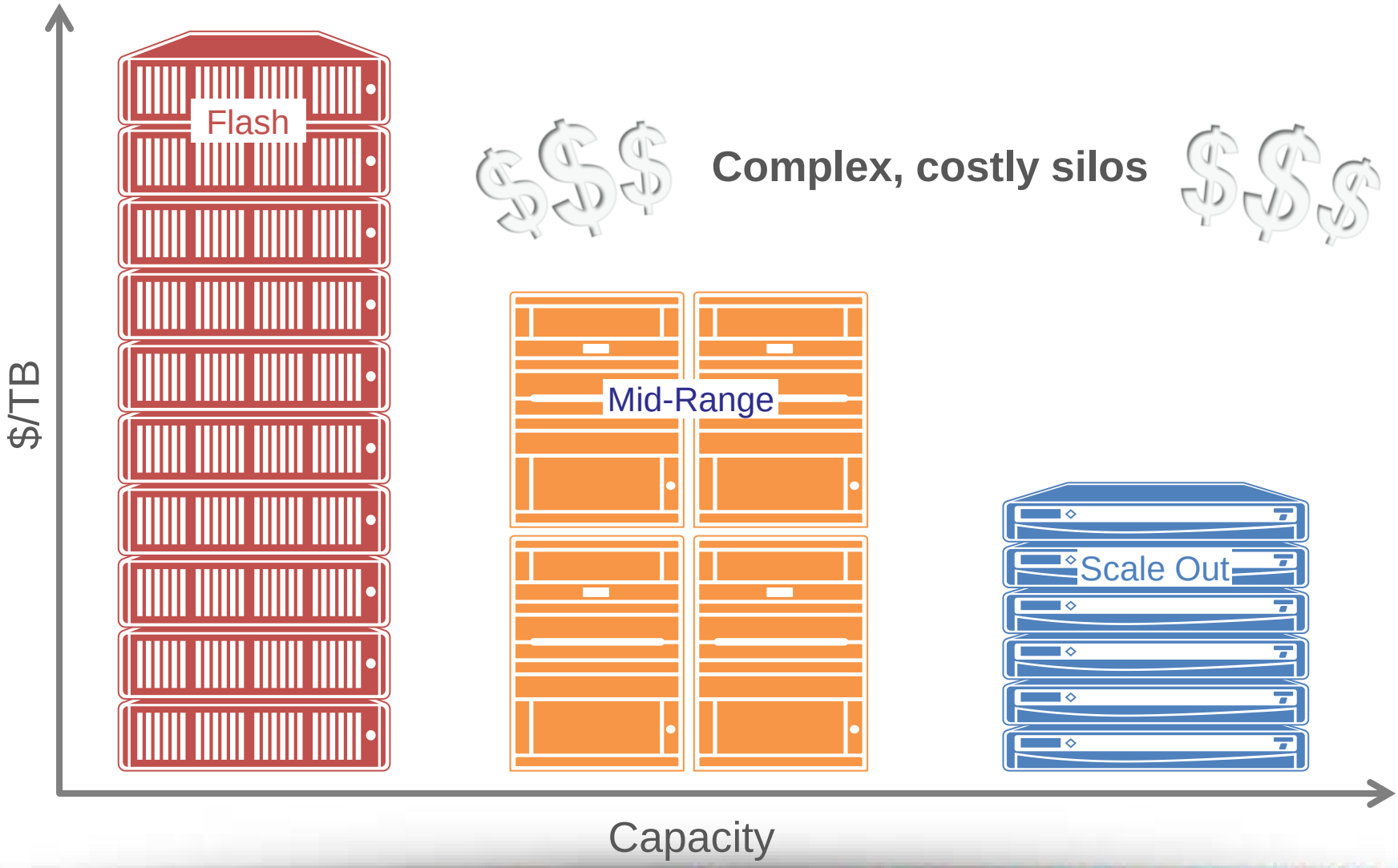
SDC
STORAGE DEVELOPER CONFERENCE
SNIA SANTA CLARA, 2015

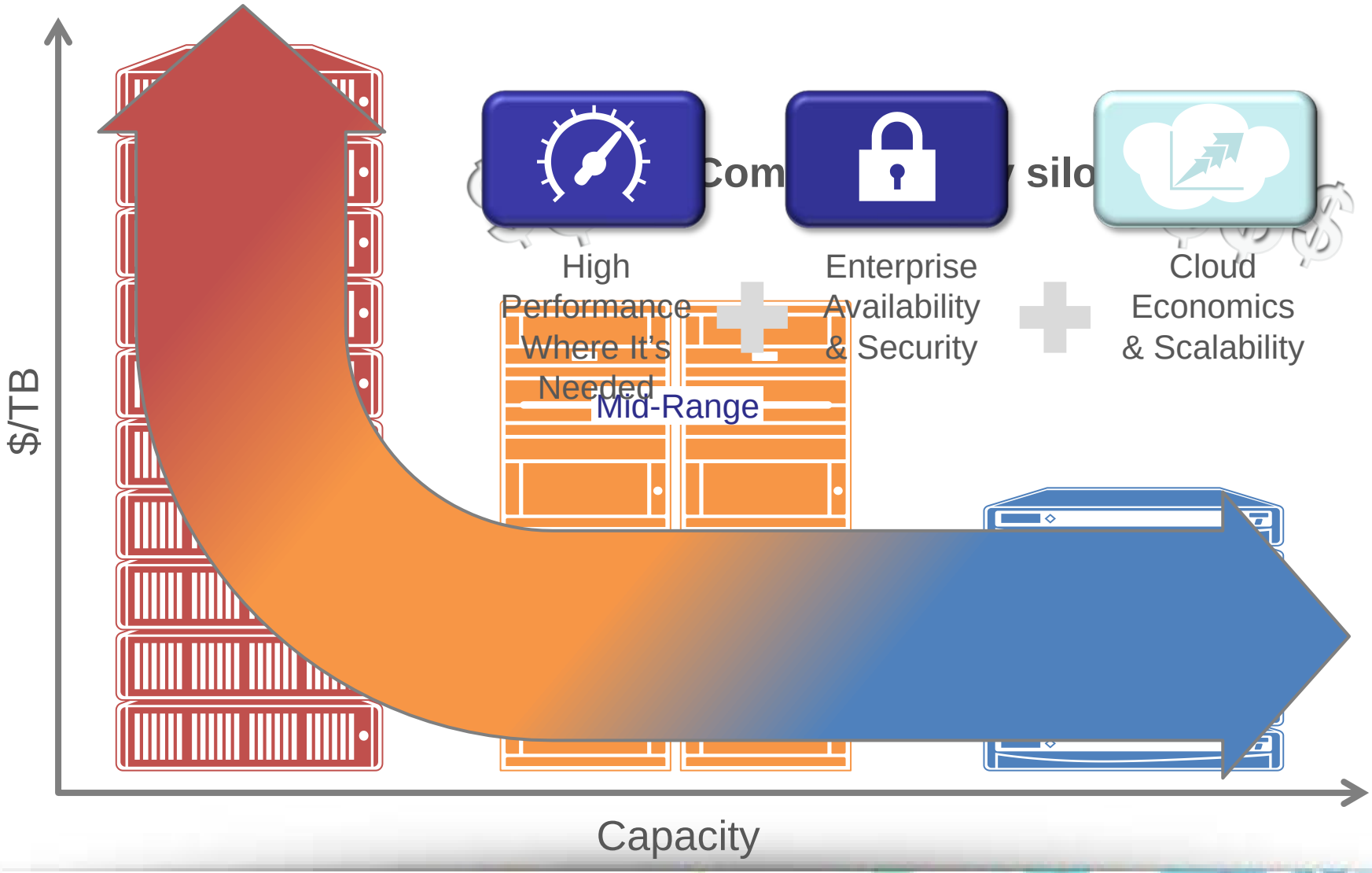
Why the storage you have is not the storage your data needs

Laz Vekiarides
ClearSky Data, Inc

Enterprise Storage Today

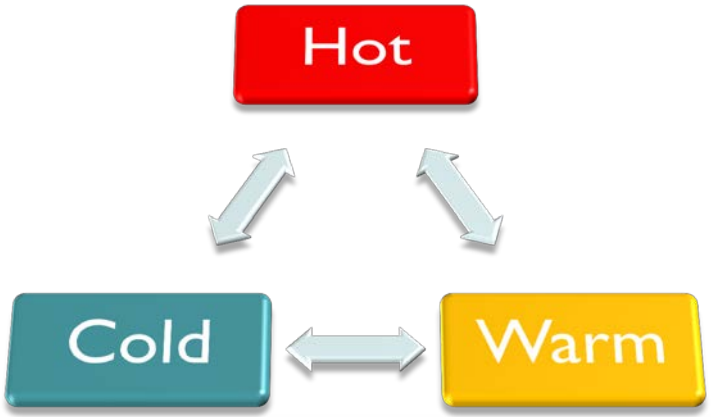


