

OPEN POSSIBILITIES.

Throughput Improvements of Hyperscaled Distributed Databases based on Precision Timing



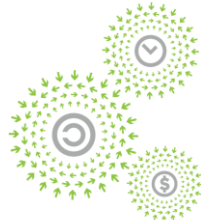
OCP
GLOBAL
SUMMIT

NOVEMBER 9-10, 2021

Throughput Improvements of Hyperscaled Distributed Databases based on Precision Timing

Georgi Chalakov, NVIDIA
Ahmad Byagowi, Facebook

OPEN POSSIBILITIES.



OPEN
PLATINUM™



Why Do We Need Synchronization in DCs?

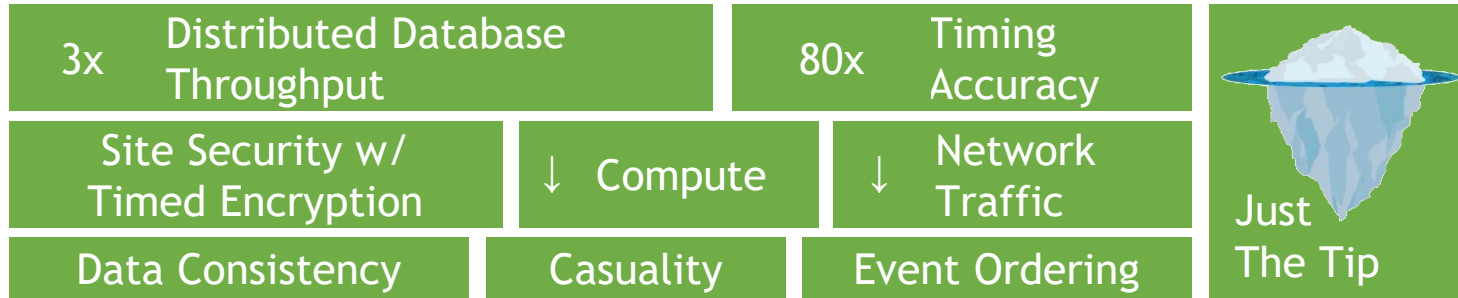


TIME
APPLIANCES

“Nanosecond-level clock synchronization enables a new spectrum of timing and delay-critical applications in data centers”

-- Google, Stanford, Exploiting a Natural Network Effect for Scalable, Fine-grained Clock Synchronization

A **Precise Time Axis** leaps applications' performance, efficiency and security



OPEN POSSIBILITIES.



External Consistency

- For any two transactions T1 and T2, if T2 starts commit after T1 finishes committing, then the timestamp for T2 is greater than the timestamp for T1



TIME
APPLIANCES

OPEN POSSIBILITIES.

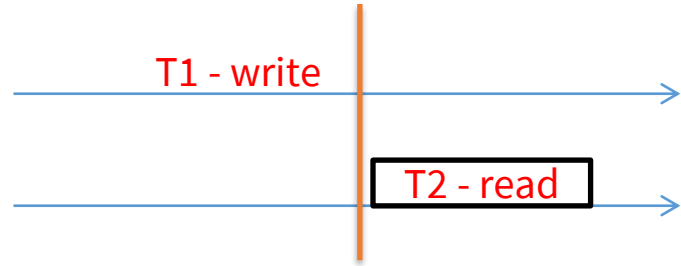
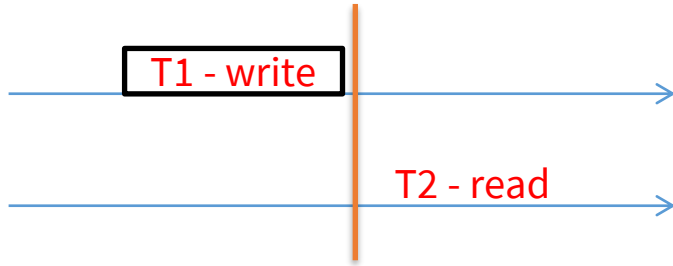


External Consistency



TIME
APPLIANCES

- Write waits proportionally to the clock uncertainty
- Read waits proportionally to the clock uncertainty.
- Or transaction restarts if external consistency is violated



OPEN POSSIBILITIES.



