

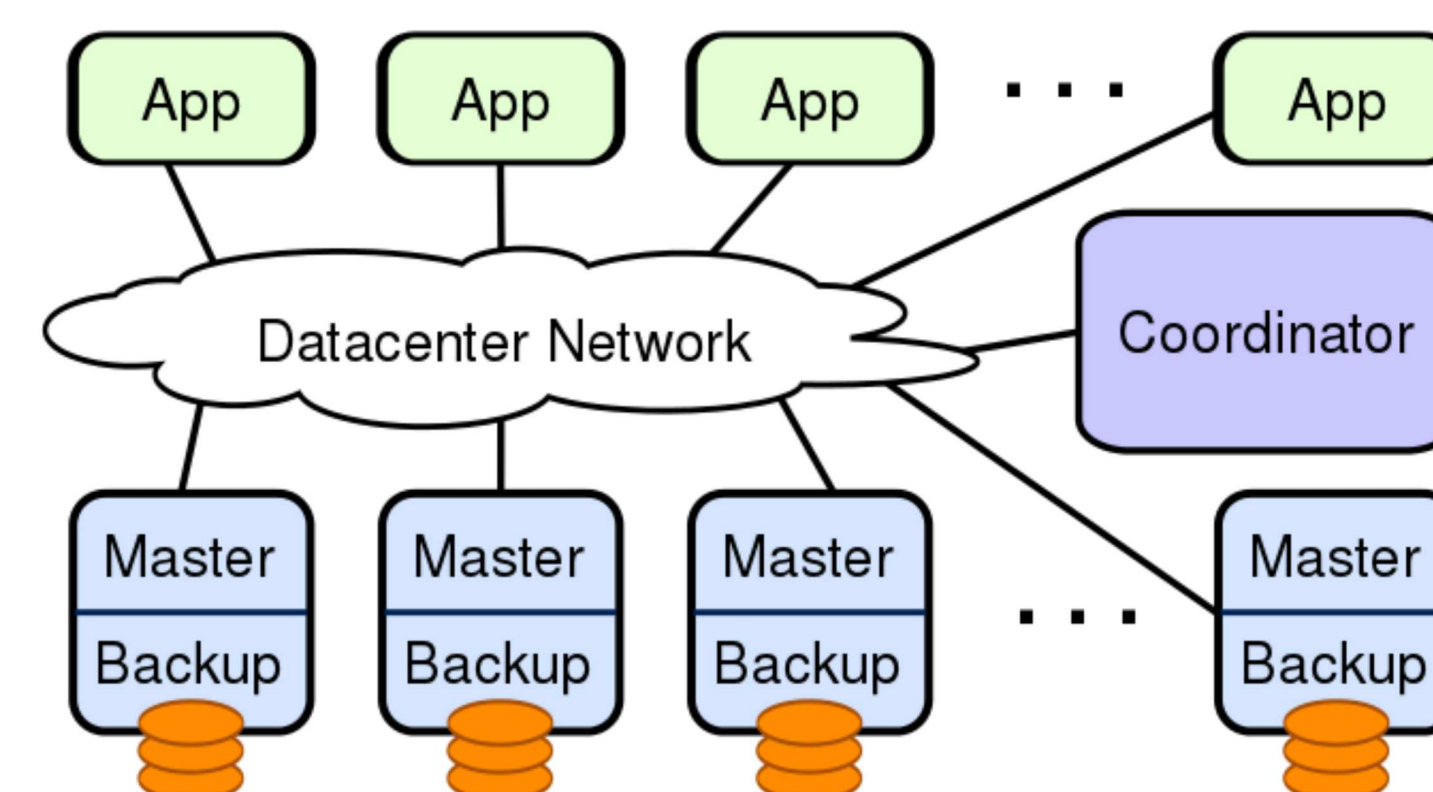
Fast Crash Recovery in RAMCloud

Diego Ongaro, Stephen M. Rumble, Ryan Stutsman,
John Ousterhout, and Mendel Rosenblum

RAMCloud

- Storage for datacenters
- All data always in RAM
 - 1,000-10,000 commodity servers
 - 32-64 GB DRAM/server
- Durable and available
- High performance
 - High throughput: **1M ops/sec/server**
 - Low-latency access: **5-10µs RPC**
- Data model similar to key-value store

