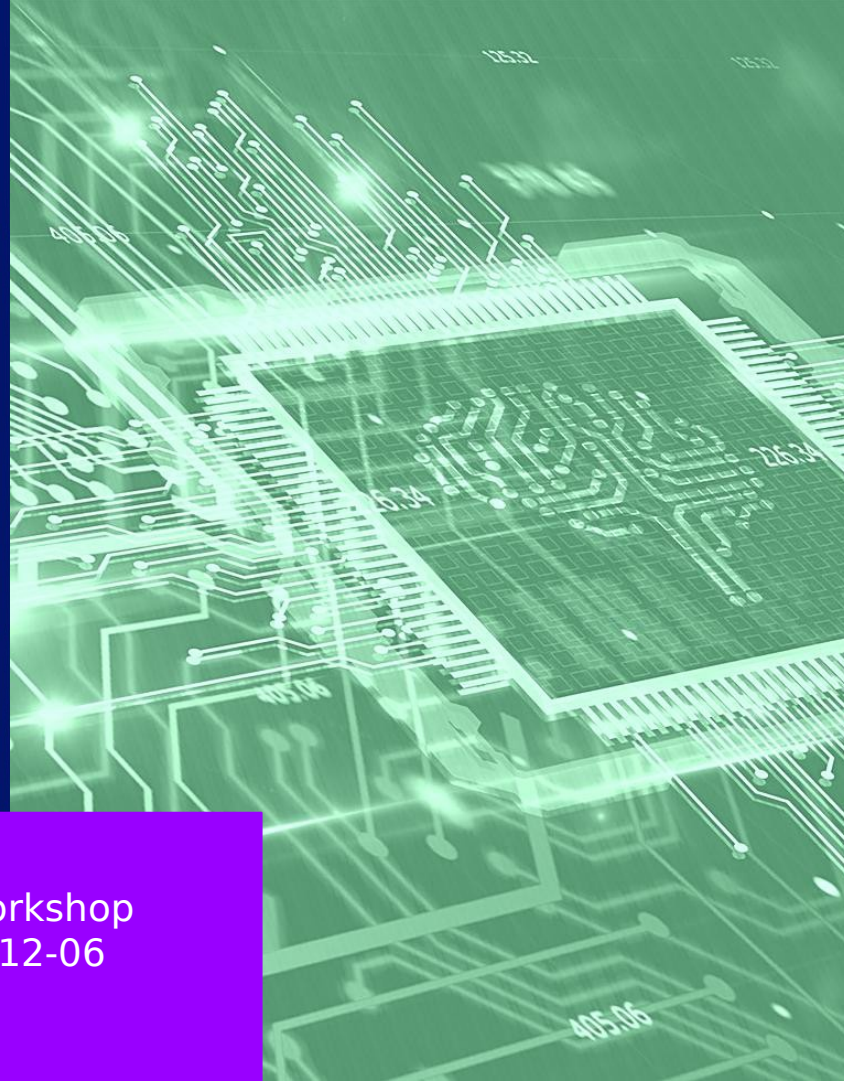
QBayLogic.

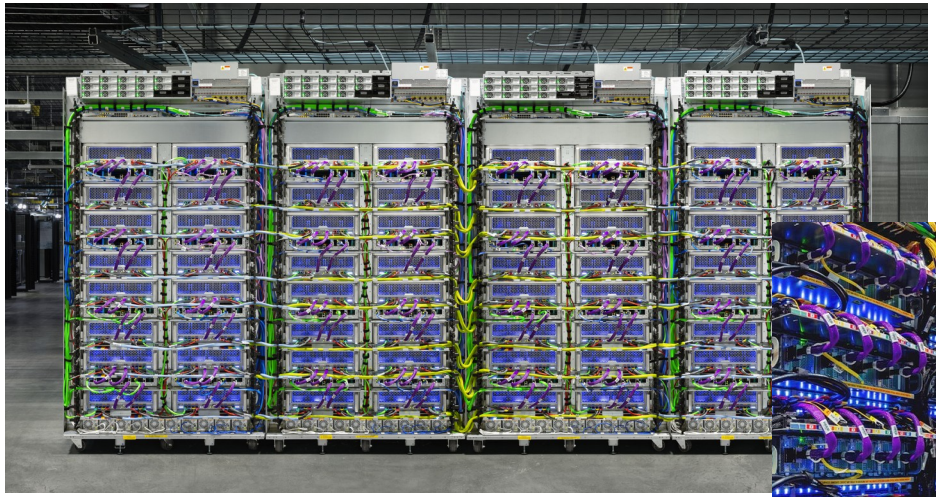
Bittide: Synchronizing Distributed Systems with Logical Time

