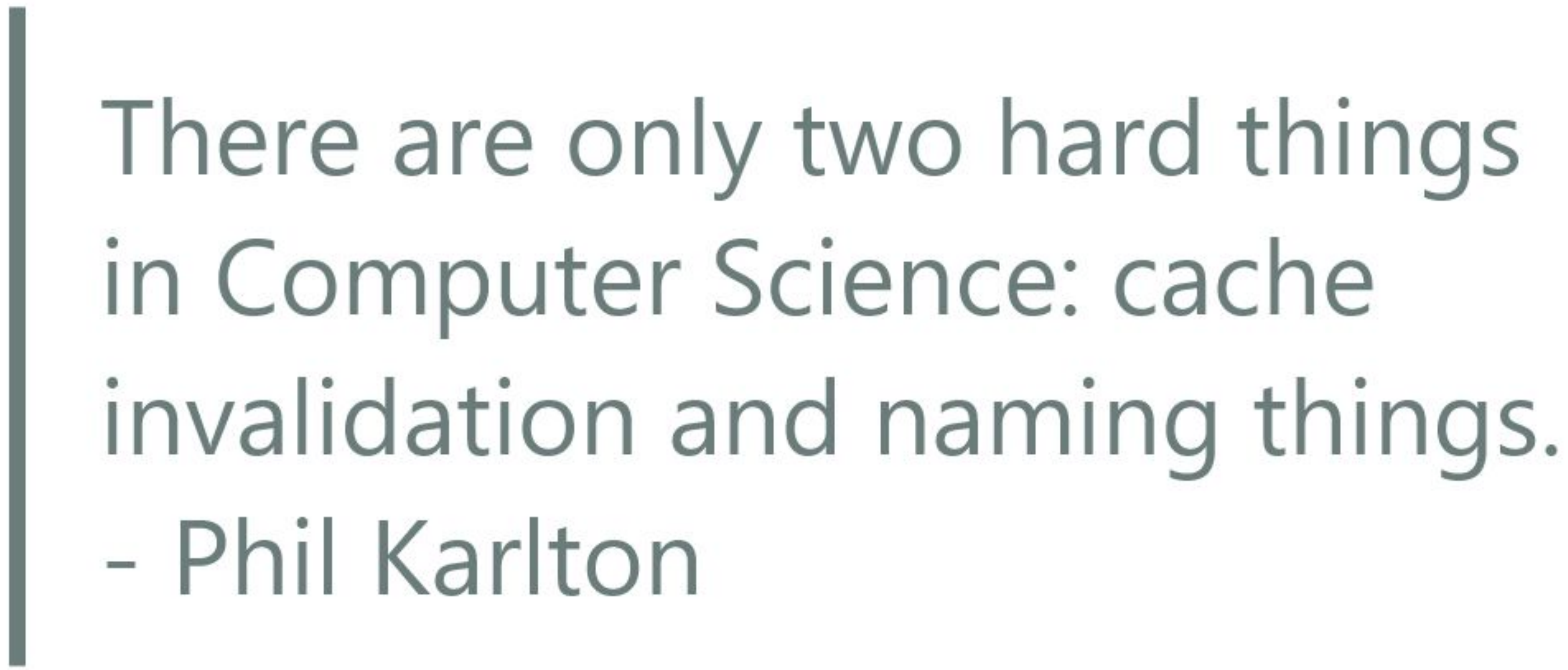


10x

IO throughput (MB)