External Consistency

- Transaction restarts if external consistency is violated

*“For example, in an experimental CockroachDB cluster of 32 servers we read 128 records, each updated every 25 ms. We found that as the clock uncertainty T was reduced from **1 ms** to **10 us** and then to **100 ns**, the retry rate fell from **99.30%** to **4.74%** and to **0.08%** in an experiment with 10,000 reads for each value of T .”*

<https://www.usenix.org/system/files/conference/nsdi18/nsdi18-geng.pdf>



TIME
APPLIANCES

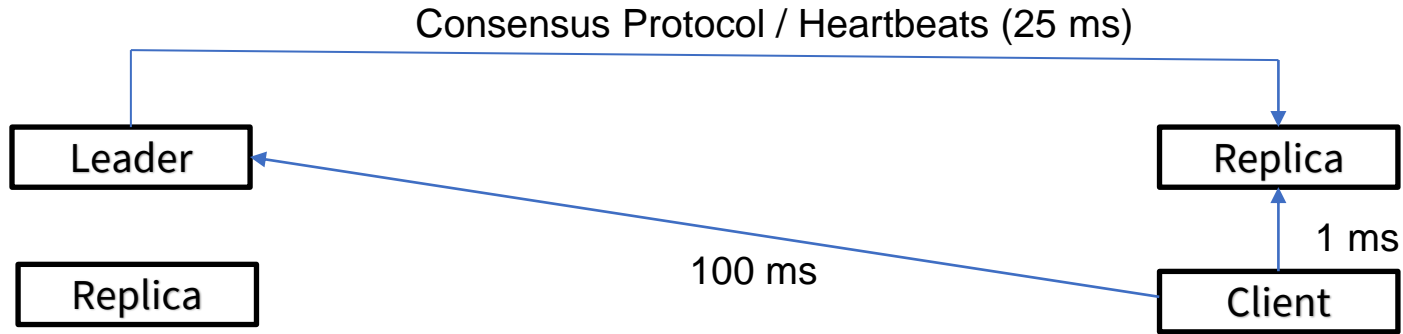
OPEN POSSIBILITIES.



Read from a Local Replica



TIME
APPLIANCES



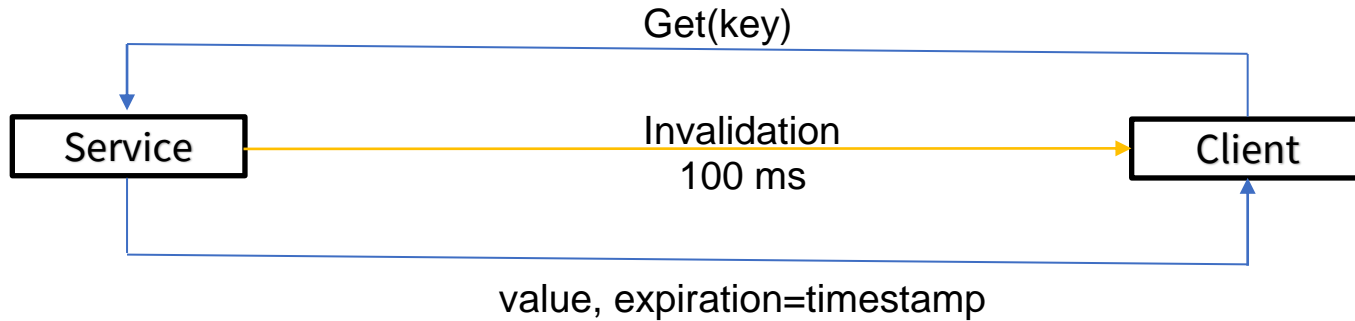
OPEN POSSIBILITIES.



Invalidate Cache Entry



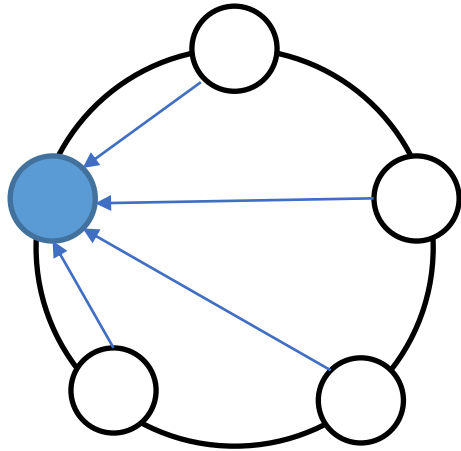
TIME
APPLIANCES



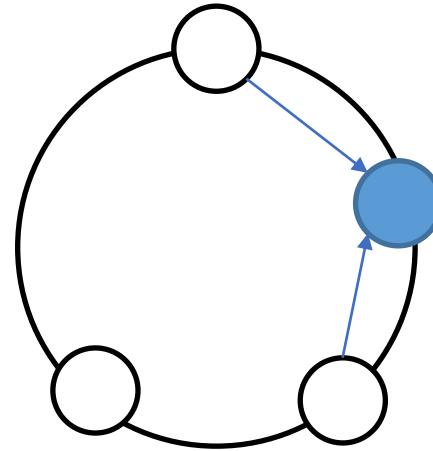
OPEN POSSIBILITIES.