What Enterprises Really Want



Tiering is a bad answer

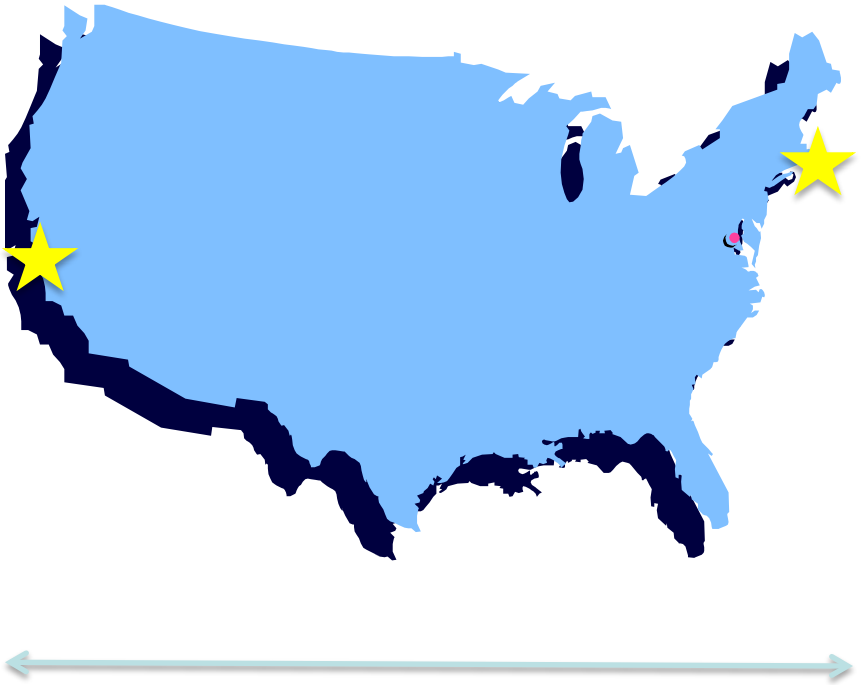
- Nothing remains static:
  - How fast does hot data cool?
  - How fast does it re-warm?
  - Is the overhead from this churn manageable?
- How can we use the cloud?



