Fresh replicas with append-only storage

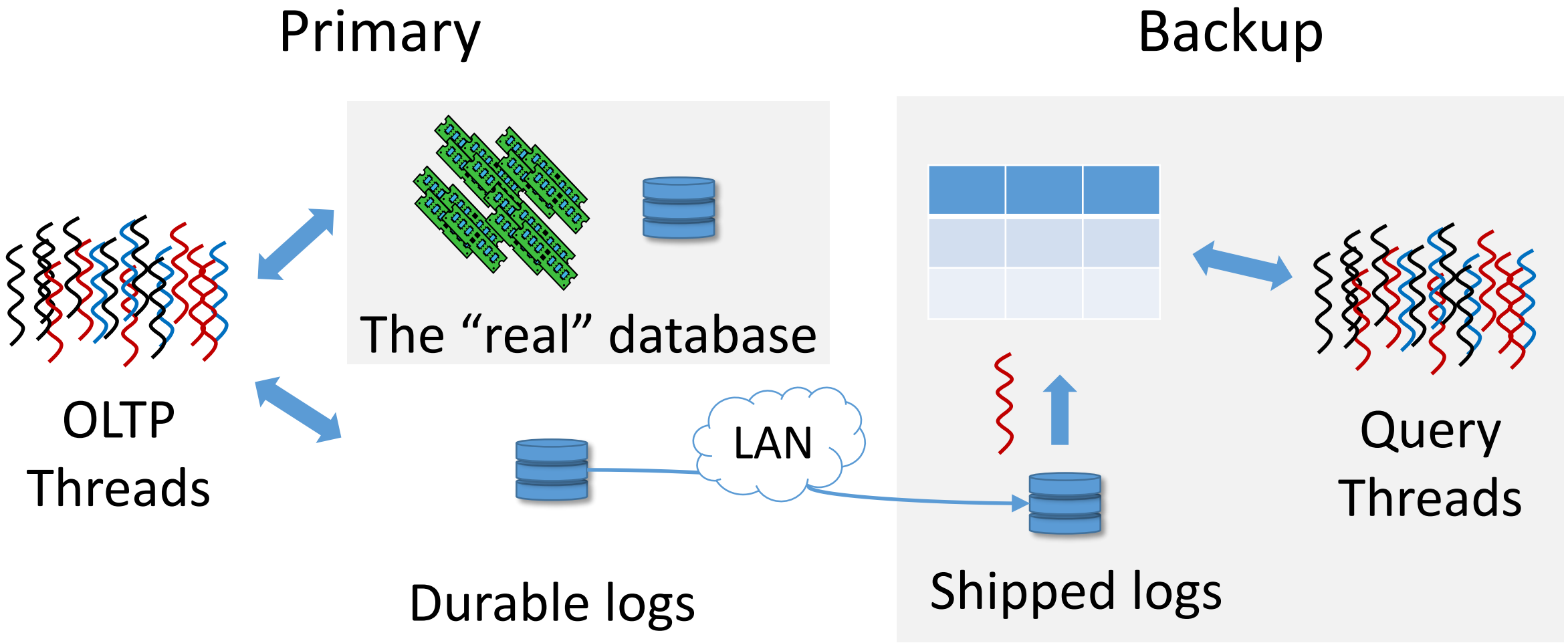
Tianzheng Wang

UNIVERSITY OF TORONTO

Ryan Johnson

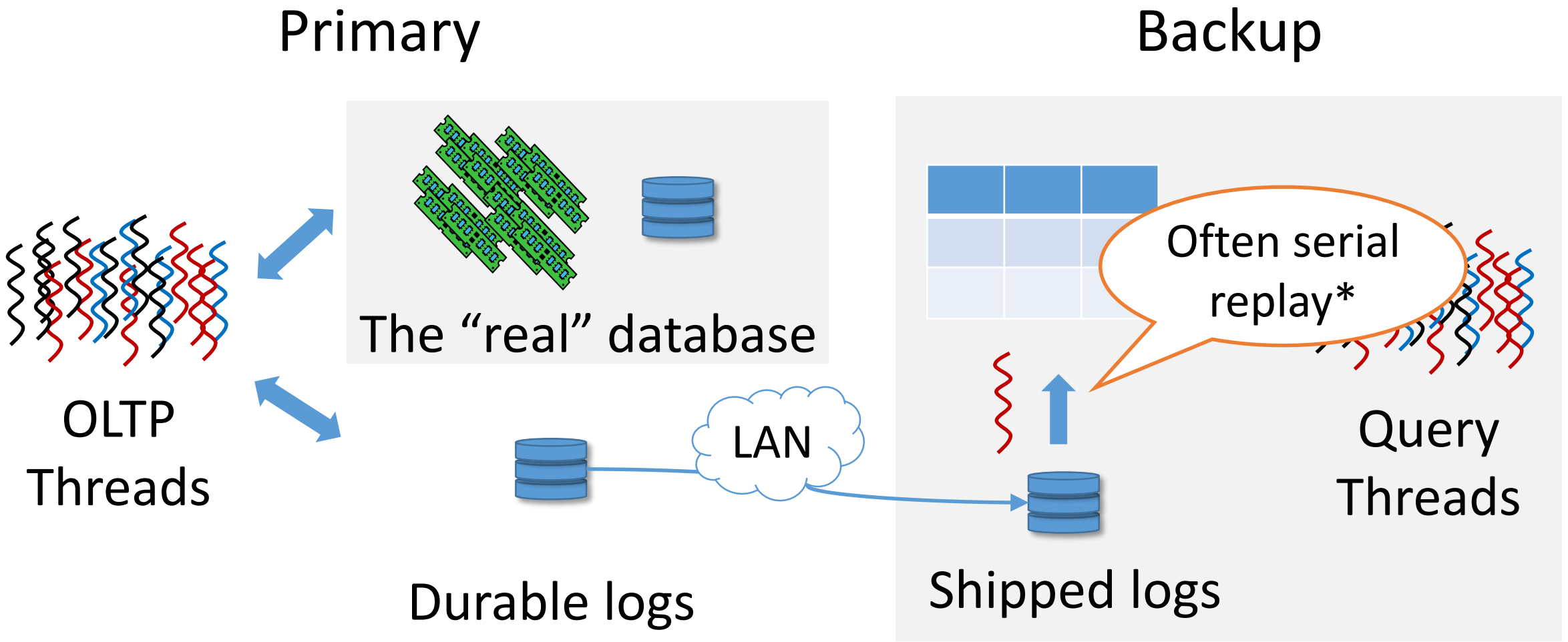