Speaker: Christiaan Baaij
Work by: Many, from multiple institutions. From QBayLogic: Martijn Bastiaan, Lucas Bollen, Lara Herzog, Felix Klein, Hidde Moll, Leon Schoorl.

FIRE workshop
2024-12-06



Data centers



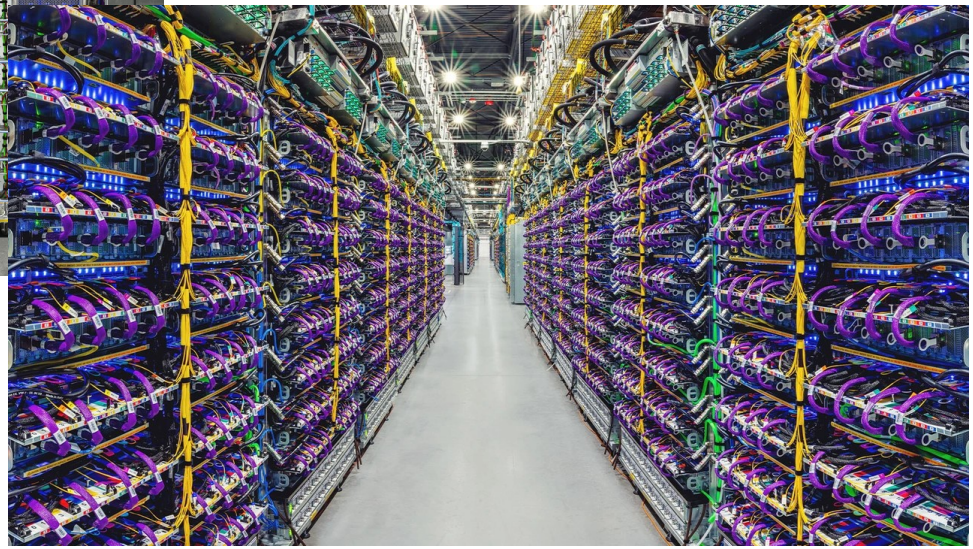


Image source: https://www.semianalysis.com/p/tpuv5e-the-new-benchmark-in-cost
Image source: https://www.datacenterknowledge.com/data-center-chips/data-center-chips-in-2024-top-trends-and-releases

Asynchronous every

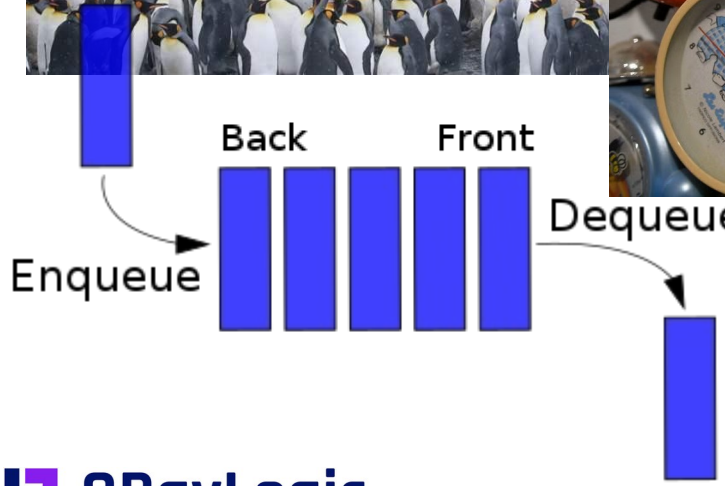


Image source: Damien Meyer/Getty Images
Image source: https://blog.oureducation.in/schedulers-in-operating-system/
Image source: The Pew Charitable Trusts
Image source: https://en.wikipedia.org/wiki/Network_scheduler

QBayLogic.