It’s the Latency, Stupid

(Apologies to Stuart Cheshire)

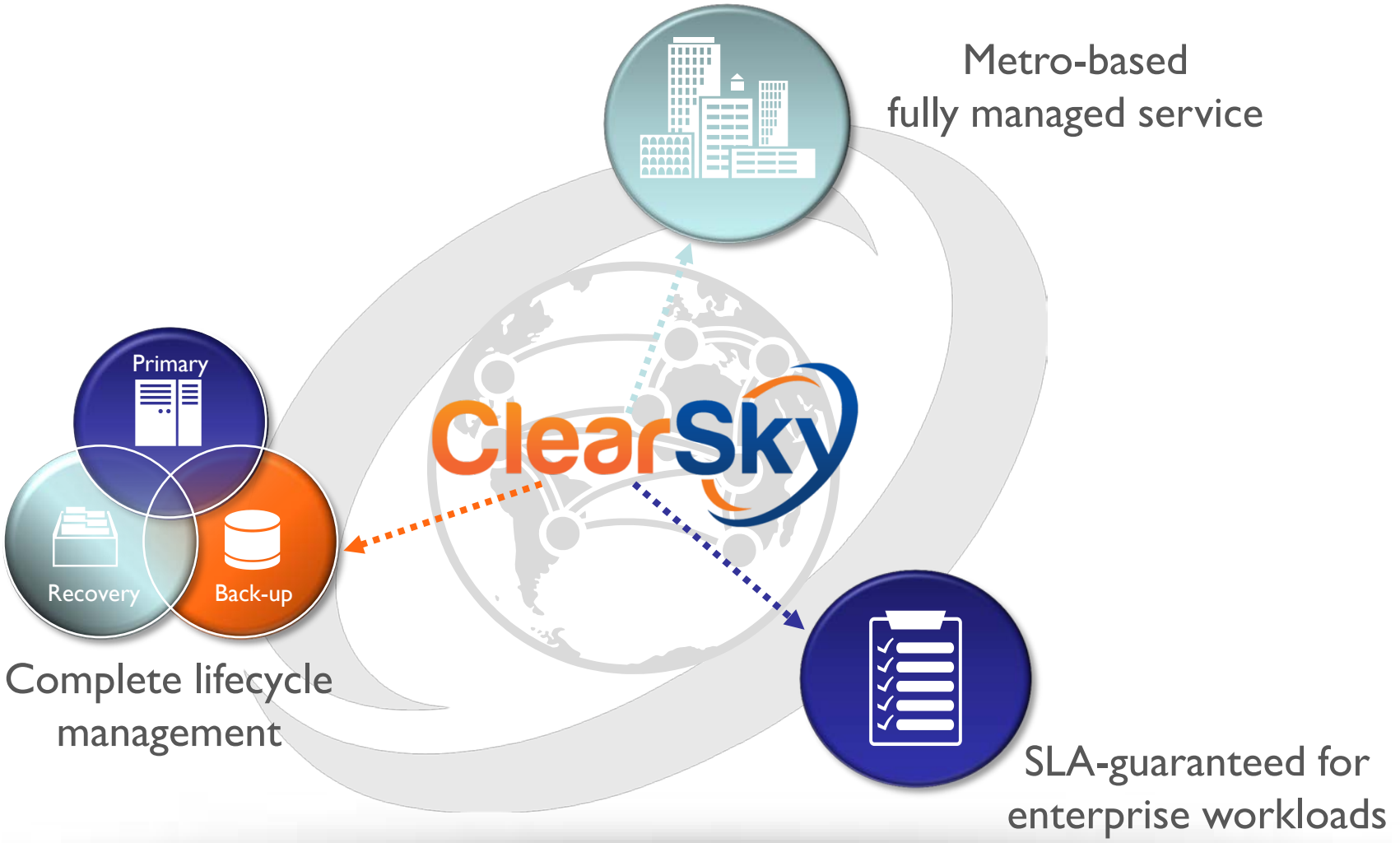
- Data travels at the speed of light
- Fast - but finite
  - $3x10^8$ meters per second
  - 186000 miles per second
- Example: Boston to San Francisco
  - 2740 miles
  - 29.4 milliseconds RT
- There are more delays
  - Light travels more slowly in fiber
  - Fiber-optic repeaters every few hundred miles
  - Switches, routers
  - Protocols, virtualization, etc.
- End result: ~70ms



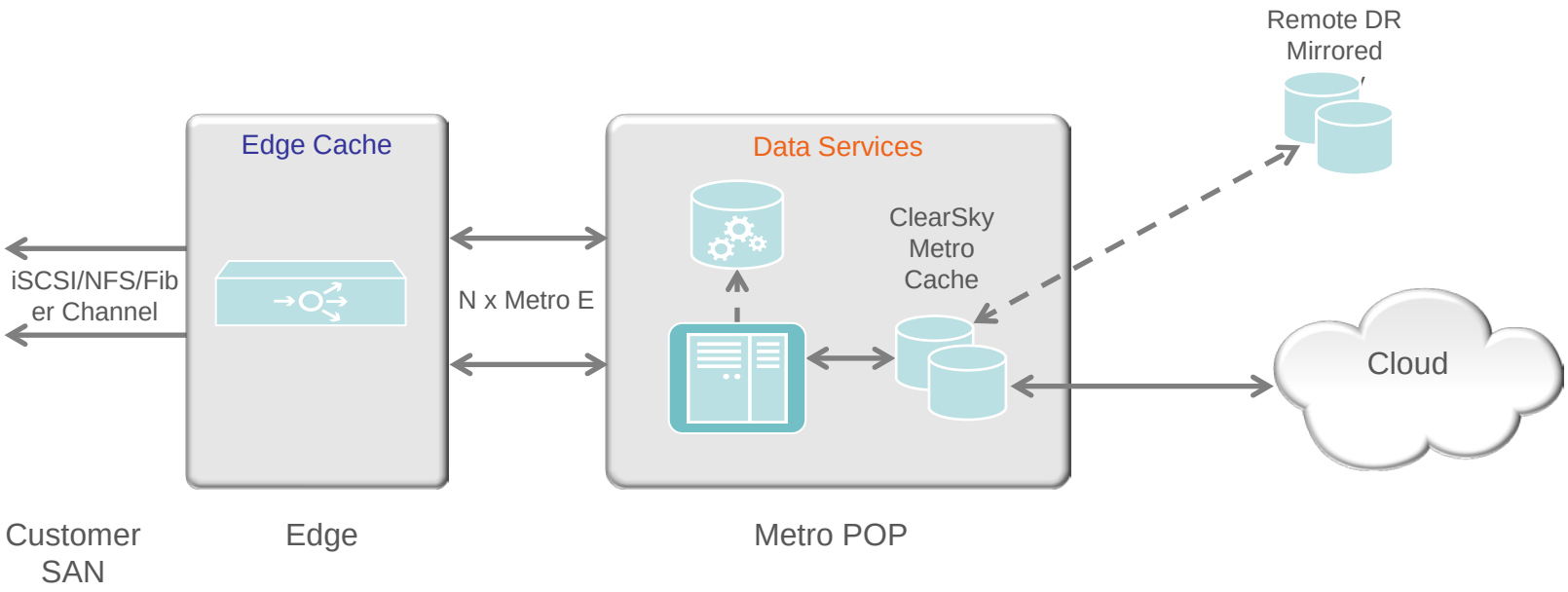
So, Where Exactly Is “The Cloud”?