Ippokratis Pandis

Hot standby: parallel-execute, serial-replay



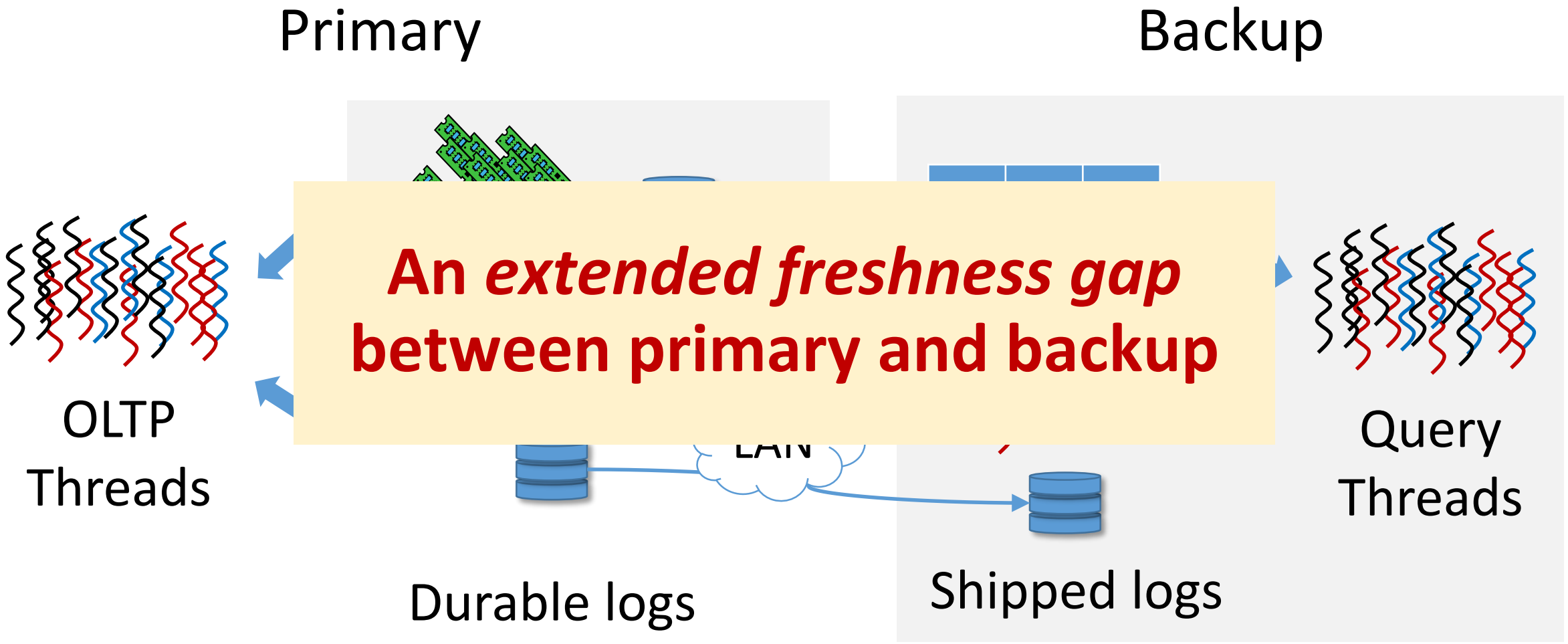
* Yang etc. KuaFu: Closing the parallelism gap in database replication, ICDE ’13.

