From 1PB to 100PB

Architectural and Performance Considerations for Massive SmartStores

Ugur Tigli

CTO | MinIO

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or plans of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results may differ materially. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only, and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionalities described or to include any such feature or functionality in a future release.

Ugur Tigli

CTO | MinIO

splunk> .conf20

Agenda

1. Infrastructure as Software
2. Disaggregation as a Principle
3. The Box Challenge
4. Why MinIO
5. Introducing Active Active Replication
6. Takeaways

Infrastructure as Software

Simple Scales

Hyper-scalers recognized that software building blocks scale better than HW. More flexible, easier to combine/configure.

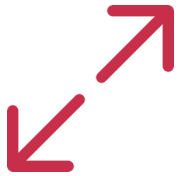
HW Becomes Commodity

Multi-PB infrastructure requires commodity HW. Software-defined solutions are superior to appliances. Failure is expected.

API + Automation Driven

Delivers elasticity, acts as a management force-multiplier and is inherently compostable.

SmartStore's Governing Principle at Box: Disaggregation

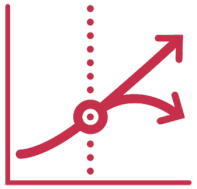


Scale

Indexers as caching + object storage as primary storage.

Scale compute + storage independently = performance.

Multiple TBs per day. Multiple PBs per Quarter.



Economics

"Classic" aggregated approach didn't make sense.

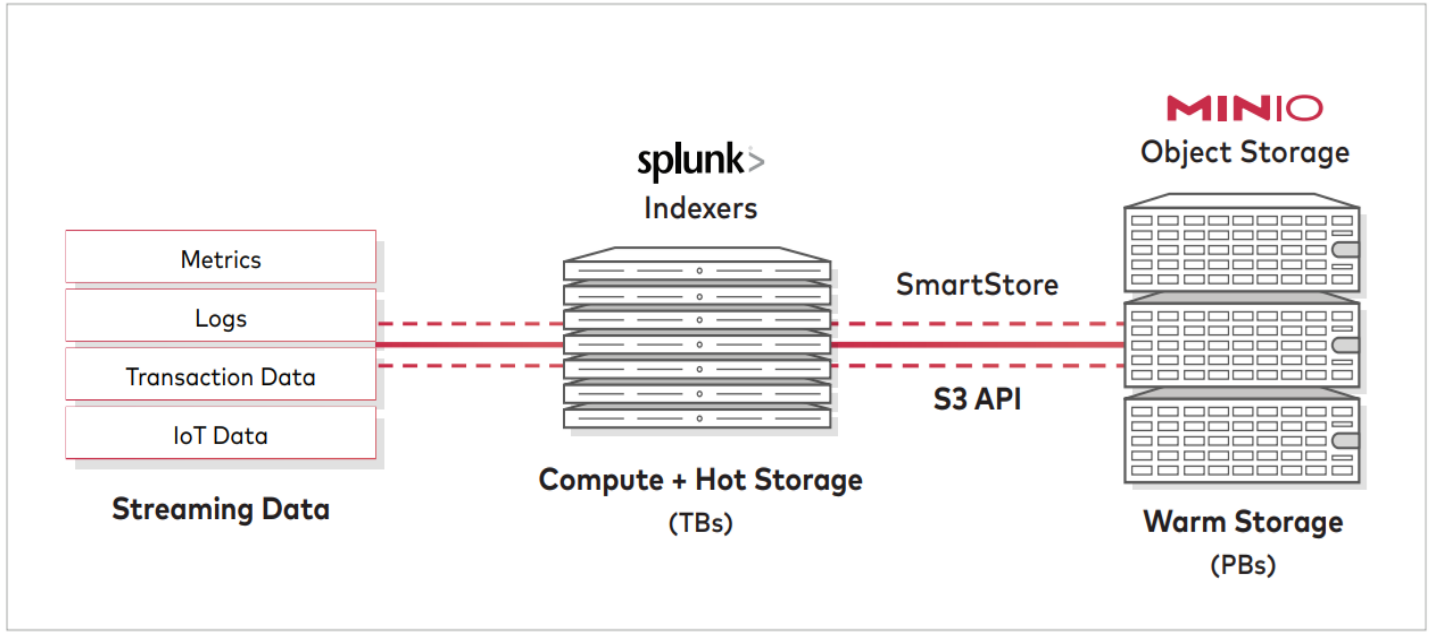
Public cloud egress costs are prohibitive past a few PBs.

Not archival – continuous reads and writes.

