Home For Gypsies – Storage for NoSQL Databases

Atish Kathpal, NetApp

NetApp

Agenda

1) Introduction on NoSQL

- Master-less and Master-slave architectures
- Data management provided by NoSQL DBs
- How is Shared Storage relevant?

2) Backup and Restore for NoSQL DBs

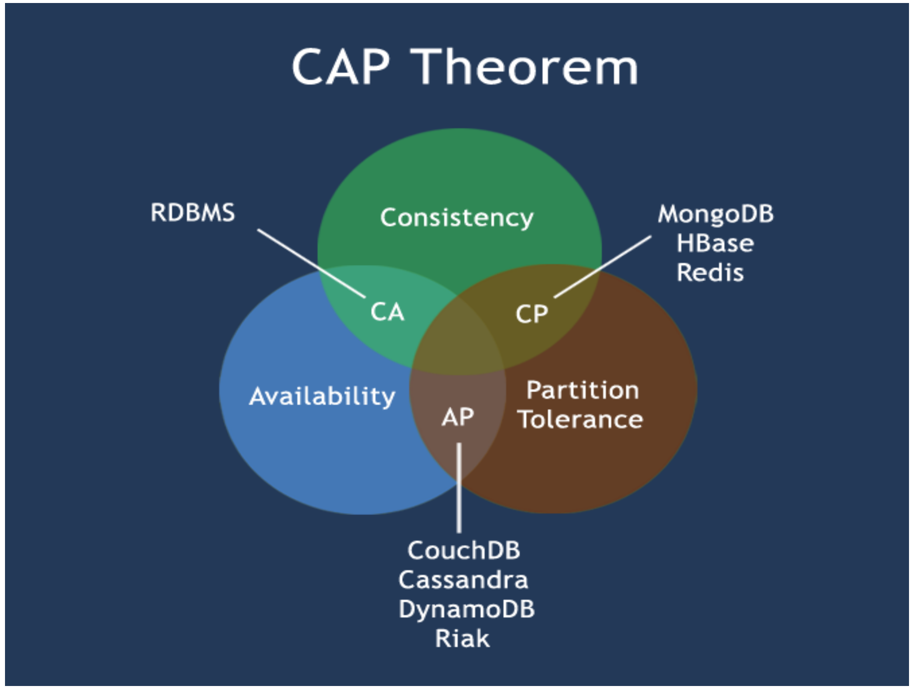
- Opportunity to leverage shared storage features
- Challenges

3) Conclusions

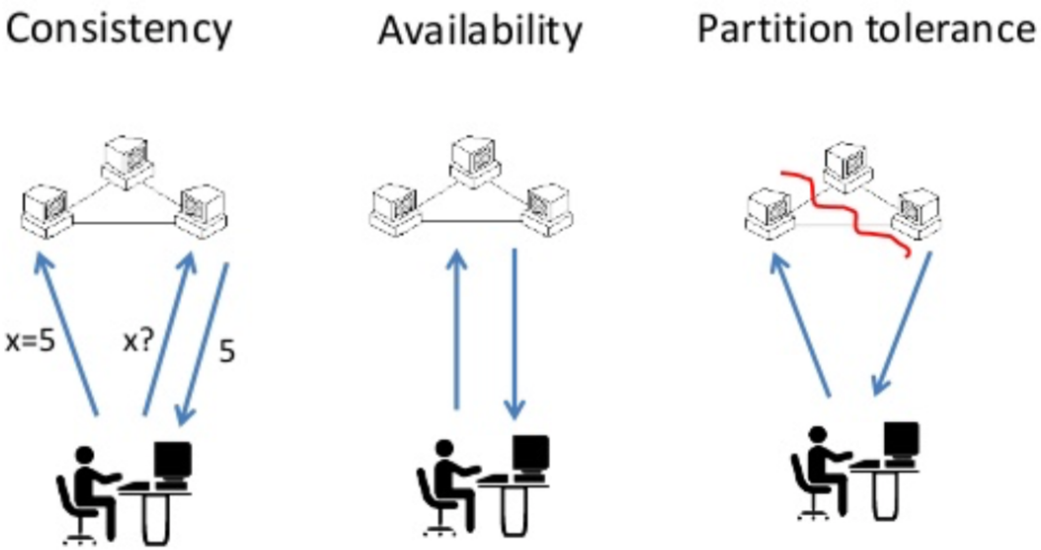
NoSQL Databases

Overview

- Different from RDBMS
  - Tunable consistency semantics
  - Vertical v/s Horizontal scale
- When scale and data availability more important than consistency
  - Big data, web-scale apps - IoT, Mobile, Analytics
  - Trade-offs: CAP theorem [1]
- Open source, commodity nodes, DAS