- Amazon East is near Ashburn, VA
- West is in Northern California
- Boston is closest to East
- Best case numbers:
  - ~10ms round trip (private line)
  - From BOS MPOP via Direct Connect Ethernet
  - Does not include time to actually access the storage
- Worst case ~150ms (IP transit)

The ClearSky Solution:
A Global Storage Network

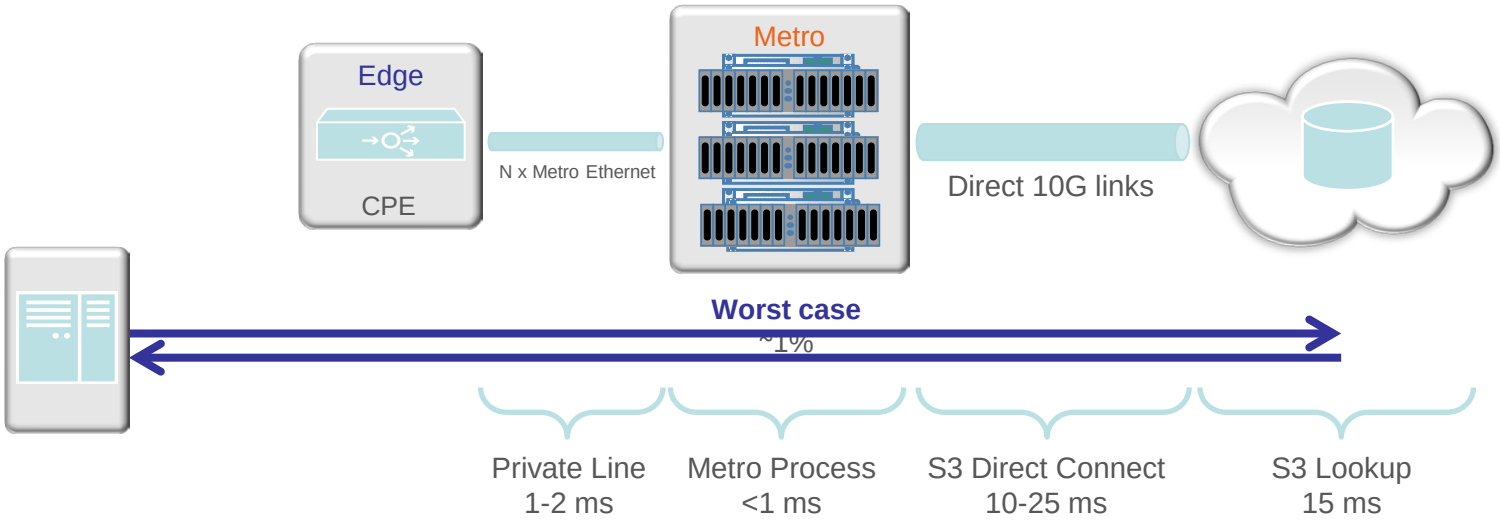


ClearSky: Geo-Distributed Data Caching



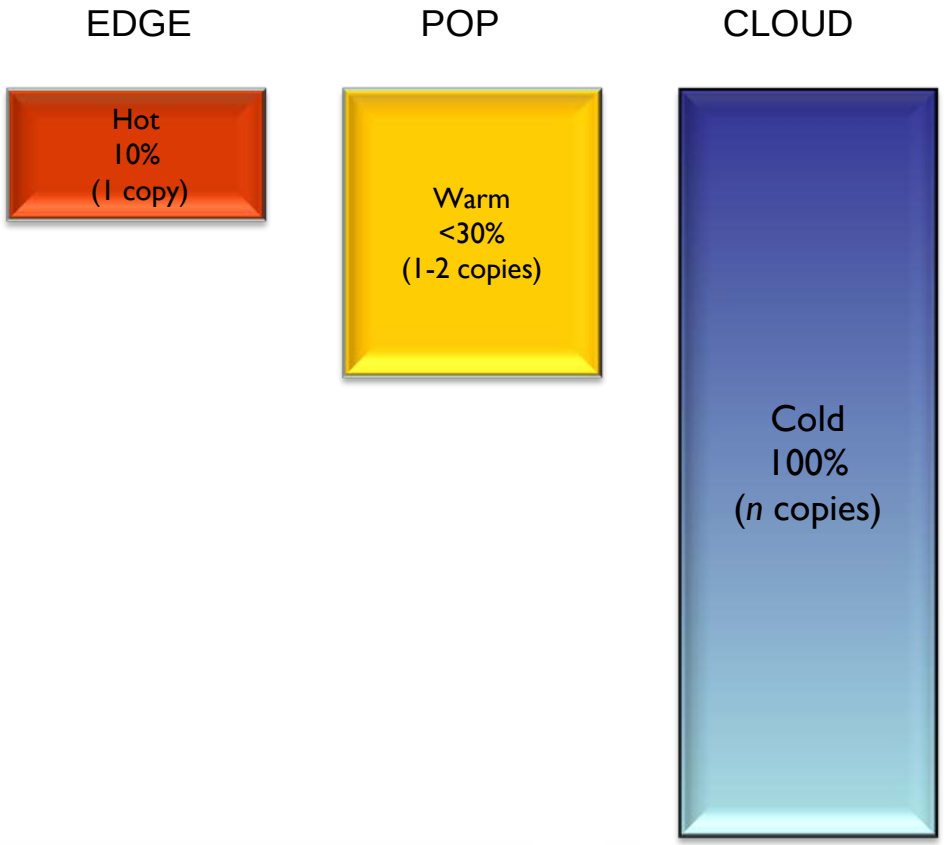
Latency Math

- Best case miss path ~25 ms
- Worst case <50 ms

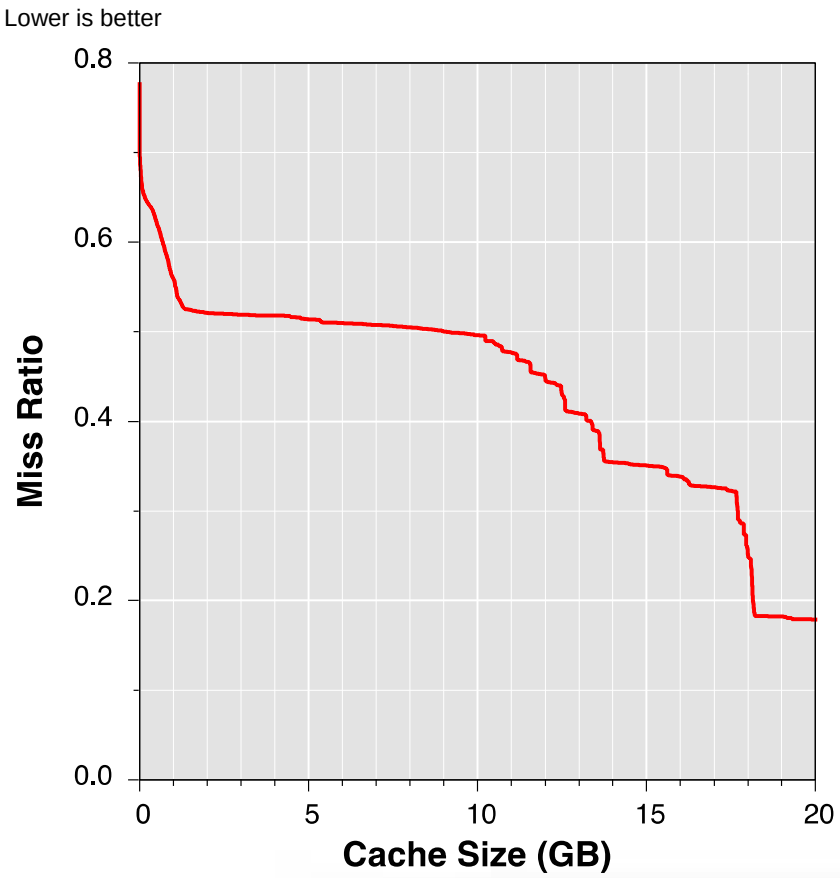


Current Space Management

- *All* managed data is migrated to a Cloud provider for durability
- All data is optimized
  - Deduplicated
  - Encrypted
- At least three tiers
  - Hot (local)
  - Warm/near-line, (POP, <2ms)
  - Cold, e.g. S3 (<20ms)
- Local appliance need only cache hot dataset (~10%)