Performance

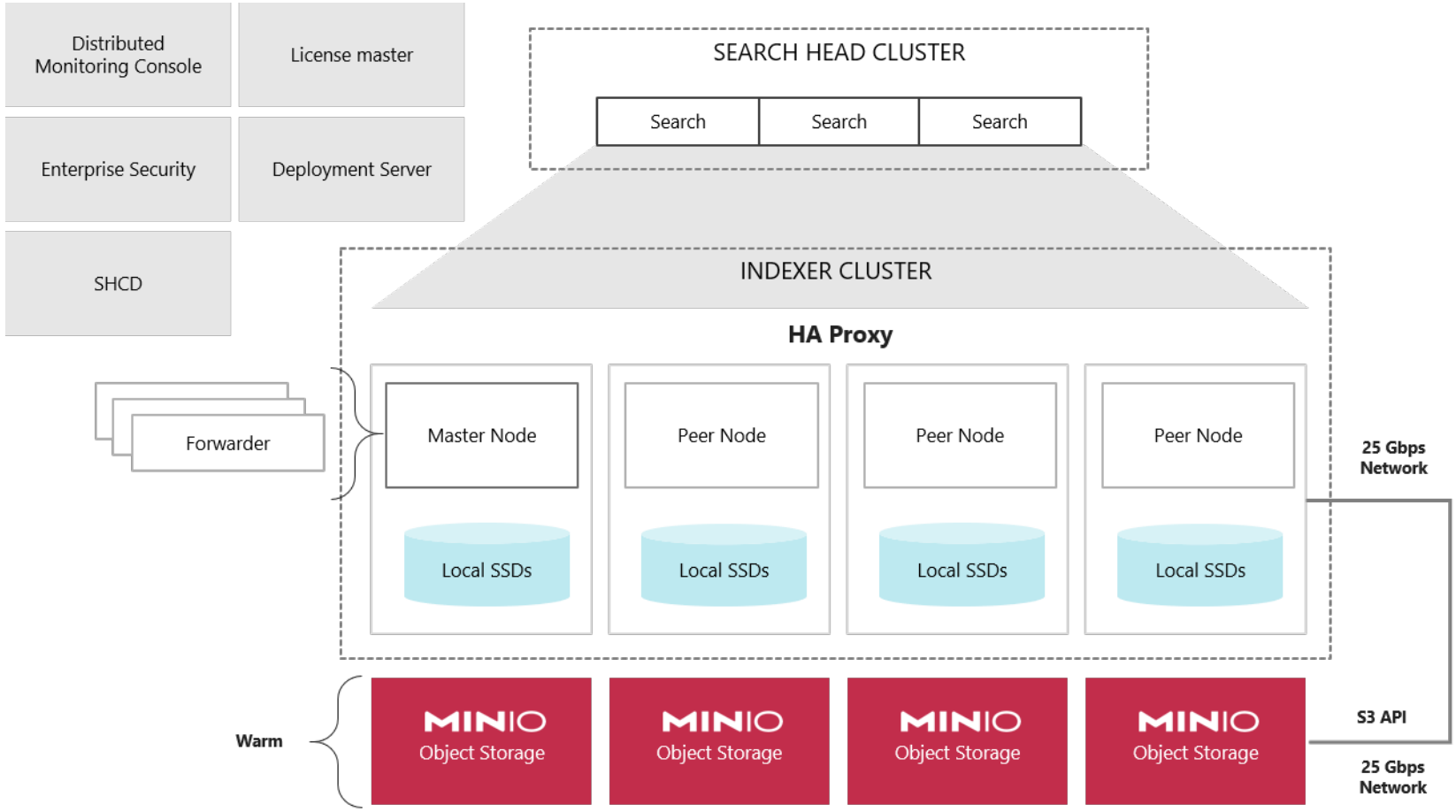
The ability to choose network and drive enable granular control against specific requirements.

Box + MinIO SmartStore Overview



Element	#
Raw Capacity	51.2 PiB
Usable Capacity	38.4 PiB
Erasure Code Settings	12/4
Nodes	16
Drives	200 x 16 TiB
Network	2x25 Gbp/s

Granular Architecture Detail



Why MinIO for Object Storage

Singular Focus

Do one thing + do it better than anyone else.

Our one thing is S3 compatible object storage.

Software Defined

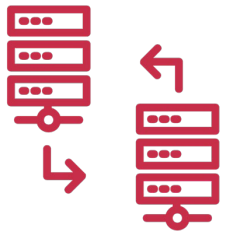
Starts with commodity HW, but also key for containerized instances.

Price per PB declines significantly at scale.

Performance

The world's fastest object store. Without massive throughput – object storage cannot serve as the primary storage tier. Significant implications for onprem vs. public cloud.

Why MinIO for Massive SmartStores



Replication Across Data Centers

The only solution for cross site, active-active replication. Continuously synchronize bucket changes. If the host fails, applications can seamlessly switch with minimal differences in state.