Tail latency

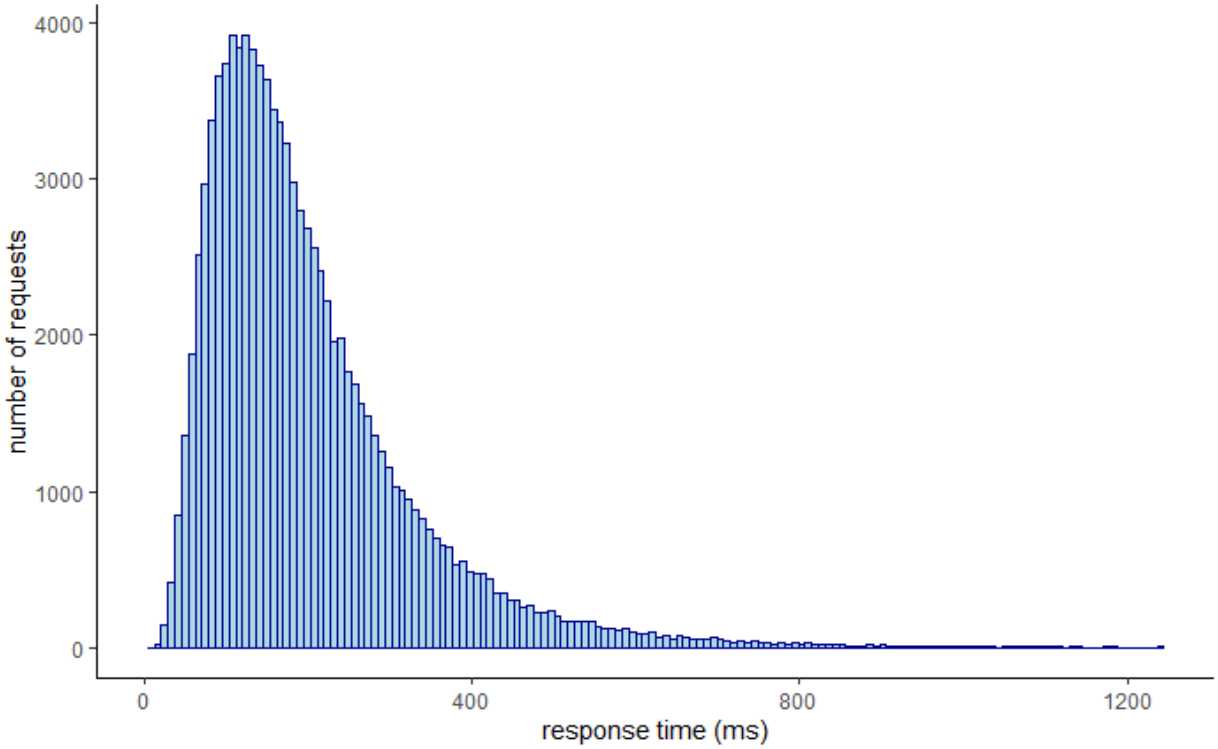


Image source: https://robertovitillo.com/why-you-should-measure-tail-latencies/

Scheduling

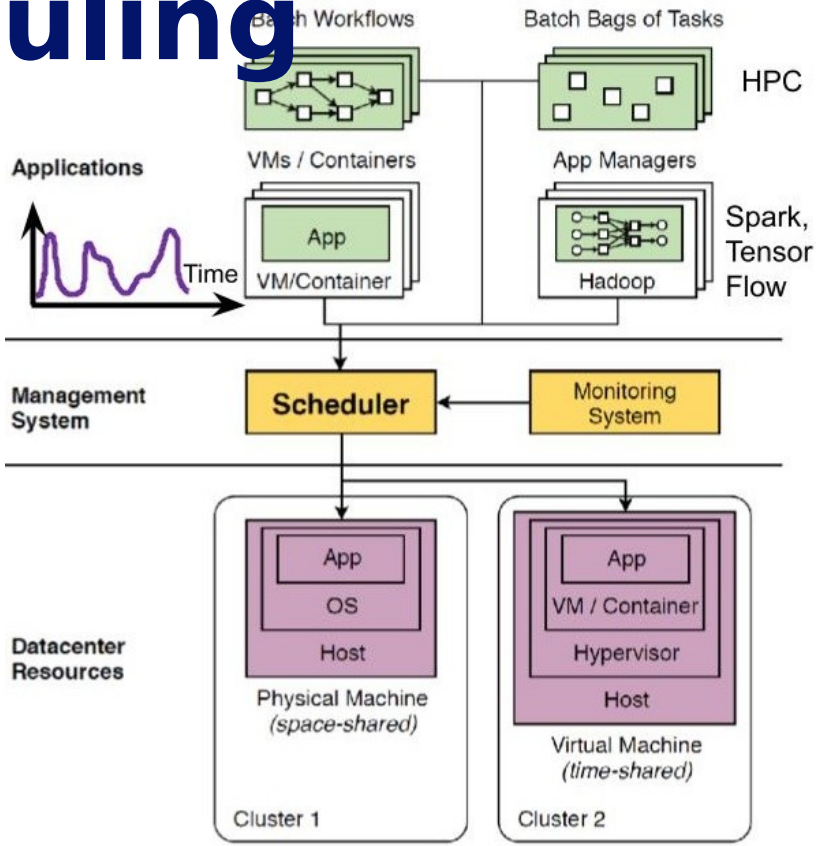


Image source: Georgios Andreadis, Fabian Mastenbroek, Vincent van Beek, Alexandru Iosup (2022) Capelin: Data-Driven Compute Capacity Procurement for Cloud Datacenters Using Portfolios of Scenarios. IEEE Trans. Parallel Distributed Syst. 33(1)

Bittide – https://bittide.io

- Synchronous execution of tasks at datacenter scale
- Without the need for a central clock or reference to universal time

- Nodes monitor neighbor's frequency
- Adjust own frequency

- Together with elastic buffers (to absorb wobbles)
- Allows for a shared sense of ***logical*** time
- A mapping from one clock to any other clock in the system

Logical time

- What it is **not**:
  - Node A sends a message at 02:00:00.000 UTC
  - Node B receives said message at 02:00:00.002 UTC
- What it **is**:
  - Node A sends a message at clock cycle 1.200.800 (reference clock Node A)
  - Node B receives said message at clock cycle 4.600.808 (reference clock Node B)
  - When A sends a message at clock cycle **X**, B will see it at the **X + 3.400.008**'th clock cycle

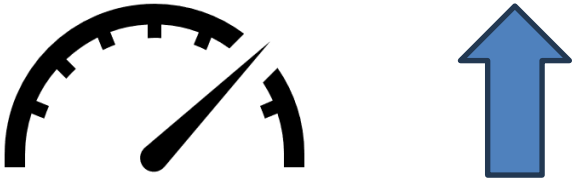
Logical time

- What it **is**:
  - Node A sends a message at clock cycle **1.200.800** (reference clock Node **A**)
  - Node B receives said message at clock cycle **100.807** (reference clock Node **B**)
  - When A sends a message at clock cycle **X**, B will see it at the **X - 1.099.993**'th clock cycle
- Differences between nodes ***cannot*** be interpreted in physical terms (wall-clock latency)
- ***Only*** round-trip times have such a physical interpretation