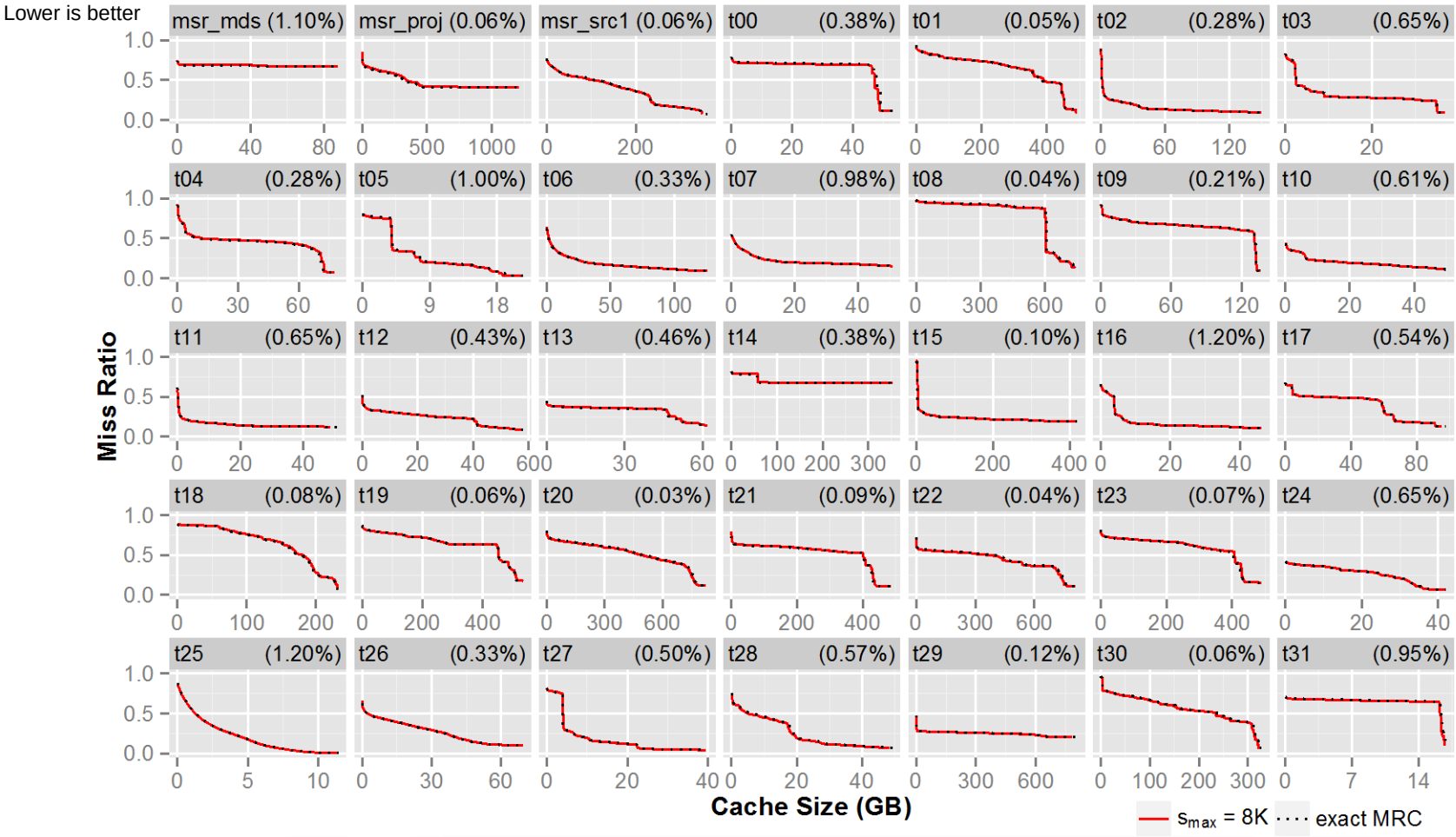
Modeling Cache Performance*



- Miss Ratio Curve (MRC)
  - Performance as *f* (size)
  - Working set knees
  - Inform allocation policy
- Reuse distance
  - Unique intervening blocks between use and reuse
  - LRU, stack algorithms