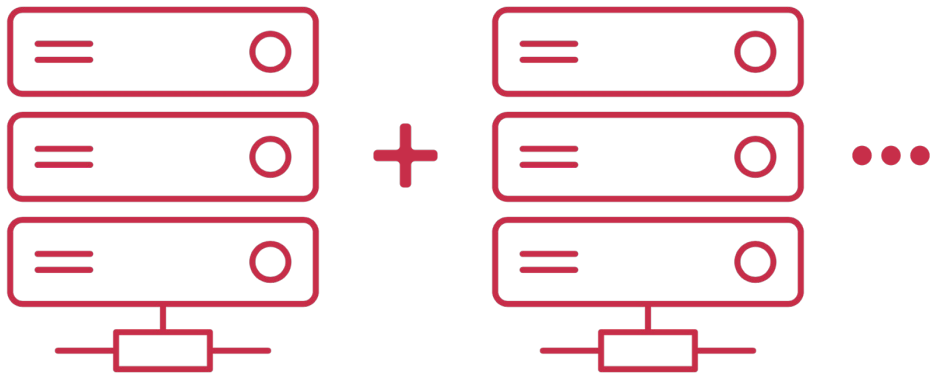
No Metadata Database

By writing metadata and objects atomically – a key inhibitor to scale is avoided. Particularly problematic with small objects.

Sidekick

By attaching a lightweight loadbalancer as a sidecar to each indexer you can eliminate a centralized loadbalancer bottleneck and DNS failover management.

Expanding Capacity

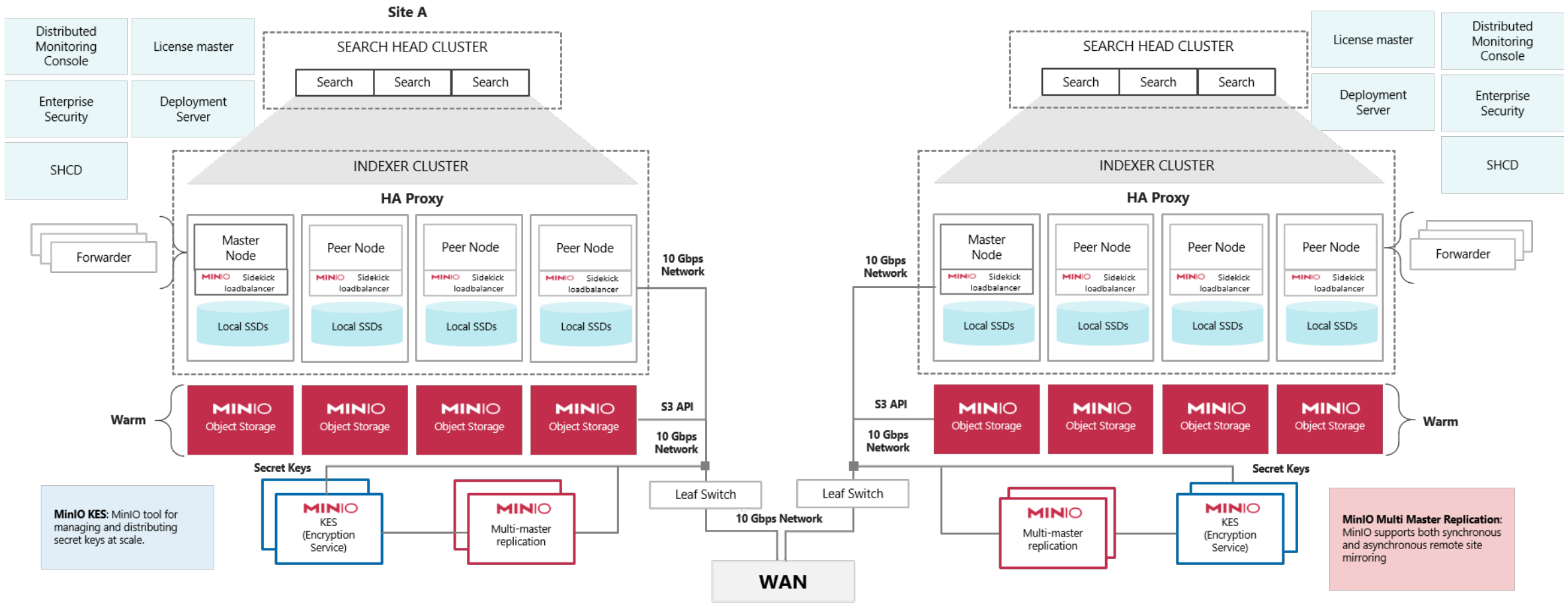


- Expansion is done by adding new zones
- Zones are simply set of new servers
- Zones allow rack awareness and heterogeneous mix of hardware
- Buckets expand to all zones
- This architecture avoids rebalancing requirements
- No limit to number of zones or servers within a zone
- Non-disruptive upgrades and updates

```
minio server http://host{1...4}/export{1...26} http://host{5...n}/export{1...26}
```

Zone 1 Zone 2

Optimal Active Active Replication



*Released by MinIO, currently undergoing Splunk testing.

splunk> .conf20

Takeaways

- Think big with your SmartStore because more data = more insight
  - You can still start small…
- Think simple with your SmartStore because simple things scale
- Think about resilience because scale magnifies risk
- Software-defined is the way of the webscalers – it should be your way too

Thank You

Please provide feedback via the

SESSION SURVEY