Source: Shopee



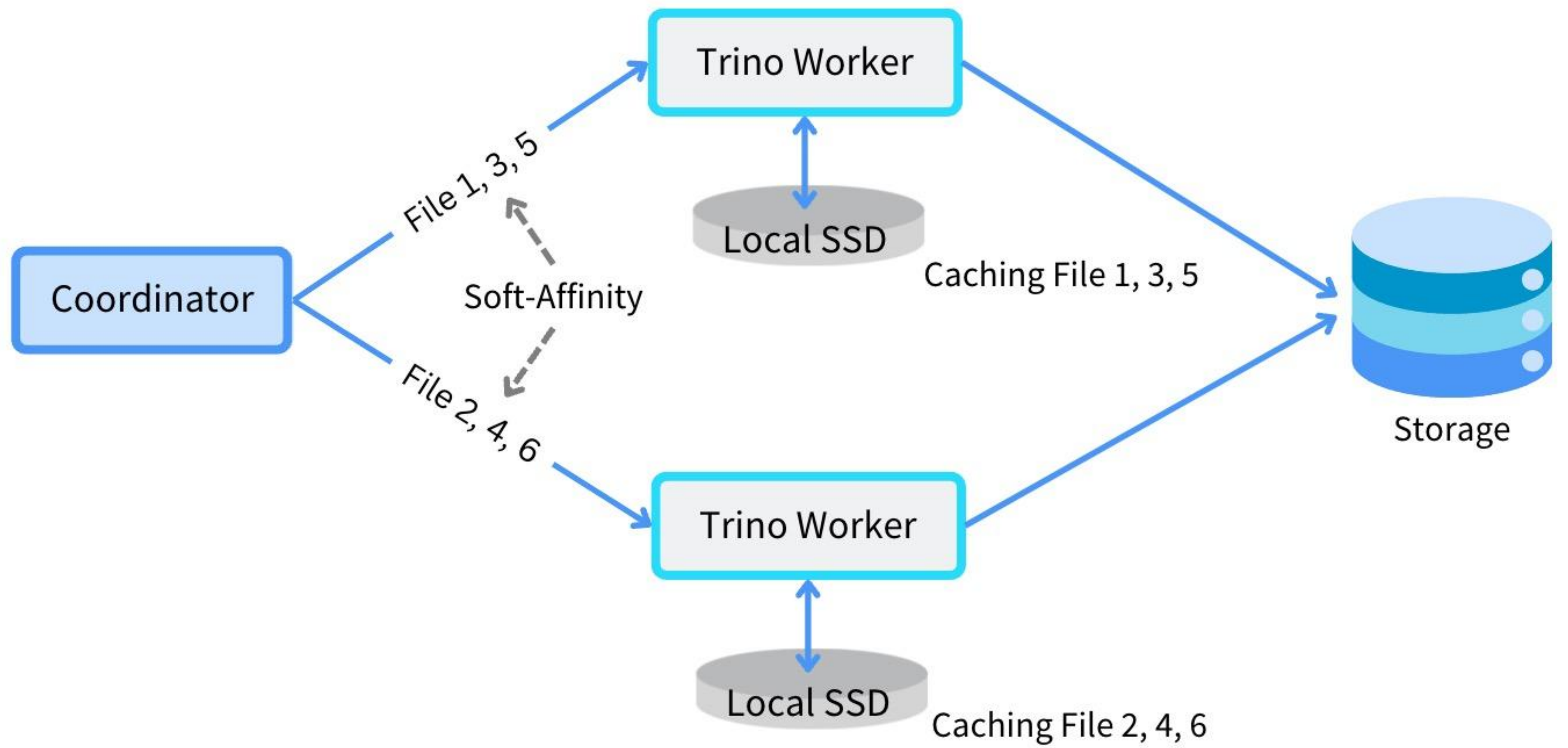
> There are only two hard things in Computer Science: cache invalidation and naming things.
> - Phil Karlton