Source: http://www.abramsimon.com/cap-theorem/



https://www.slideshare.net/GrishaWeintraub/cap-28353551

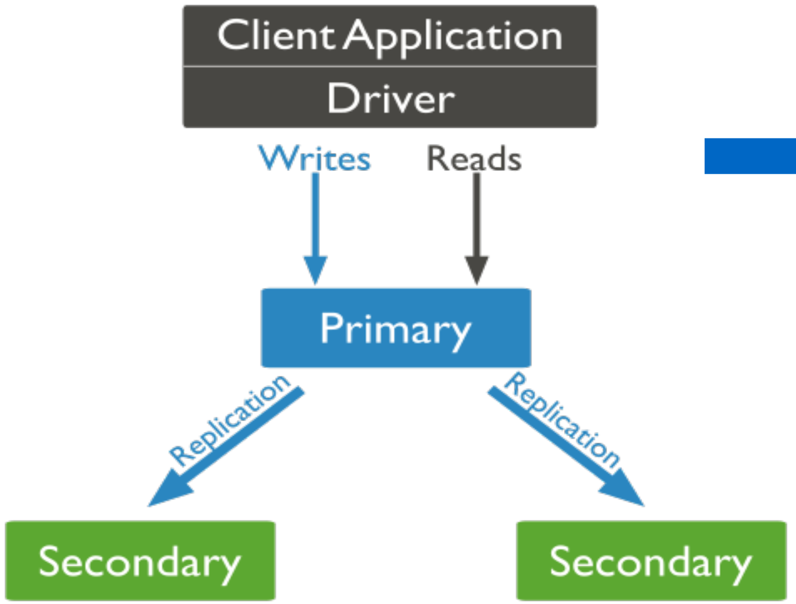
[1] Brewer, Eric. "A certain freedom: thoughts on the CAP theorem." *Proceedings of the 29th ACM SIGACT-SIGOPS symposium on Principles of distributed computing*. ACM, 2010.

NetApp

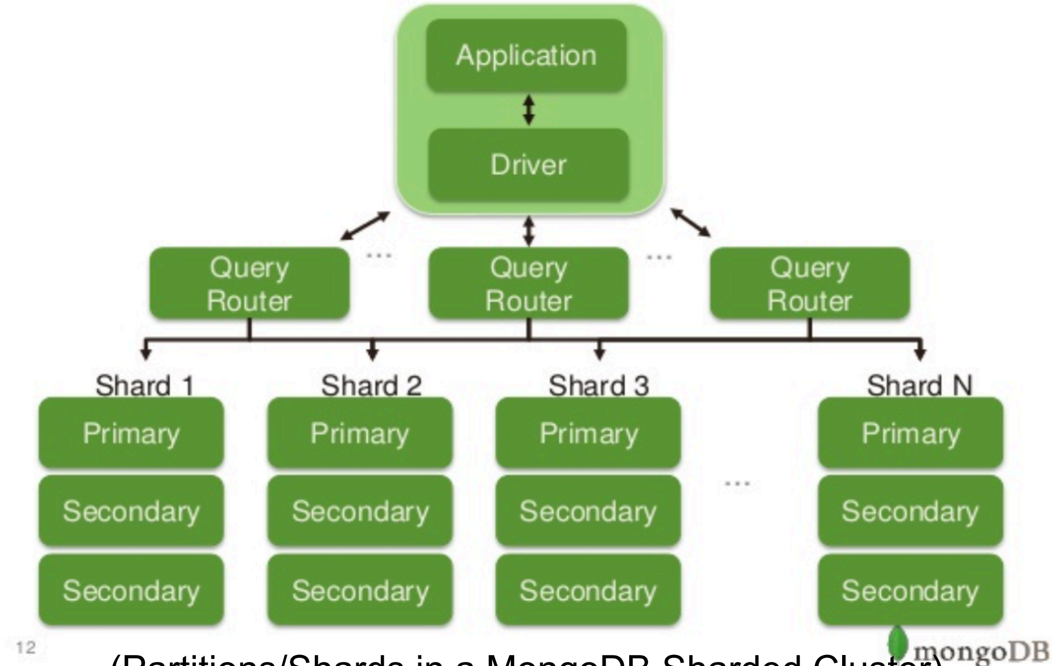
Master-Slave Databases

NoSQL DB classification

- All writes to a partition, first written to the master node
  - Thus subsequent reads involving the Primary are always consistent
  - Loss of primary node leads to shard/partition-unavailability until new leader is elected
  - MongoDB, Redis fall in this category