*Courtesy of Irfan Ahmad & CloudPhysics

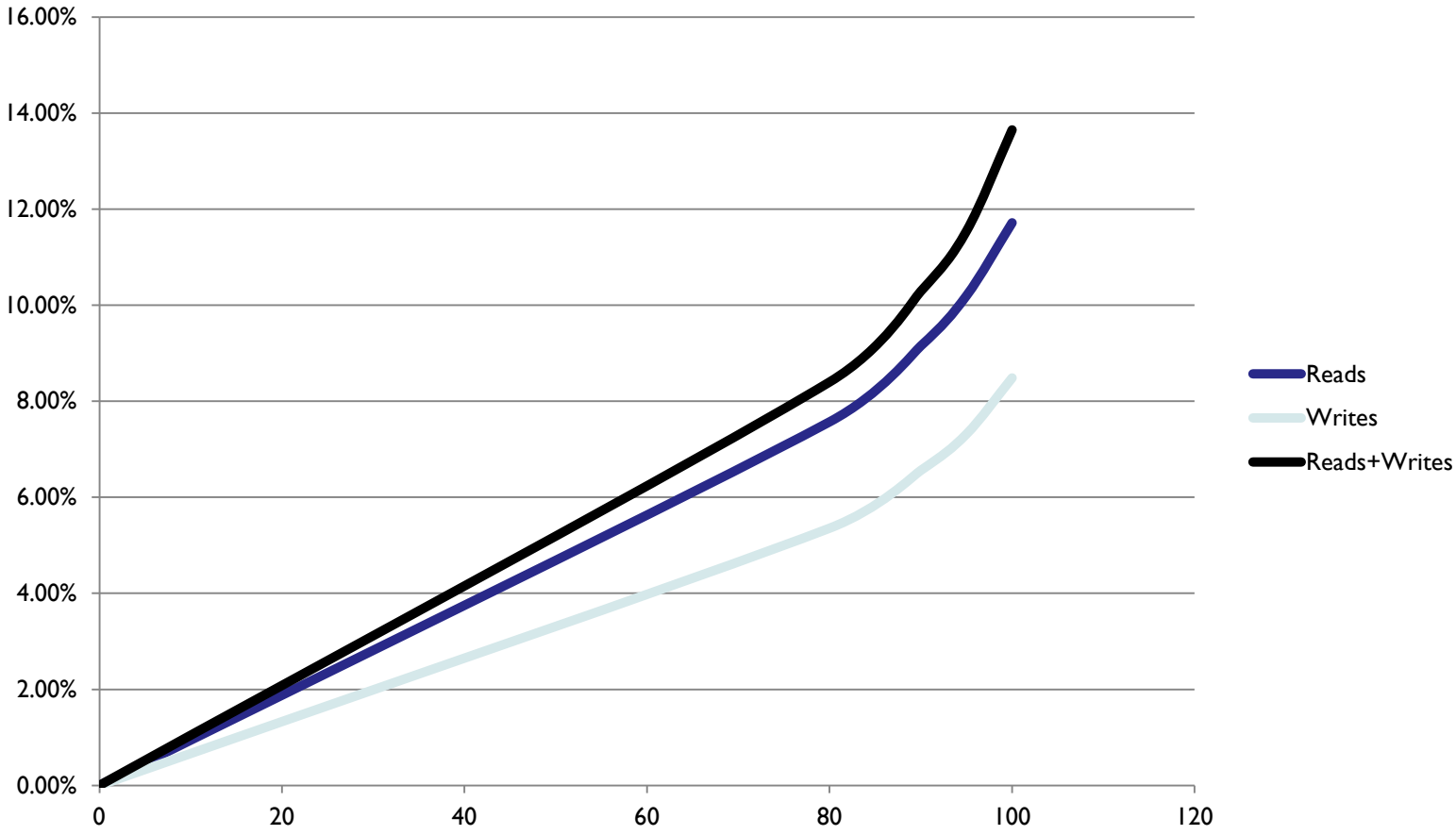
MRCs from Customer Workloads



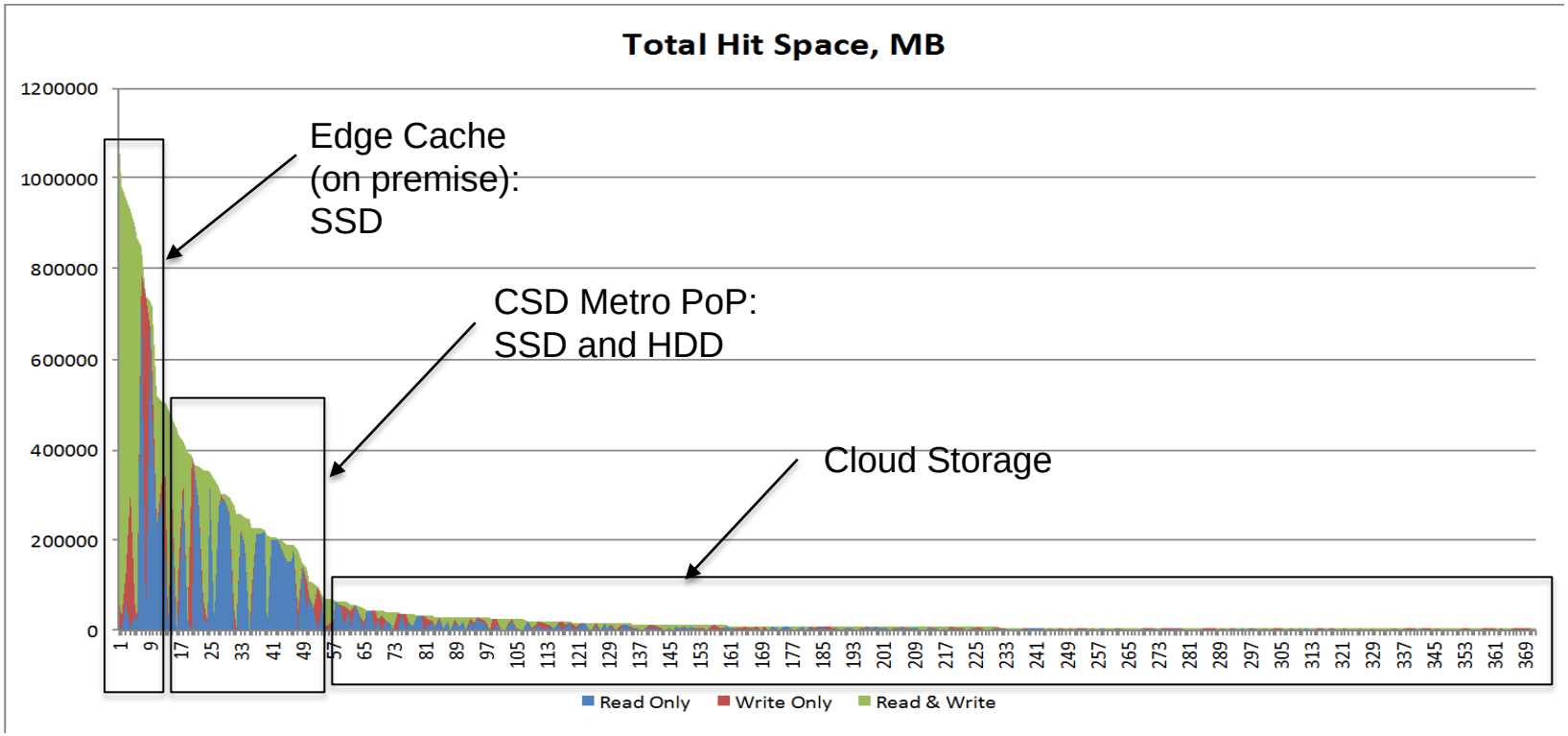
Customer Heat Map Data Collector

- Sizing tool built for VMware environments
- Collected 3-9 days per workload, most workloads analyzed for 7 days:
  - >1400 virtual disks on >800 VMs