Source: https://martinfowler.com/bliki/TwoHardThings.html

Technical Highlights

High Cache Hit Rate

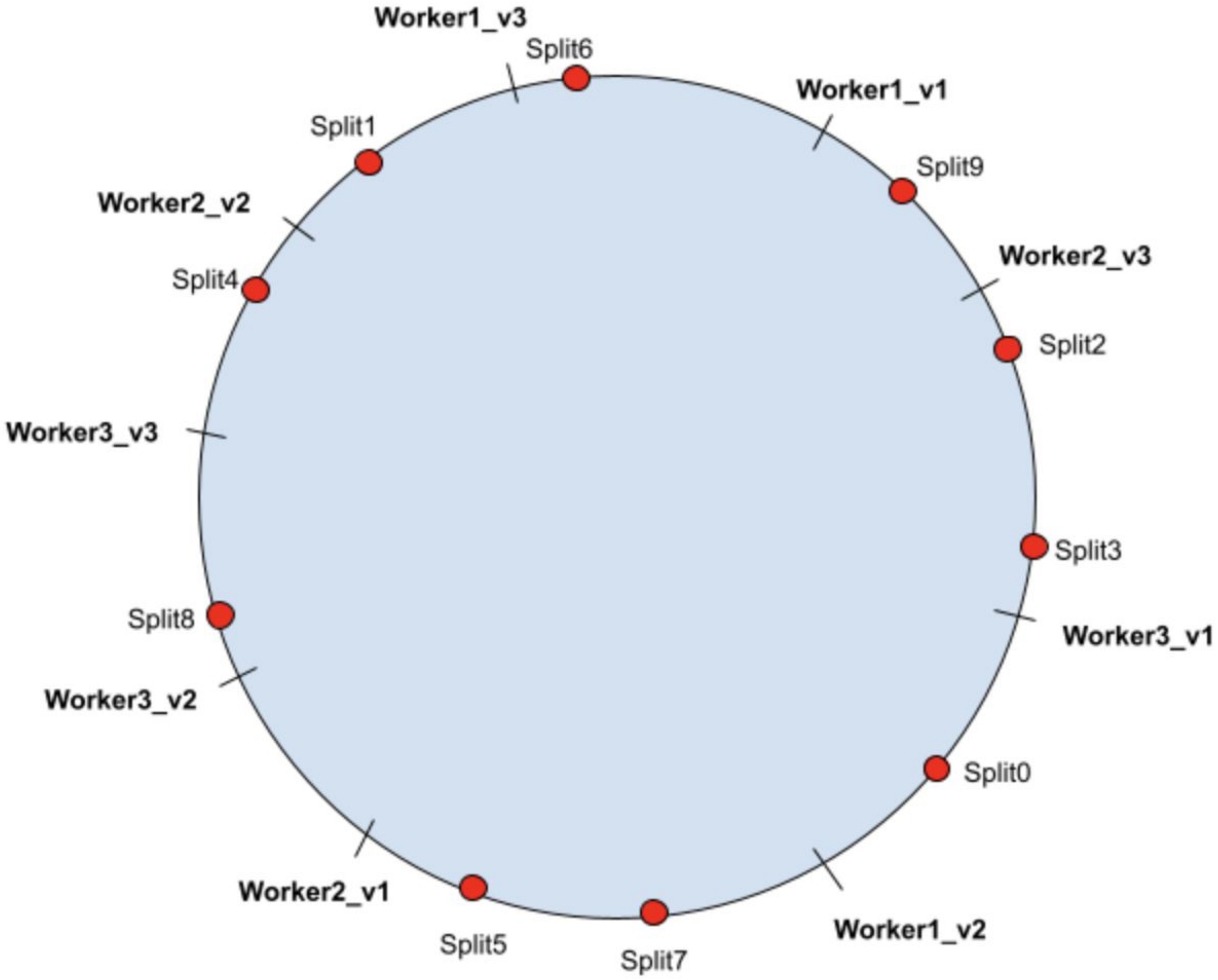
Soft-Affinity Scheduling Mechanism



Cluster Elasticity

Implement **Consistent Hashing**

- Minimize the number of split relocation when adding or removing workers

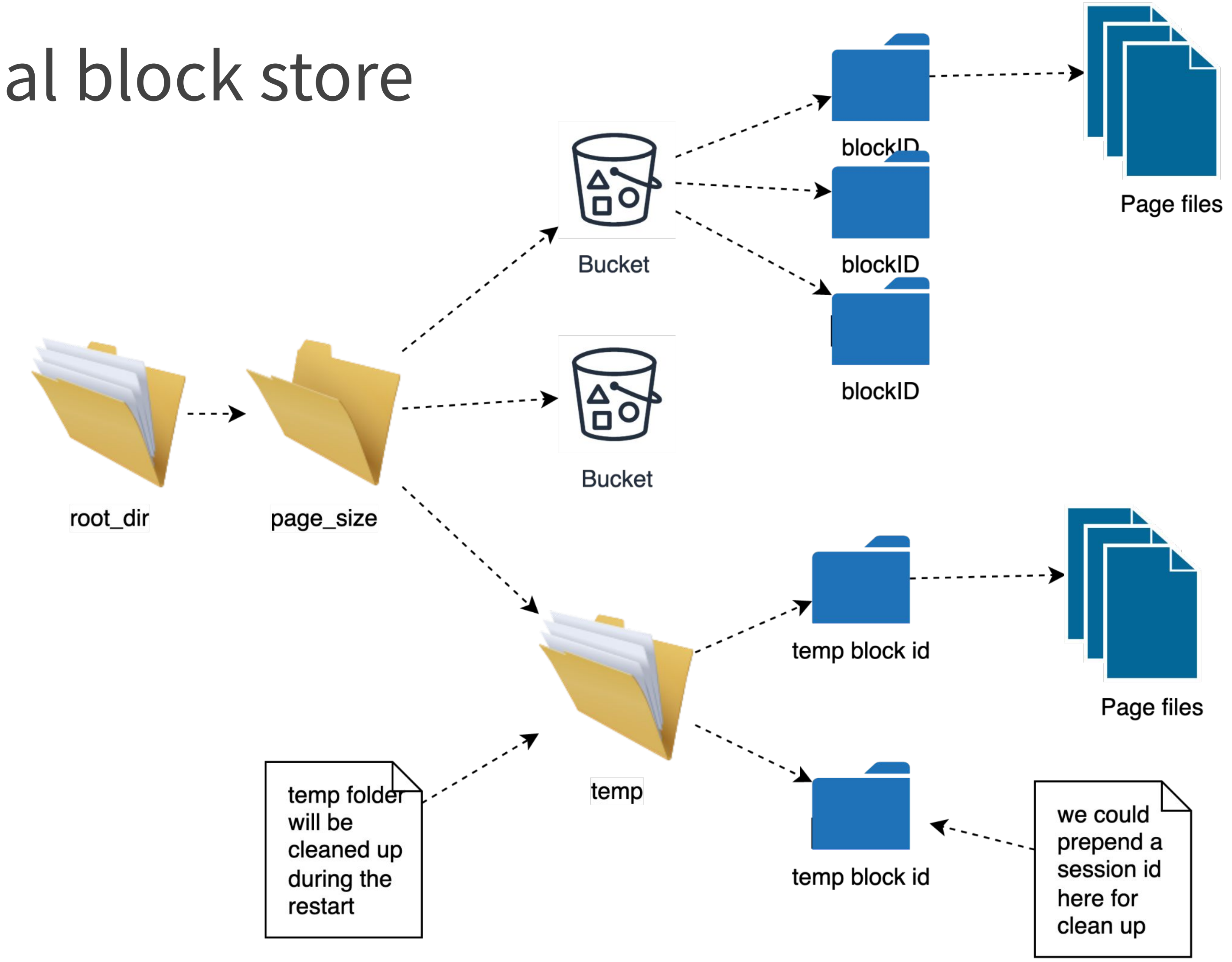


Source: https://www.alluxio.io/blog/using-consistent-hashing-in-presto-to-improve-caching-data-locality-in-dynamic-clusters/

Cache Storage Efficiency

Trino Data Cache is powered by **Alluxio's page store**

- Battle-tested in many tech giants
- Much less read amplification than the traditional block store
- Support LRU and FIFO cache eviction policy
- Support customized cache admission policy