Buffer control – Above midpoint

Buffer fill level above midpoint

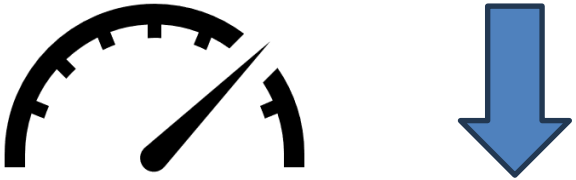
Increase clock frequency



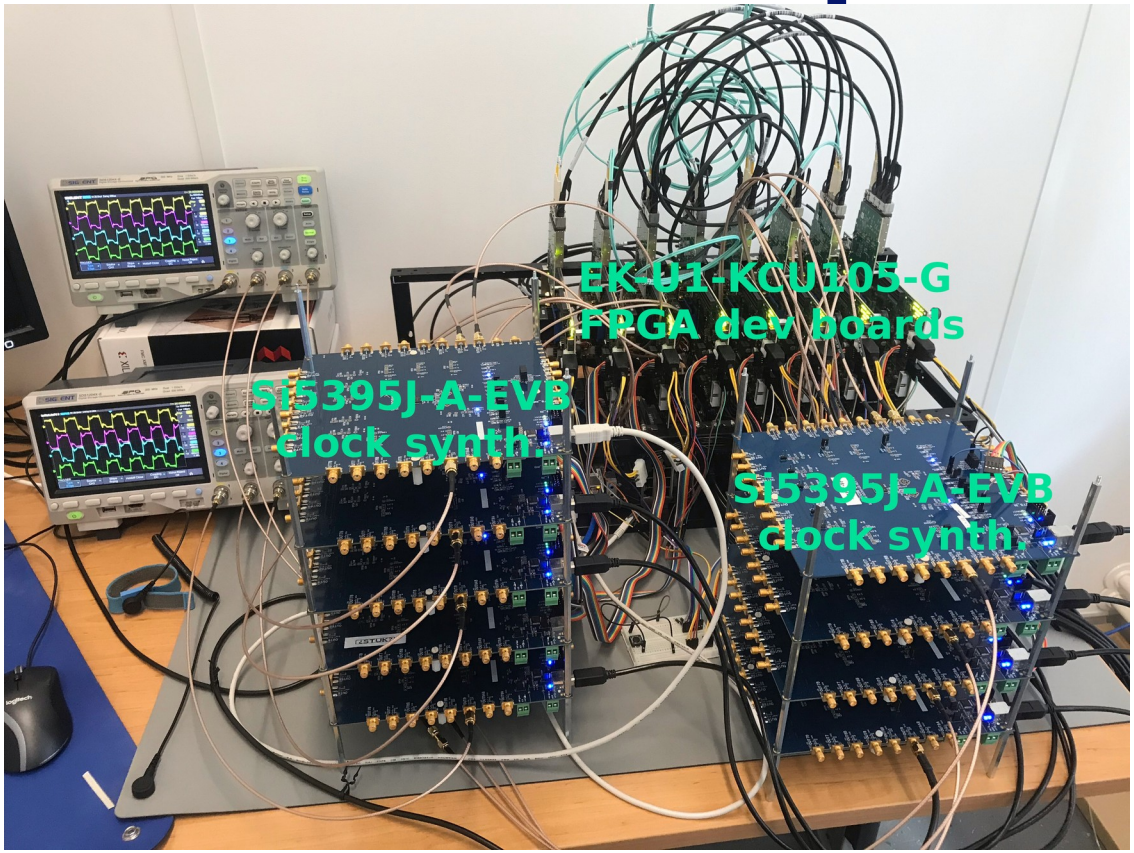
Buffer control – Below midpoint

Buffer fill level below midpoint

Decrease clock frequency



Experimental setup



QBayLogic.

Conclusions - https://bittide.io

- **Bittide System:** Ahead-of-time time-multiplexing of datacenter-scale distributed applications at clock-cycle accuracy.
- **Efficient:** No communication overhead (in-band signaling)
- **Inexpensive:** Relatively minor additional hardware needed
  - PLL with a digitally controlled oscillator (DCO) mode
  - Convenient: Transceivers with CDR; but those are already prevalent for modern networks (10Gbit/s connections and up)