  - Logical size of all workloads 27.4TB
  - Allocated space 18.9TB (68%)
  - Avg Read IOPS 5.2K, write IOPS 5.9K
- Performance & latency averages:
  - Read IO 36KB, write IO 110KB
  - Read latency 9.7ms, write latency 4.5ms

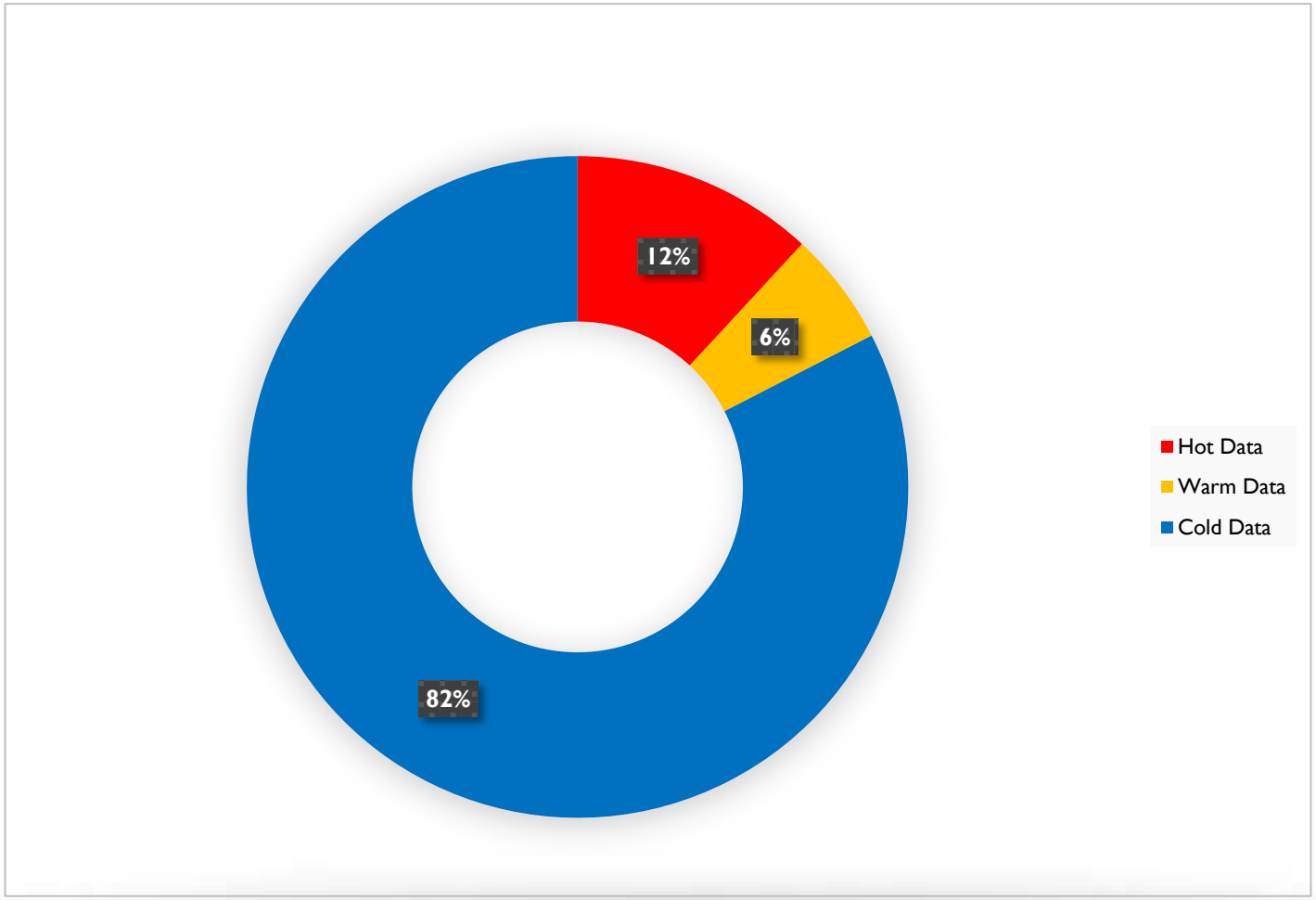
Miss Ratio Curves (>1400 virtual disks)