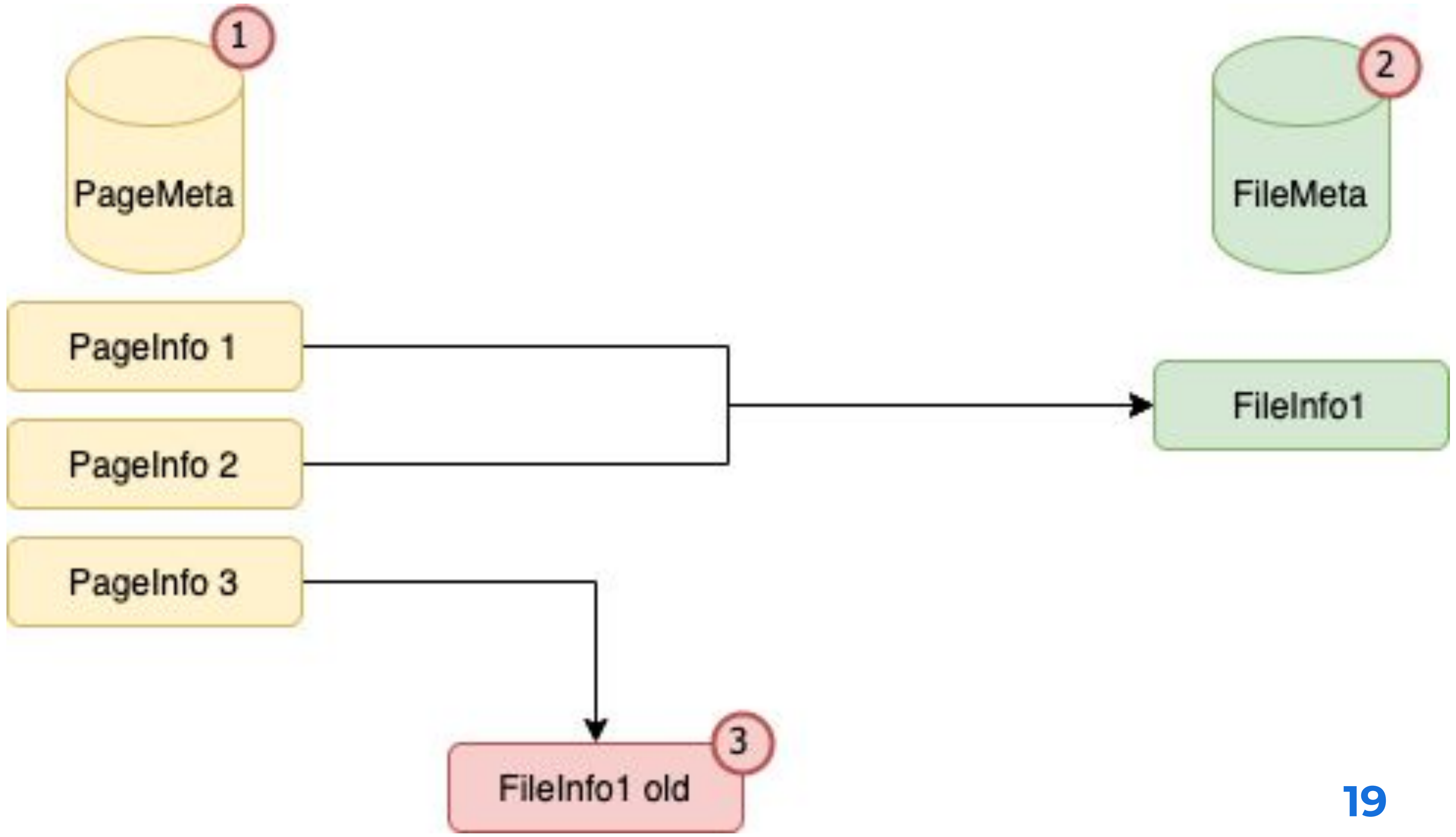


temp folder will be cleaned up during the restart

we could prepend a session id here for clean up

Data Consistency

- Get the lastModificationTime from File Status
- Generate the Cache Identifier
  - md5(file_path + lastModificationTime)
- The stale data will get evicted