Hot standby: parallel-execute, serial-replay

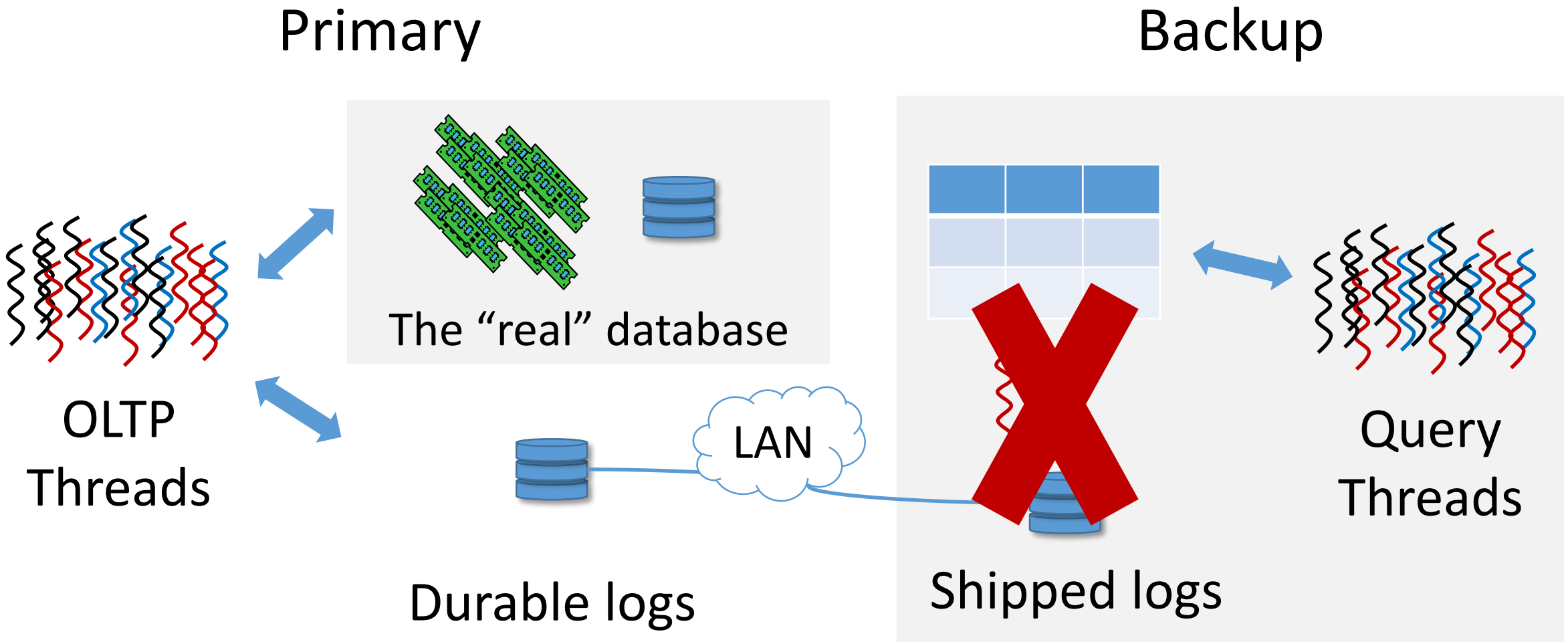


* Yang etc. KuaFu: Closing the parallelism