Datacenters and Time Synchronization

Ahmad Byagowi (OCP-TAP)

Open Compute Project
Time Appliance Project (TAP)

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

- Mission:

Mission Statement

1. Create specifications and references for **Data Center Timing** appliances, applications and networking infrastructure
2. Promote openness in **Timing Appliances** and interfaces through open-source implementations

- Work Streams:

	Project	Objective
#1	Open Time Server	Development of an open time server for DC and Edge systems
#2	Data Center PTP Profile	Development of a PTP Profile tailored for data center applications
#3	Precision Time API	Time APIs to disseminate the time error (error bound) and bring accurate time to the user space
#4	Oscillators	Classification and measuring of oscillators

Why Do We Need Synchronization in Data Centers?

“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

3x Distributed Database Throughput

80x Timing Accuracy

Site Security w/ Timed Encryption

↓ Compute

↓ Network Traffic

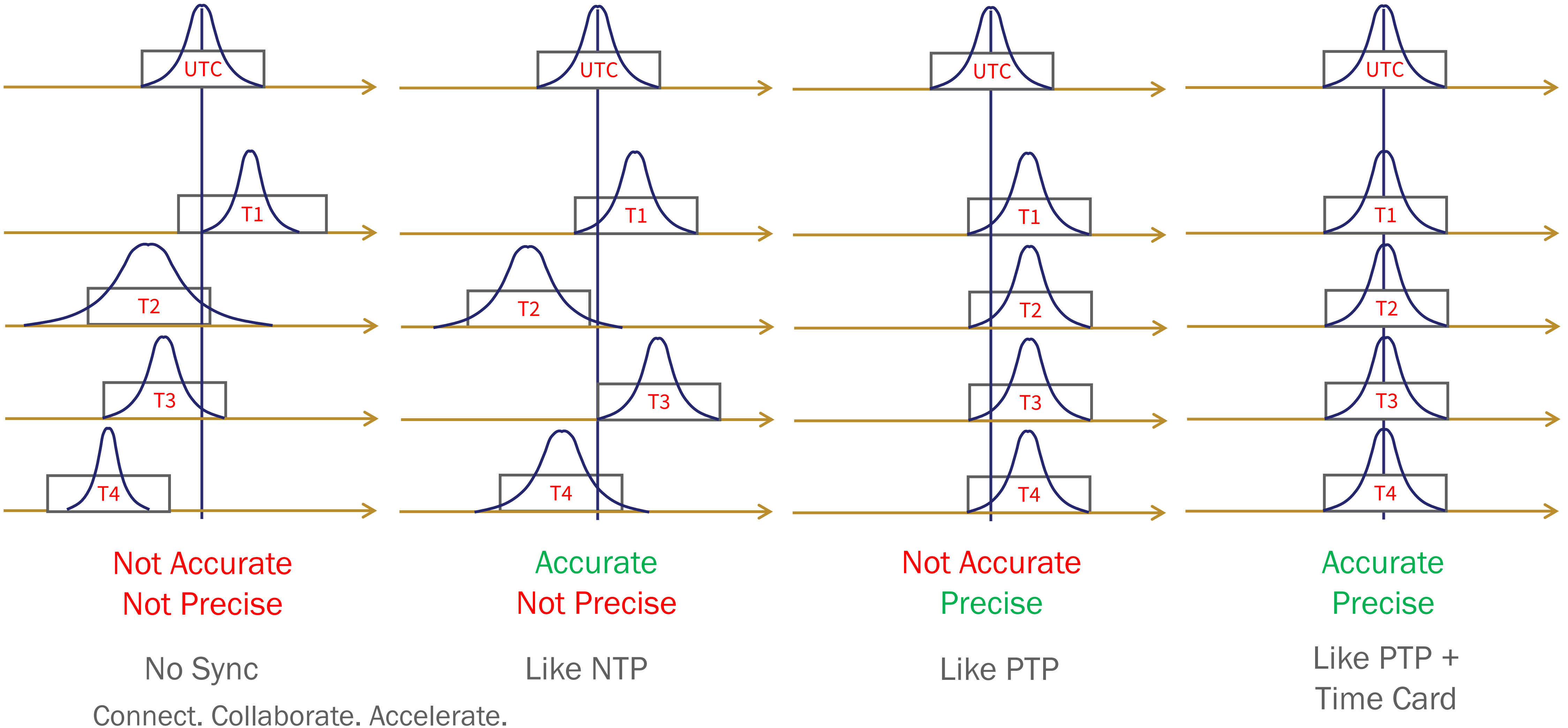
Data Consistency