Single shard



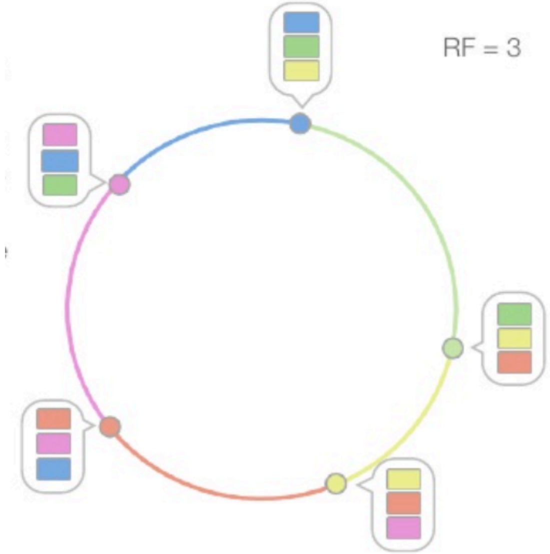
(Partitions/Shards in a MongoDB Sharded Cluster)

https://www.slideshare.net/mongodb/sharding-v-final

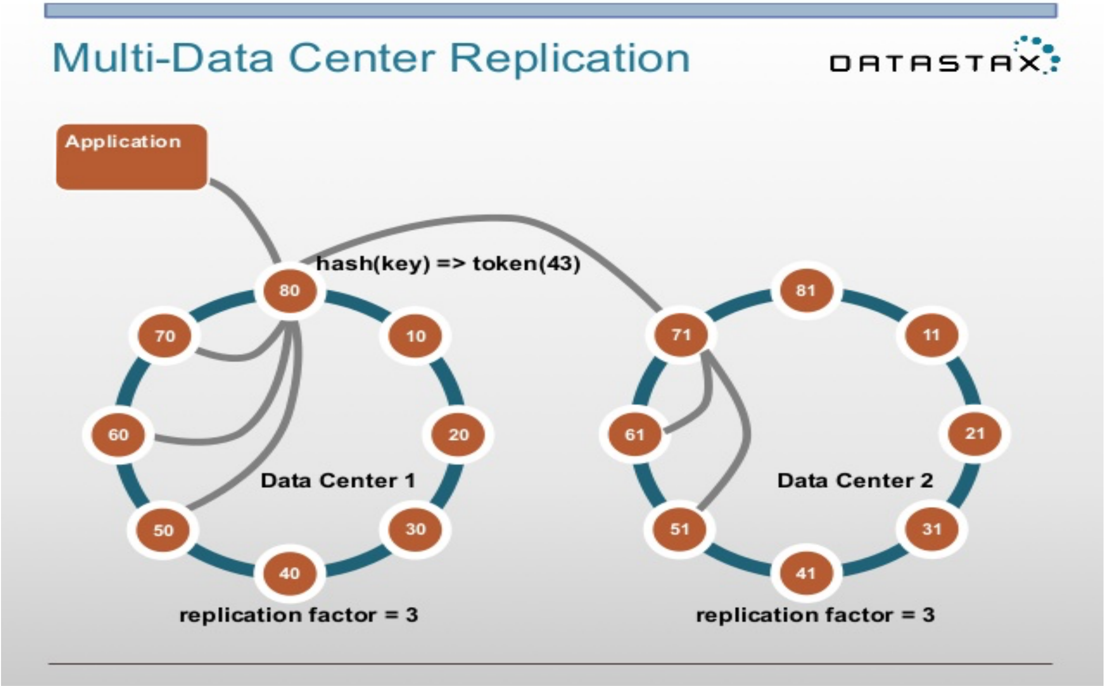
Master-Less Databases

NoSQL DB classification

- Data is scattered across nodes using consistent hashing techniques
  - Writes streamed simultaneously to all nodes that the data hashes into
  - Eventual consistency: Unavailability of a destination node does not lead to write-failure, data is *eventually* replicated to the node when it becomes available, or gets hashed into a peer node
  - Examples: Cassandra, CouchBase