gap in database replication, ICDE ’13.

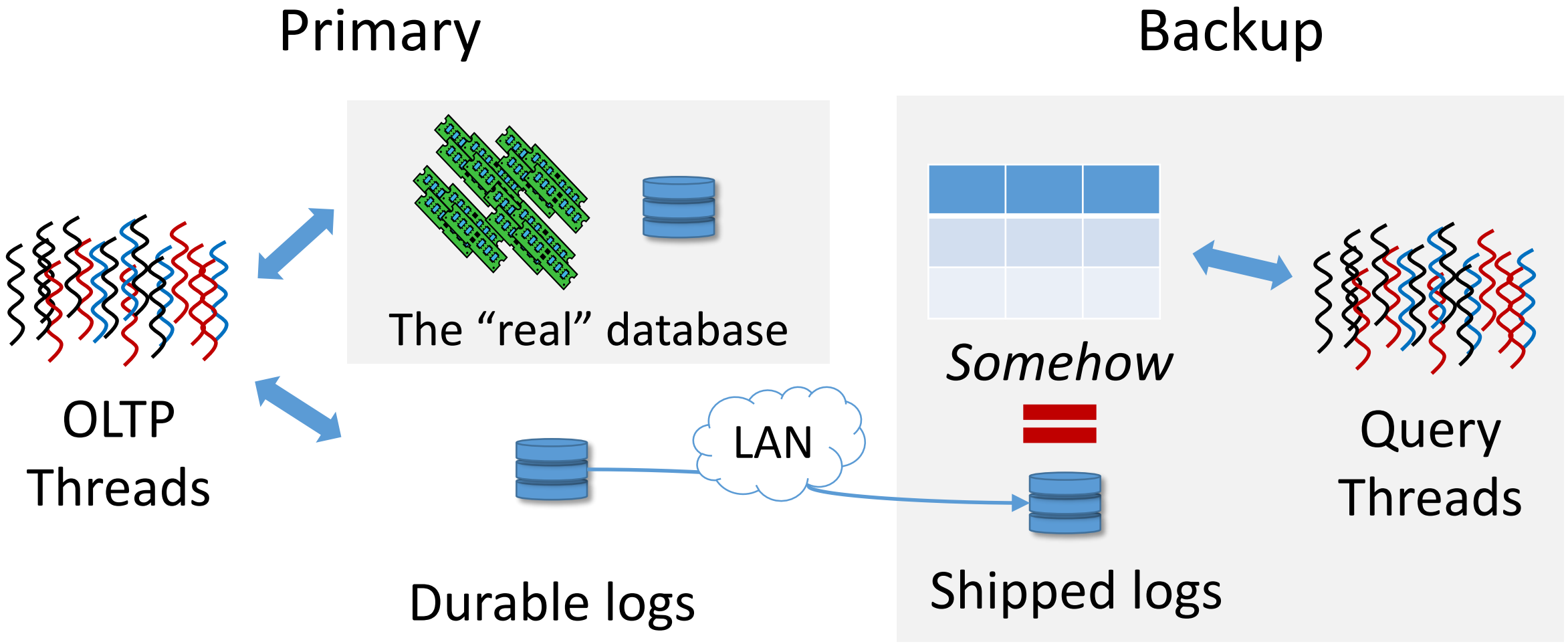
Hot standby: parallel-execute, serial-replay