Casuality

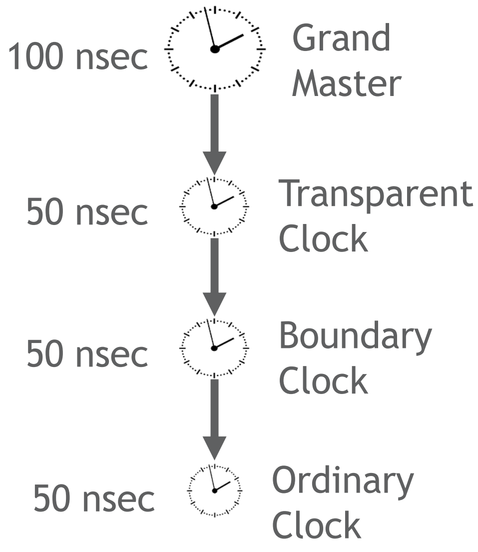
Event Ordering

Just The Tip

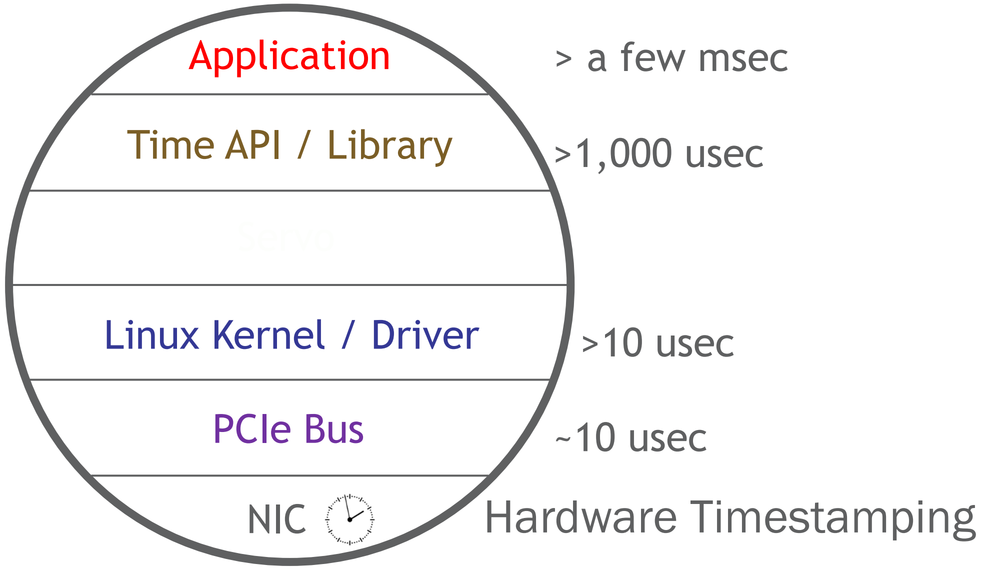
Accuracy vs Precision in Sync



Precision Time Protocol

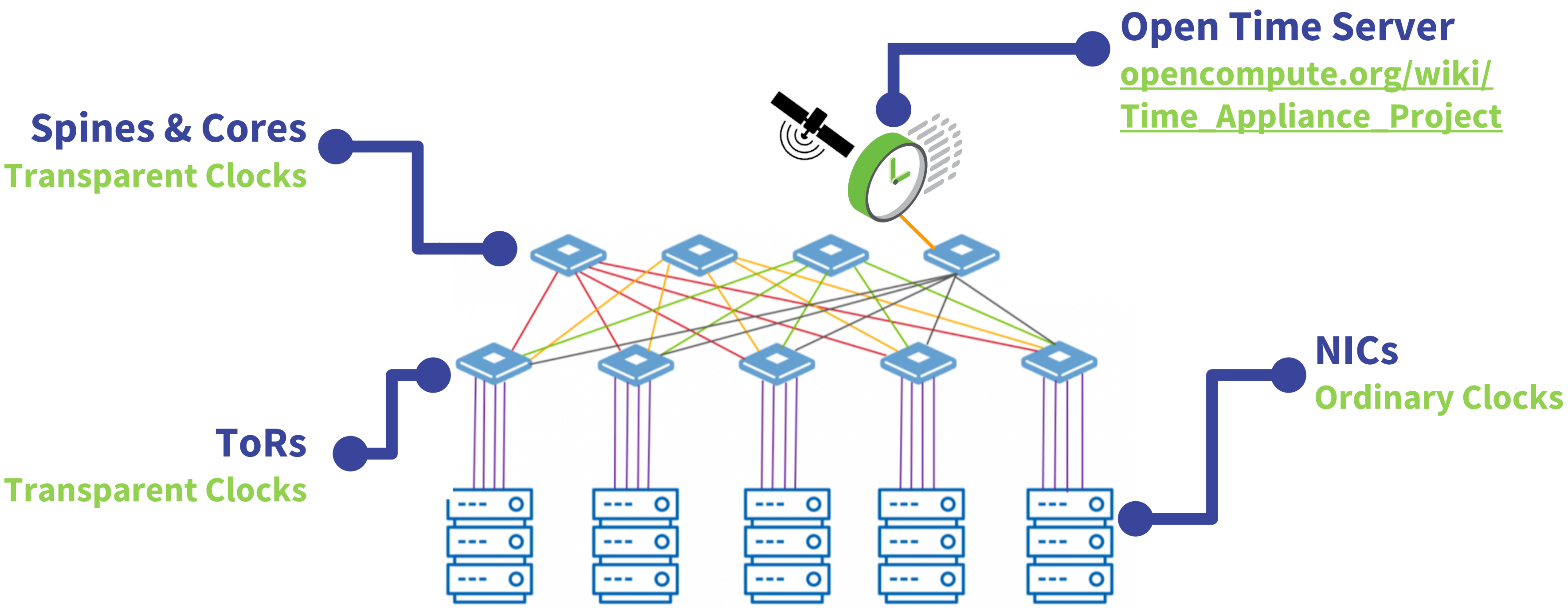


Network Hardware



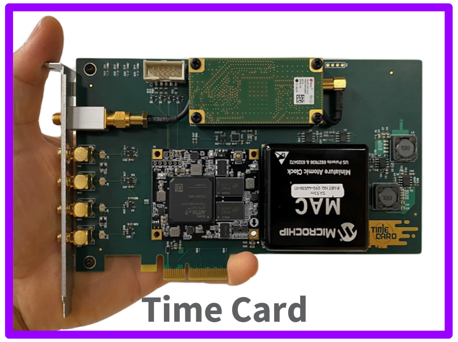
Server Hardware

Synchronization in Data Center



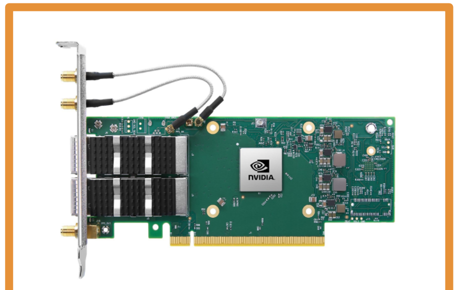
OPEN Compute Project®

Open Time Server

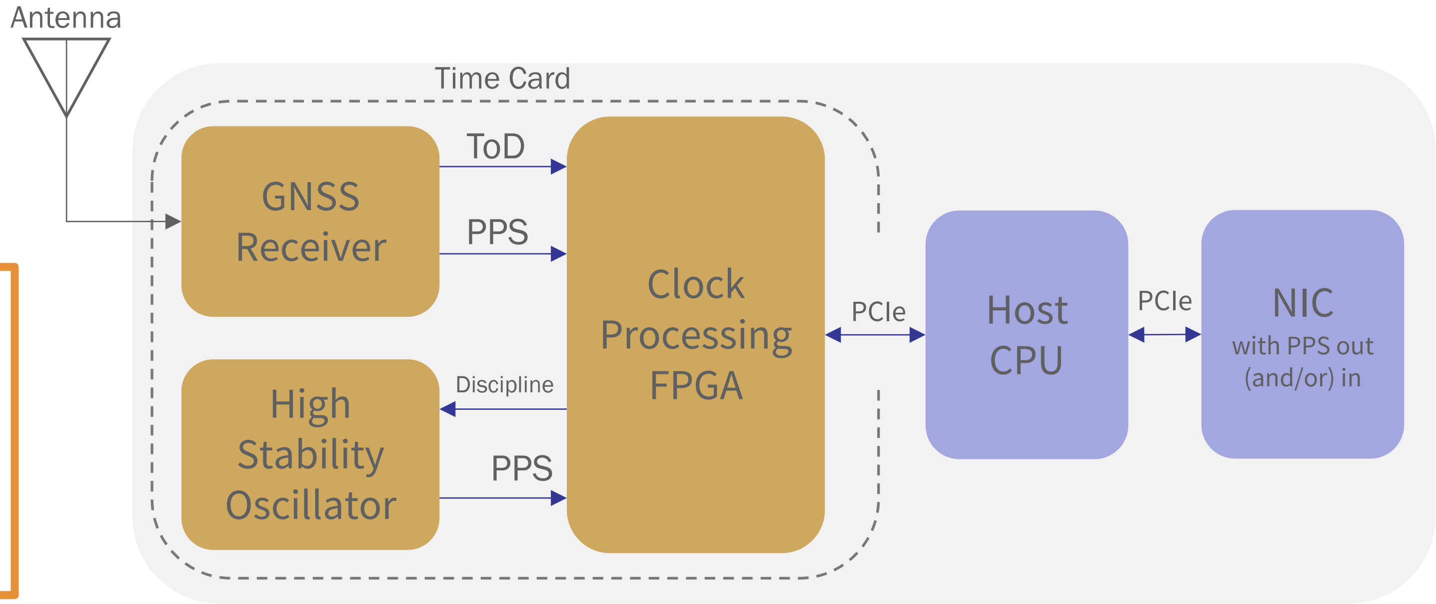


Time Card

HPE DL380



NVIDIA ConnectX-6 Dx



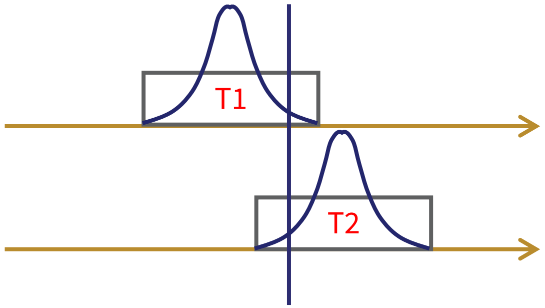
Use Case: Network Telemetry

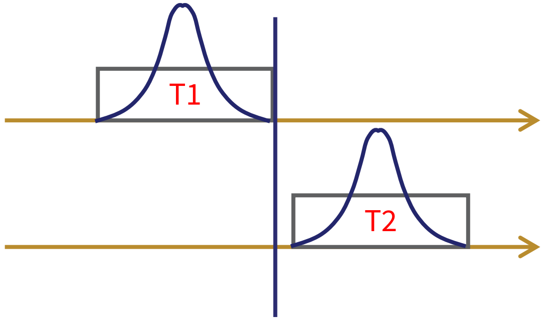
- Constantly pings machines
  - If machine doesn’t respond, it must take an action.
- Why not do pings based on Hardware Timestamps
  - SING = Synchronous Pings
  - One way delay measurements
- In-Network Telemetry
  - Improve Congestion recognitions
  - Improve Congestion Control mechanisms
- End-to-End Precision: <100ns
  - Want to measure one way latency

https://engineering.fb.com/2016/02/18/core-data/netnorad-troubleshooting-networks-via-end-to-end-probing/