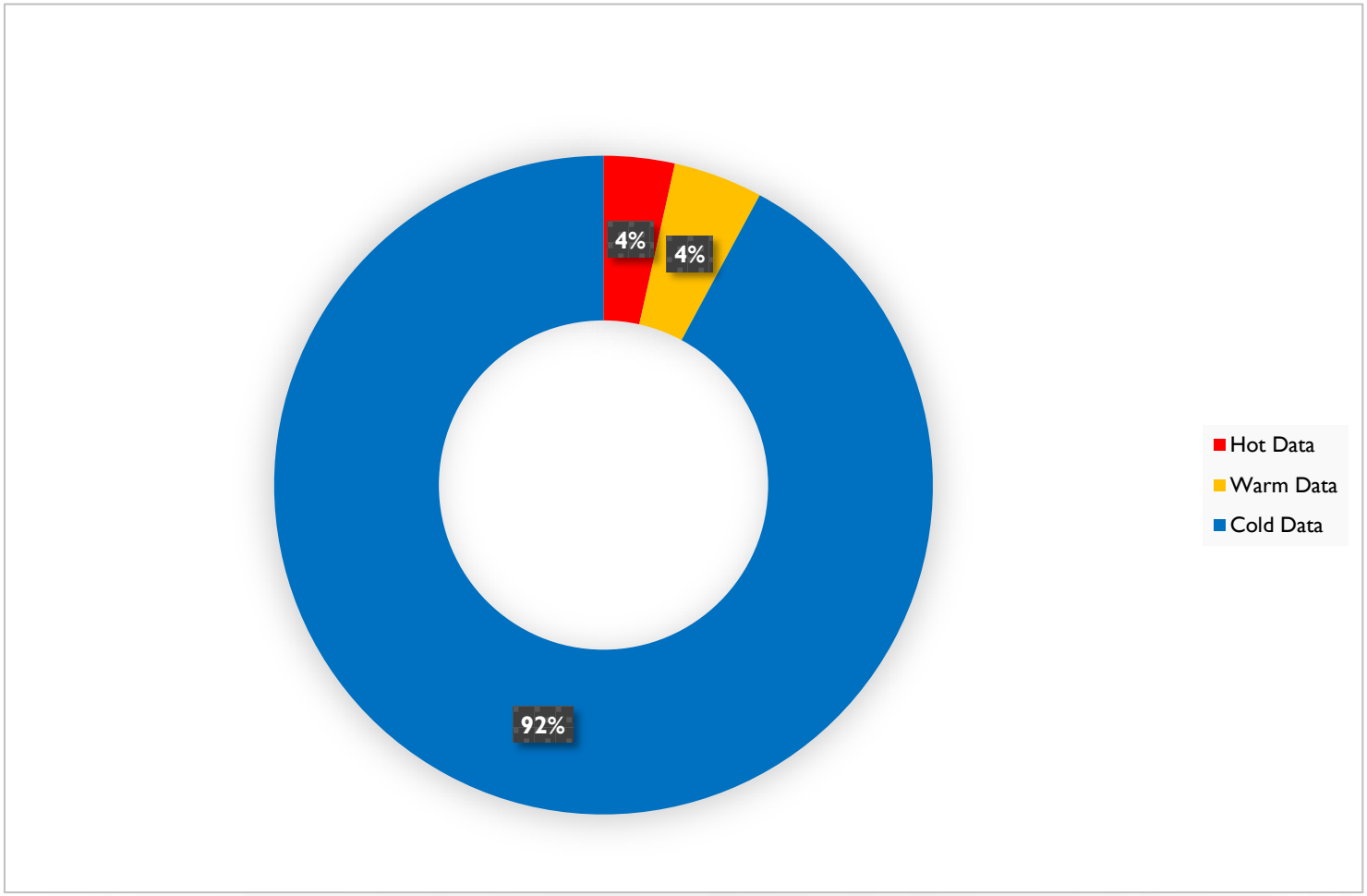
Importance of The Warm Tier



Heat Map Example: Production cluster



Example 2: Test / Dev / Beta / Xen



Yes. It Can Work

- Data access is very tiered
- Small amounts of flash can yield disproportionate performance benefits
- Variation of latencies must be bounded
  - Single tier cache in front of high latency storage cant work
  - Bounding network latency is as important as bounding media latency