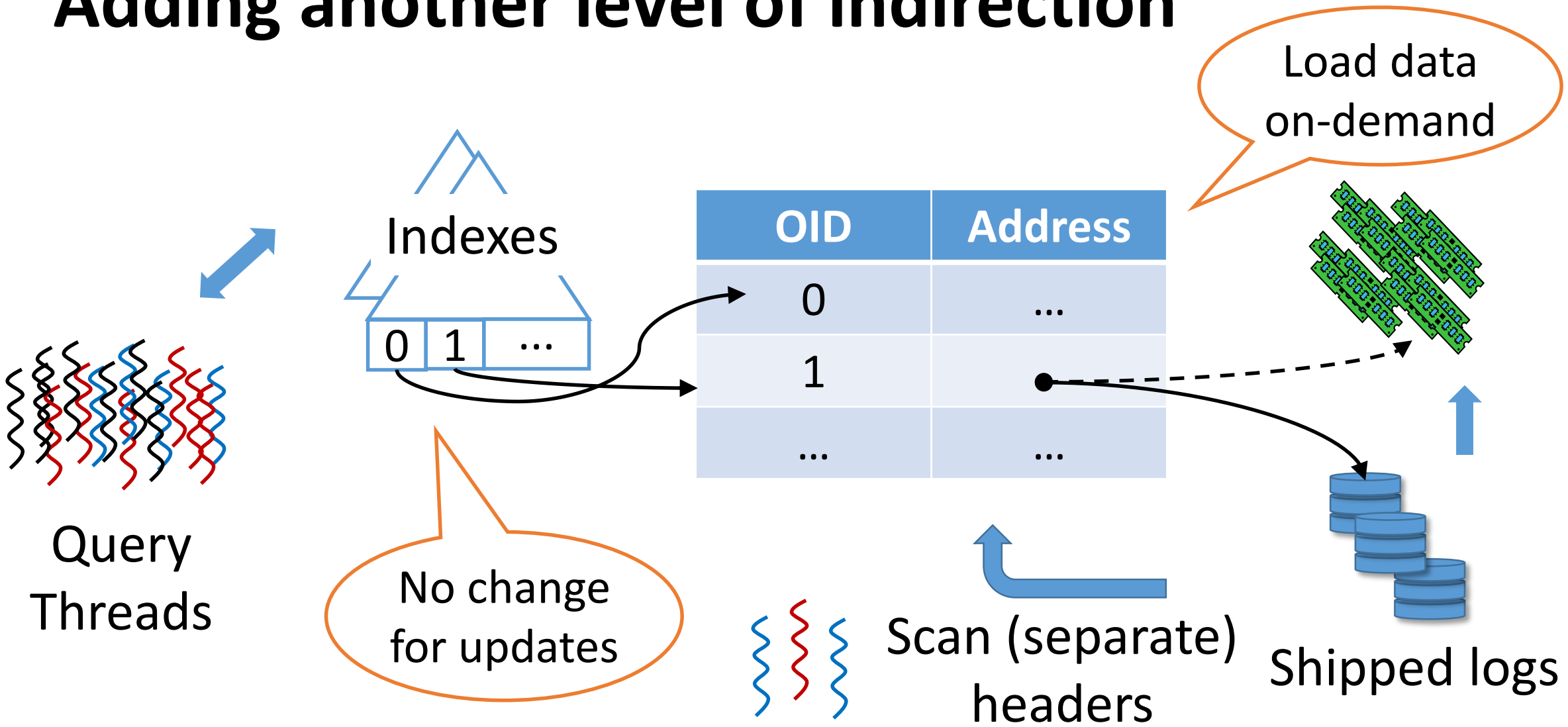
What if the log is the database?



Reduced/little replay ➔ fresh replicas



Adding another level of indirection



Key enablers and conclusion

- High-speed network
  - RDMA over Infiniband
- Non-volatile memories
  - Ship the log buffer once it’s durable
  - NV-DIMMs (battery + DRAM + flash) will suffice
- Append-only storage + new hardware = fresh replicas
  - Related work: Corfu [Balakrishnan ’12], Hyder [Bernstein ’13], Indirection array [Sadoghi ’14], logging in NVRAM [Wang ’14], etc.
  - Combining these techniques: - very lightweight “replay”