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



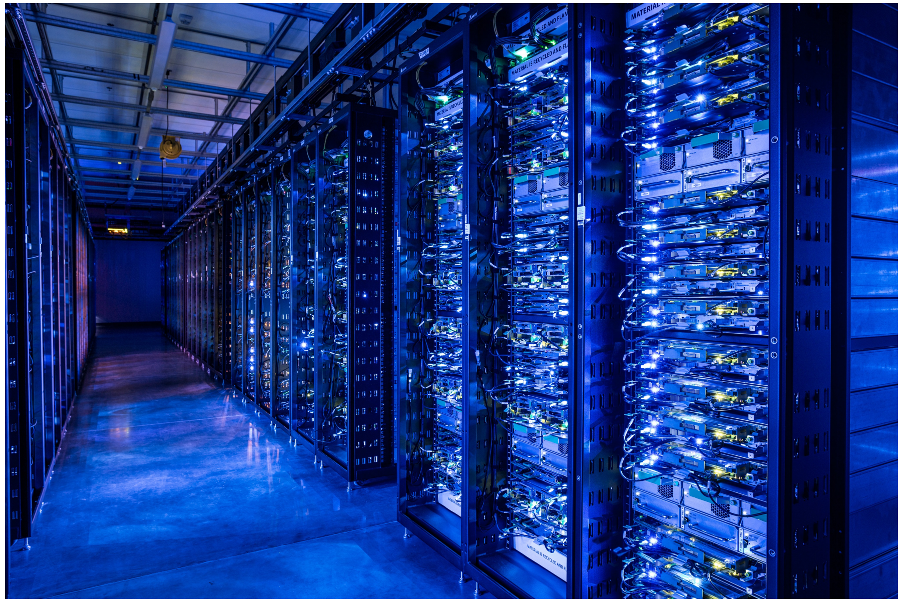


Use Case: Distributed AI

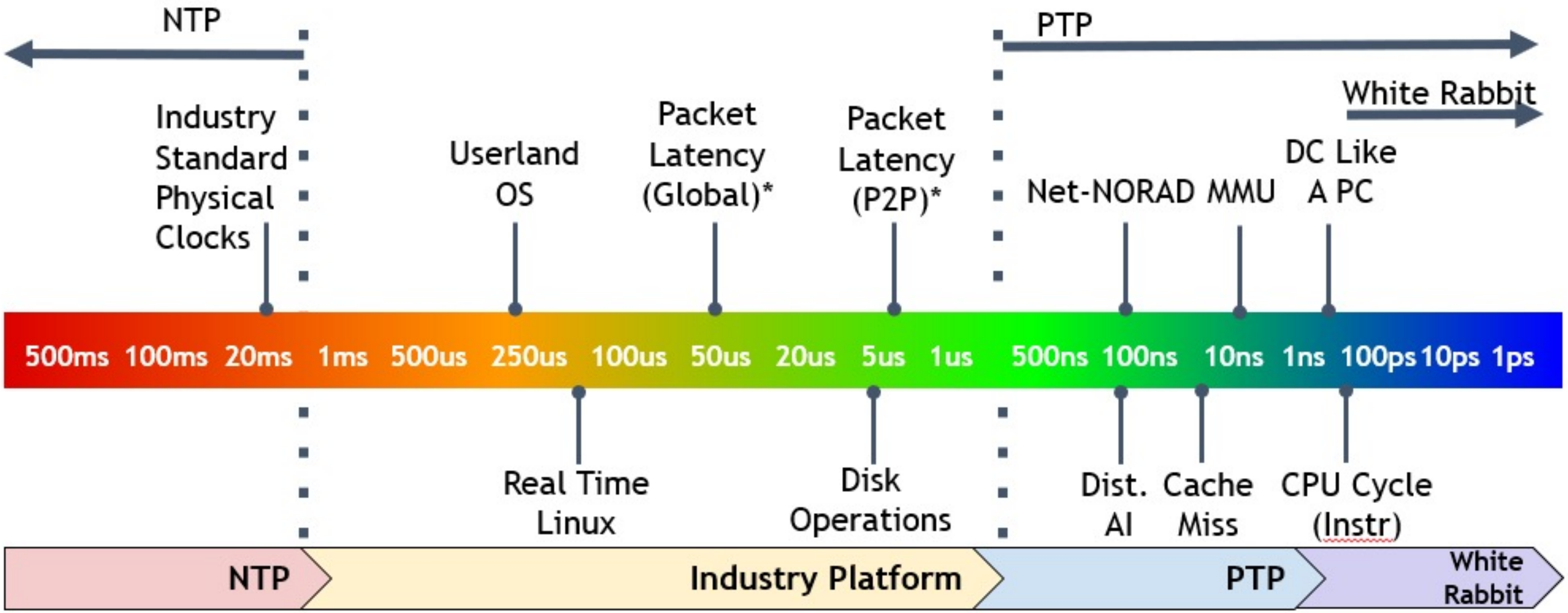
- Resource Intensive to move data to one machine or cluster
- With the right precision, you can train in many places
- Then use the timestamps to merge the results
- Advantages:
  - Reduces data center traffic/congestion
  - Save Resources
- Requires end-to-end precision of <100ns
  - Across the data center
  - Globally

Use Case: Multicore Systems Across the Network

- Data Center Network is the Fabric
  - Ultra-Path Interconnect (UPI) over the network
  - Input-Output Memory Management Unit (IOMMU) over the network
- Can we program a DC like a PC?
  - We know how to program a Personal Computer well.
  - Precise time can help us program the Data Center Better
  - All DC equipment follows the same precise time vector
- Benefit:
  - Current data center loads are far from 100%
  - Determinism: If you know when everything happens, the load could be closer to 100%
- Requires End-to-End Precision of <10ns



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.

erate.

Thank You

Connect. Collaborate. Accelerate.

OPEN Compute Project®