Cluster Configuration

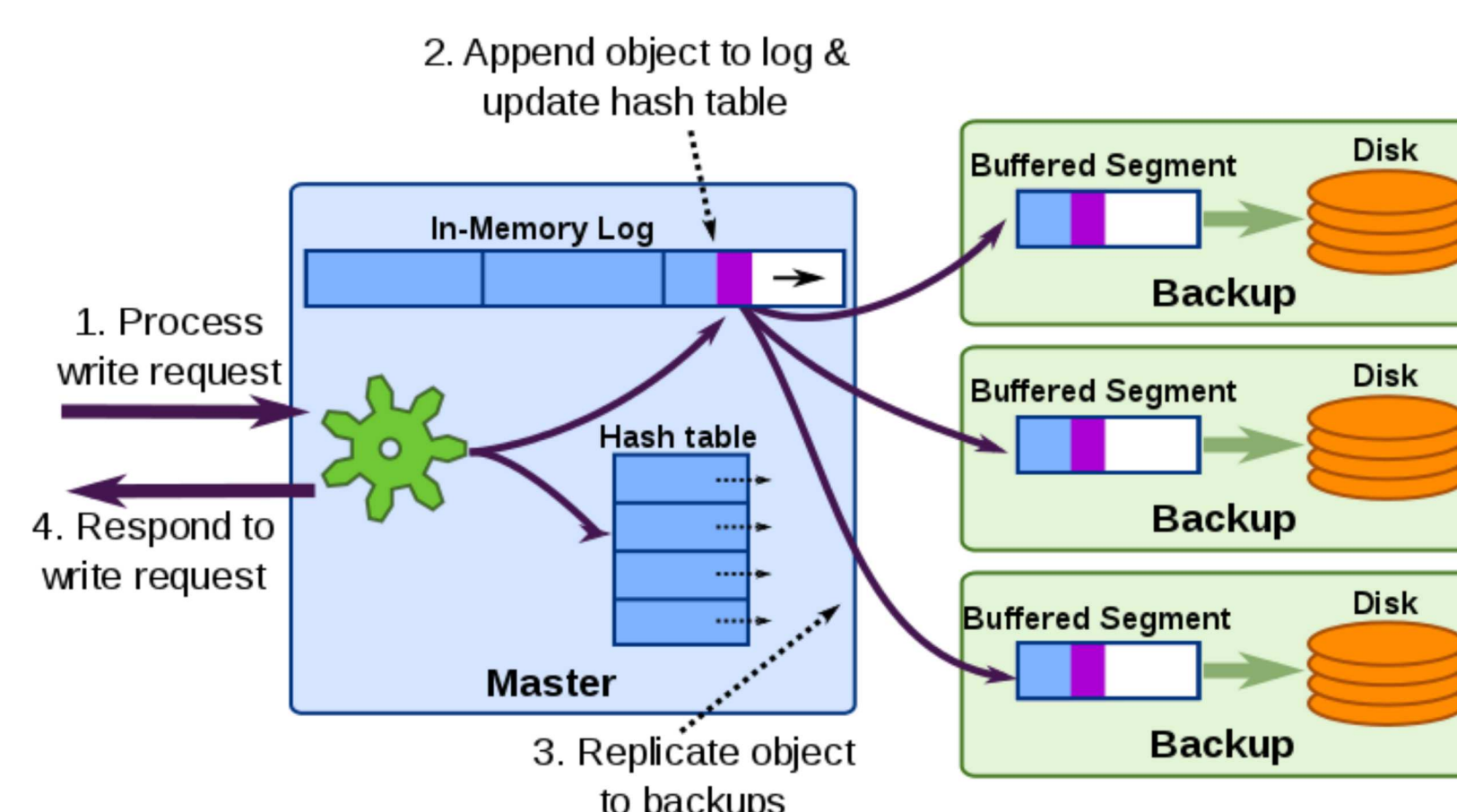


- Applications access RAMCloud using RPCs
- Coordinator manages server pool
- Masters store application data
- Backups store copies on disk/flash

Why Fast Recovery?