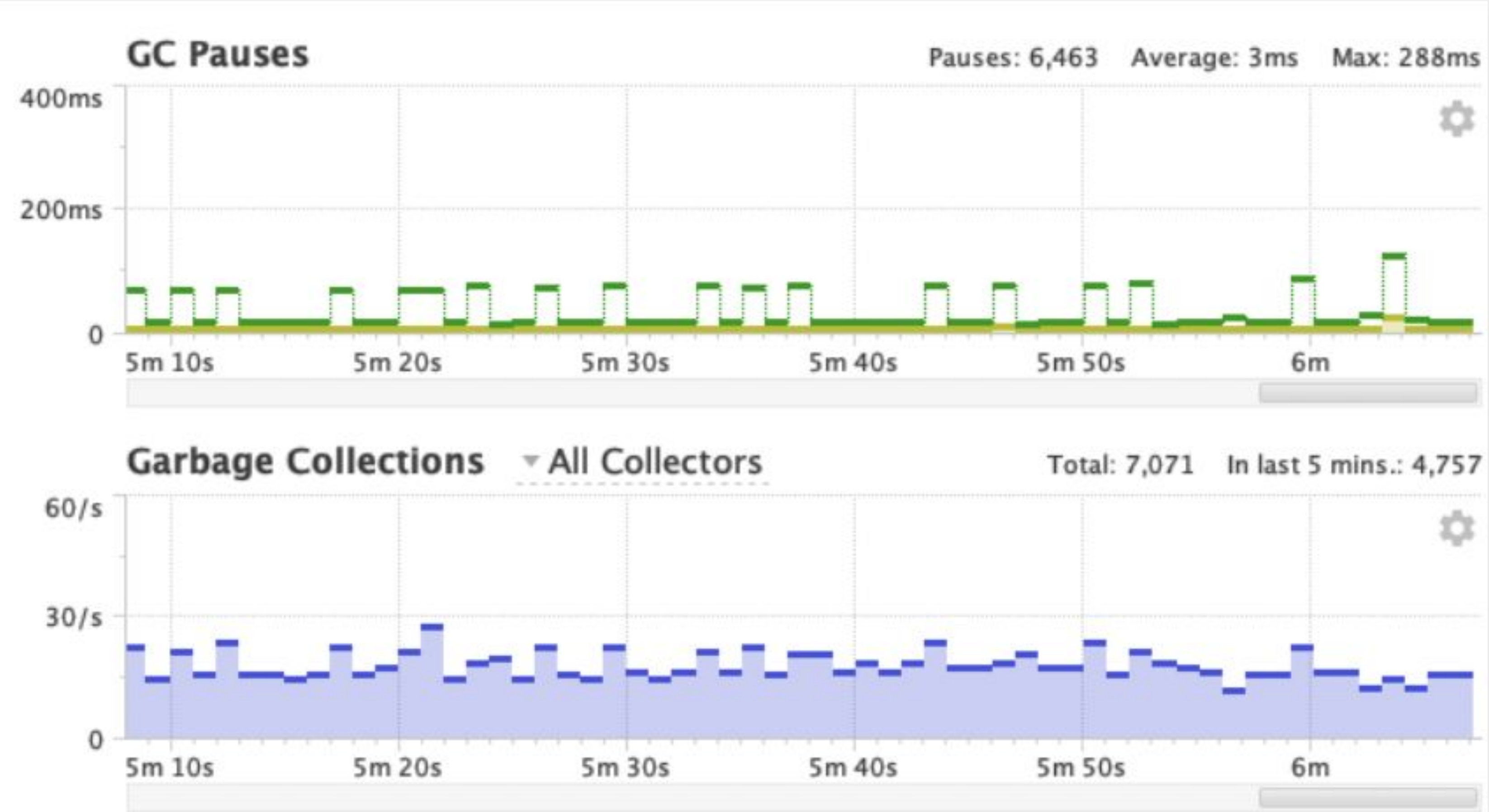
Ongoing Work

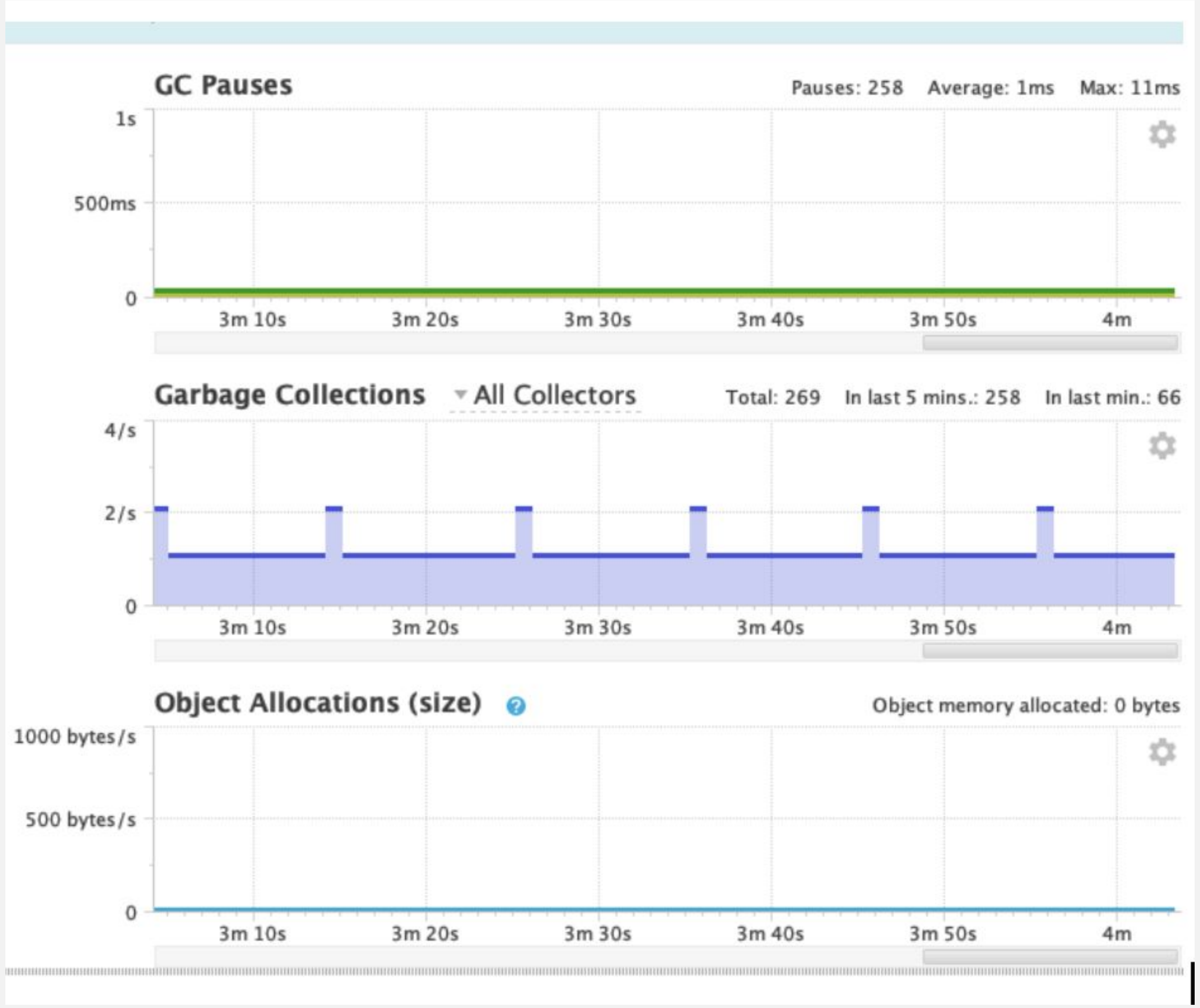
Semantic Cache

- Zero Read Amplification
- Performance Gain: 8% Less CPU Usage

Native/Off-heap Cache

- Much less GC pressure
- Less CPU Usage





Distributed Cache

- Scalability
- High Availability
- Performance
  - Better cache hit rate
  - Optimized for positioned read

Thank You

JOIN THE CONVERSATION
ON SLACK
ALLUXIO.IO/SLACK

www.alluxio.io

linkedin.com/alluxio

twitter.com/alluxio

slackin.alluxio.io