Leader Election



New leader waits for the duration of the lease + clock uncertainty timespan

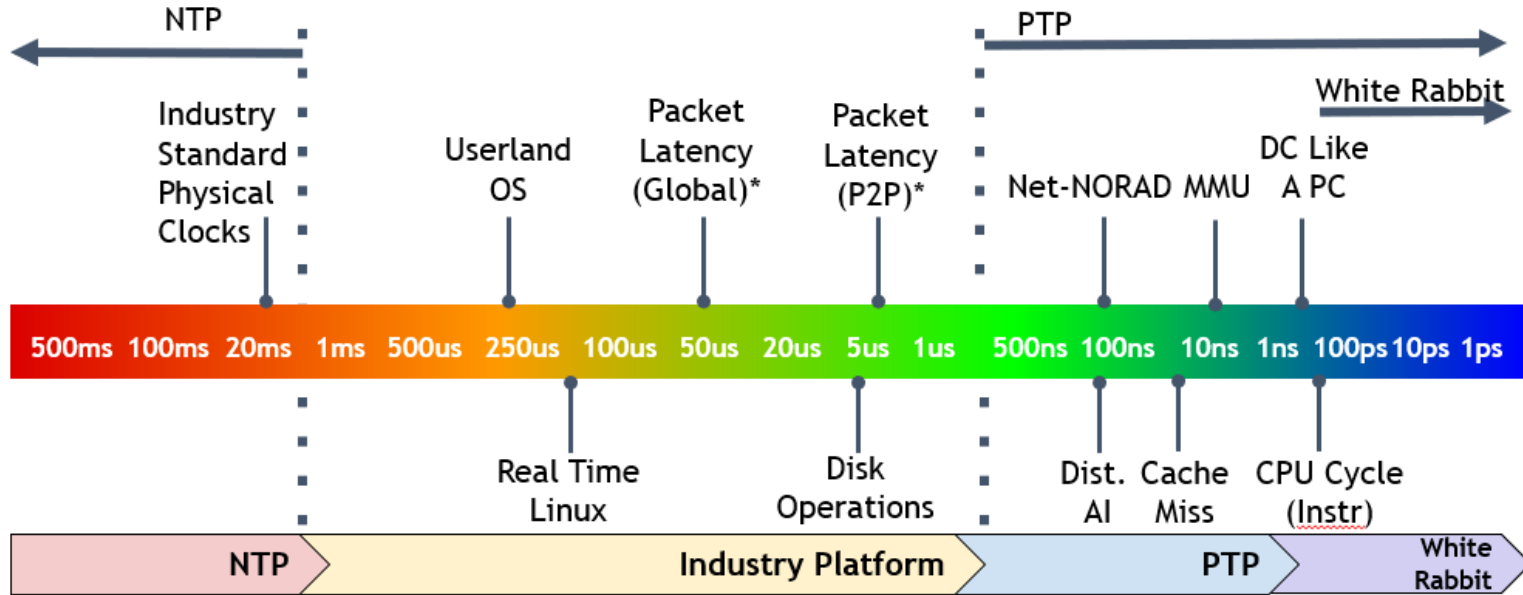


TIME
APPLIANCES

OPEN POSSIBILITIES.



Time Precision Today and Tomorrow



- Global – Data Center CPU to another Data Center CPU around the world
- P2P – CPU to another CPU in the same rack with minimum latency.