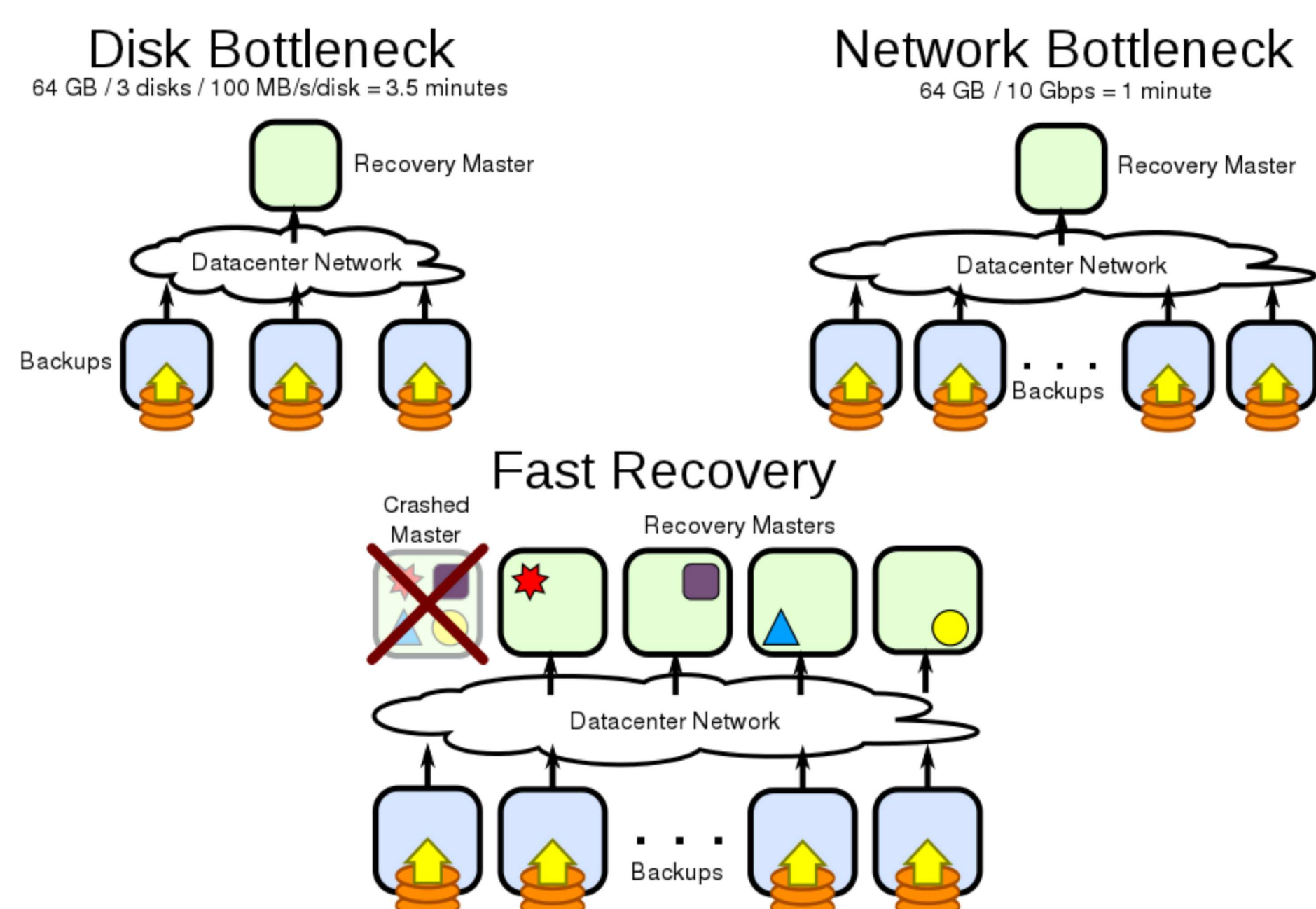
- Durability goals:
 - Small impact on performance
 - Minimum cost and energy
- Keep replicas in DRAM of other servers?
 - Triples cost and energy usage
 - Power failures are still a problem
- RAMCloud's approach: fast recovery
 - 1 copy in DRAM, backup copies on disk/flash
 - Hypothesis: failures will not be noticed