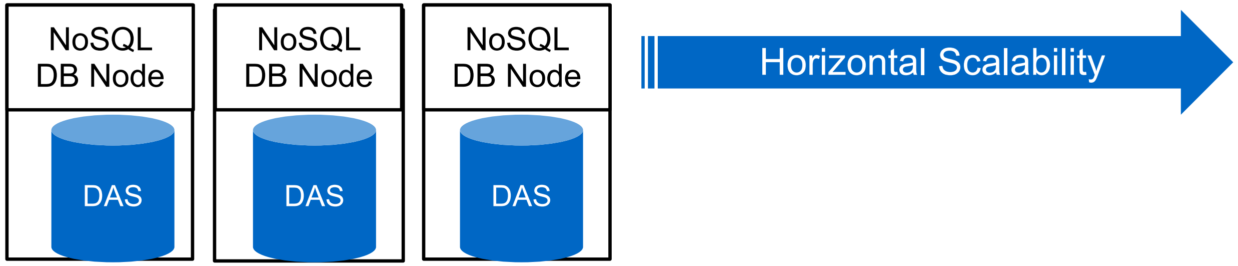
Data spread across nodes

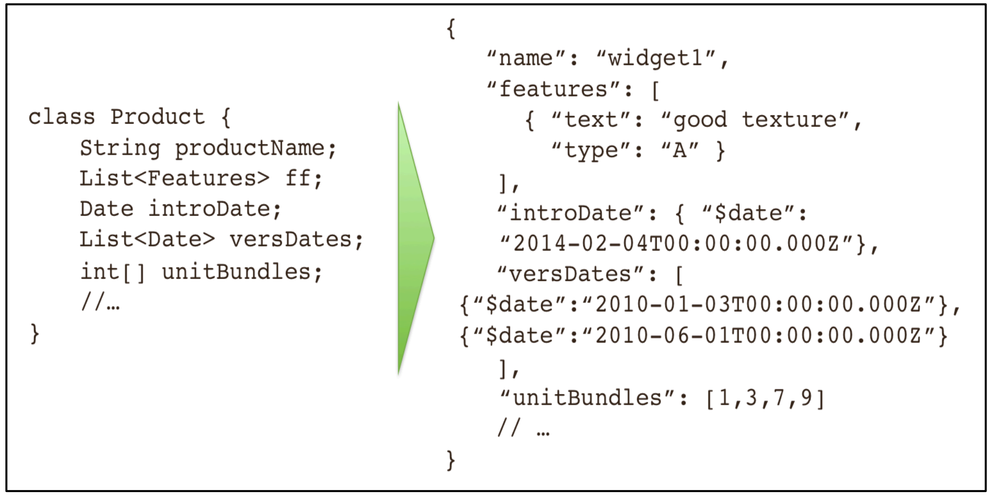


https://www.slideshare.net/pcmanus/cassandra-4345401, https://www.slideshare.net/planetcassandra/solr-cassandra-searching-cassandra-with-datastax-enterprise,

NoSQL DBs – Built-in Data management

- Performance through horizontal scale-out
  - Commodity compute nodes with DAS
- Replication: high-availability and fault-tolerance
- Cluster Management, across data centers
- Inline compression, Encryption
- Developer friendliness
  - Support for different data formats and schemas
  - Integrations with analytics engines like Hadoop and Spark





In-memory objects can easily map into JSON-documents => flexible schema

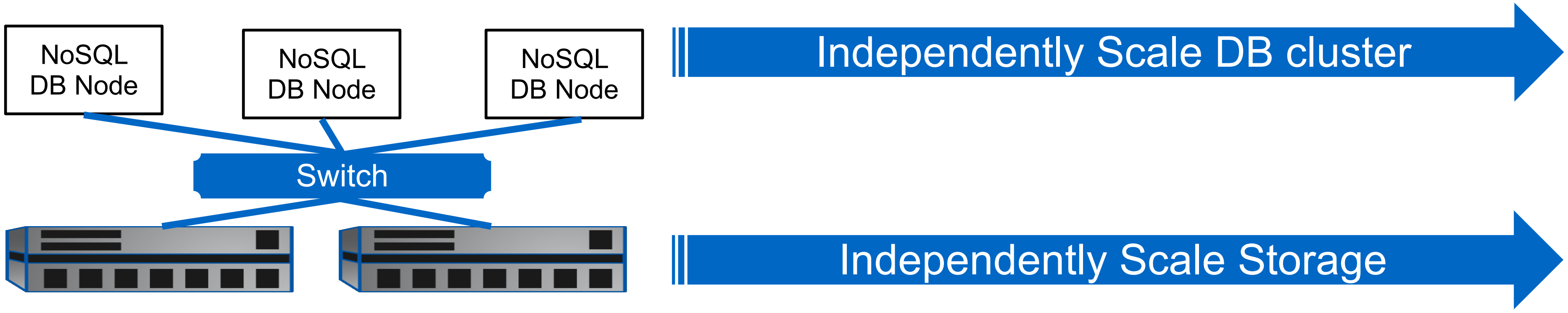
What additional value can shared storage bring in?