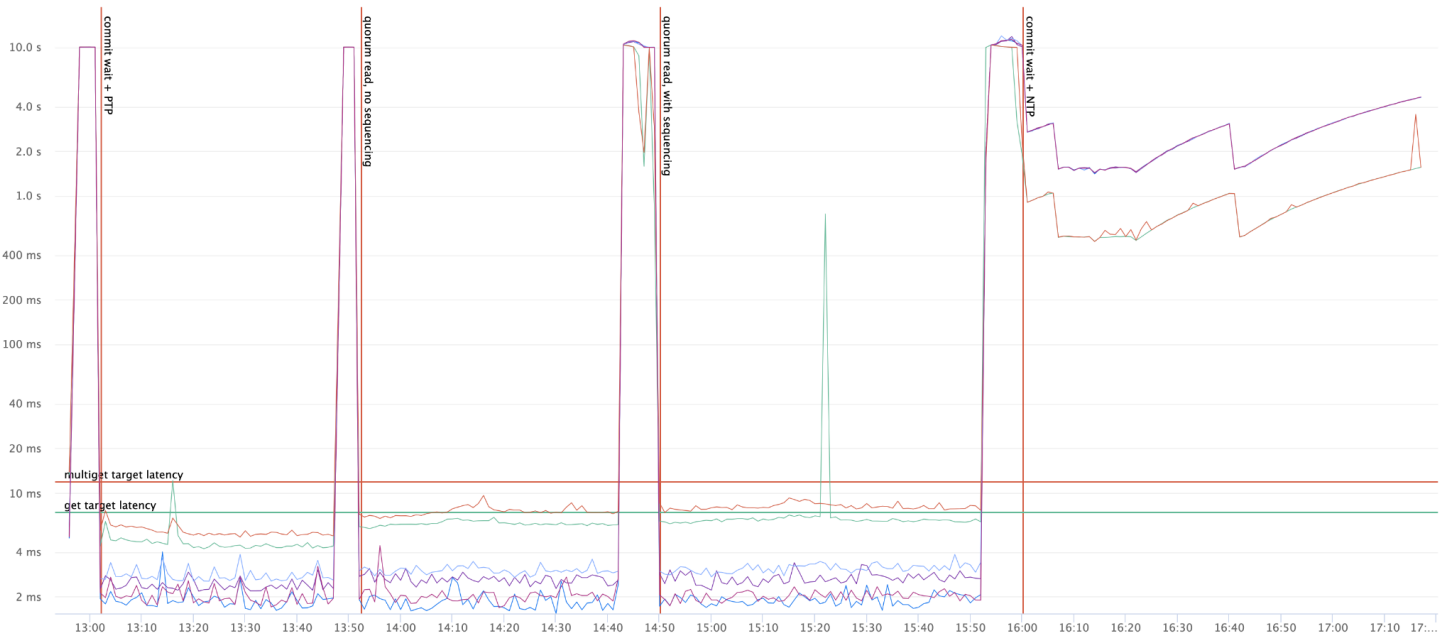
OPEN POSSIBILITIES.



Distributed Database



TIME
APPLIANCES



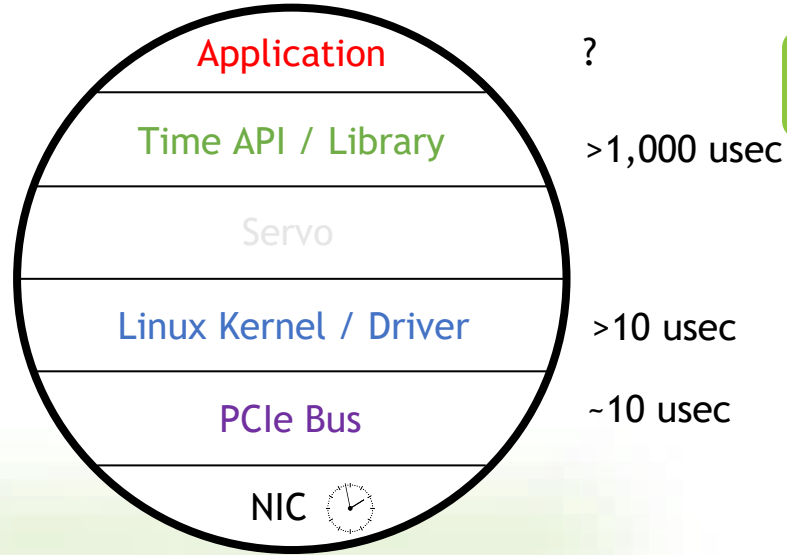
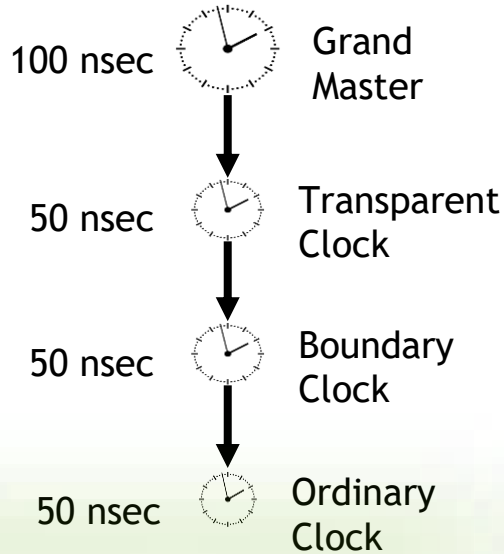
OPEN POSSIBILITIES.



Error Bound



TIME
APPLIANCES



Network Hardware

Software / Application

OPEN POSSIBILITIES.



Call to Action

- How to get involved in the Project?

- TAP Project

Precision Time API - Time APIs to disseminate the time error bound and bring accurate time to the user space processes.

- POC: Georgi Chalakov

- Where to find additional information

https://www.opencompute.org/wiki/Time_Appliances_Project



TIME
APPLIANCES

OPEN POSSIBILITIES.



Open Discussion



OCP
GLOBAL
SUMMIT

NOVEMBER 9-10, 2021