Durability

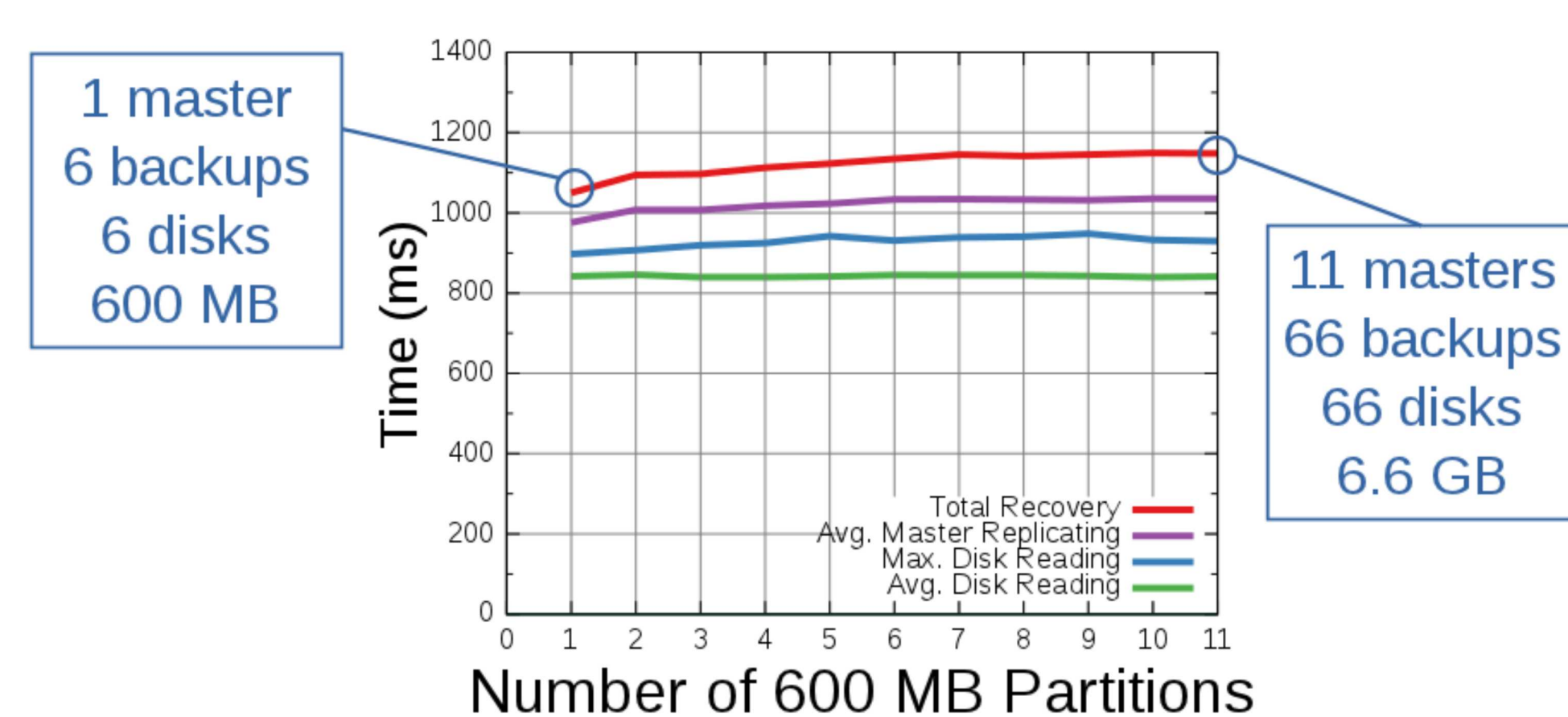


- Masters replicate writes to backups immediately
- Backups buffer it and flush to disk/flash in batch
 - Need auxiliary power source for these buffers

Recovery Idea



Evaluation



- 36-node cluster, 32 Gbps Infiniband network
- Recovered 6.6 GB in ~1 second
- Expect to recover entire master in 1-2s at scale