[Image source: https://www.mongodb.com](https://www.mongodb.com)

NetApp

Shared-Storage value-adds

- Independent *scaling of compute and storage*



- Consolidation of storage implies *easier storage resource management*
  - Reduced cluster management costs

Shared-Storage key value-adds (continued)

- All Flash Arrays
  - Performance at *reduced cluster size*
  - Can run mixed workloads without performance impact

Can support mixed workloads: NoSQL clusters, RDBMS, Hadoop, Data warehousing etc.

Legacy HDD storage arrays

- Less space
- Less power
- Lower TCO
- Predictable performance

All flash storage array

So shared storage can provide value as primary tier, however what about data protection and secondary storage?

NetApp

Backup and Restore

Relevance of Shared Storage

Why Backup/Restore NoSQL DBs?

Customers are directly ingesting critical into NoSQL

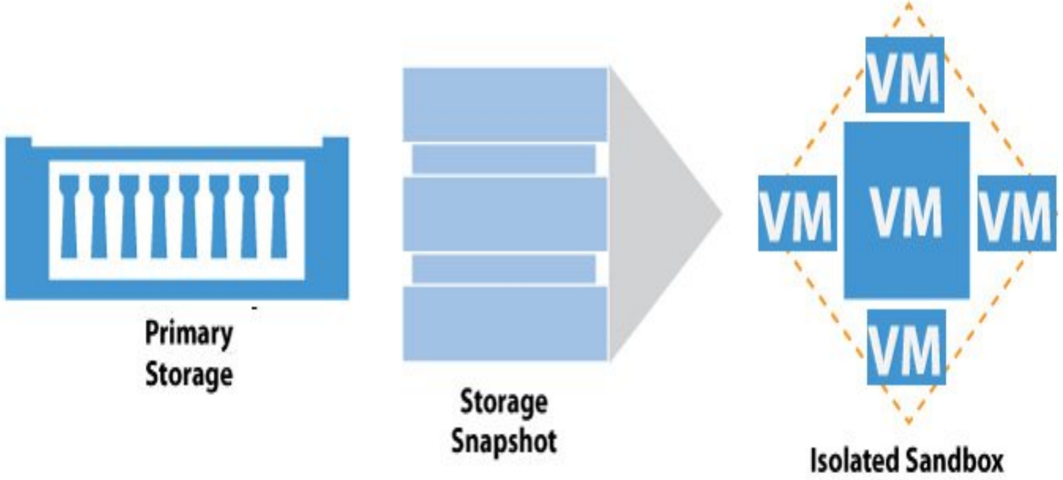
Security breach are on the rise e.g. **ransomware attacks** on MongoDB [2] and recent **WannaCrypt** exploits

“Fat-finger” errors eventually propagate to replicas

Ransomware

Sandbox deployments for **test/dev**

- Bring up shadow clusters of different cardinality (from production cluster snapshots)



Compliance and regulatory requirements

IDC, 2016 report [3] lists data-protection and retention as one of the top infrastructural requirements for NoSQL

[1] http://thehackernews.com/2017/01/secure-mongodb-database.html [2] Nadkarni A., Polyglot Persistence: Insights on NoSQL Adoption and the Resulting Impact on Infrastructure. IDC. 2016 Feb.

Images: https://arstechnica.com/security/2017/01/more-than-10000-online-databases-taken-hostage-by-ransomware-attackers/, https://www.veeam.com/netapp-snapshot-snapvault-snapmirror-integration.html

Existing Open-source Utilities - Limitations

- Utilities like mongodump, mongorestore are in-adequate
  - **Operates on per-node basis**
  - **Copy based** solution (expensive for TBs of data)
  - **High restore times** due to copy-in
- Cassandra backup utilities
  - Keeps a hard link of data files on disk (**storage overhead**)
  - Requires **expensive repair steps** upon restore
- Such utilities **require separate automation** to take cluster wide backups
  - Suffer from **failure scenarios**

Shared storage to the rescue? – Address pain points of copy-based backup and repair after restore

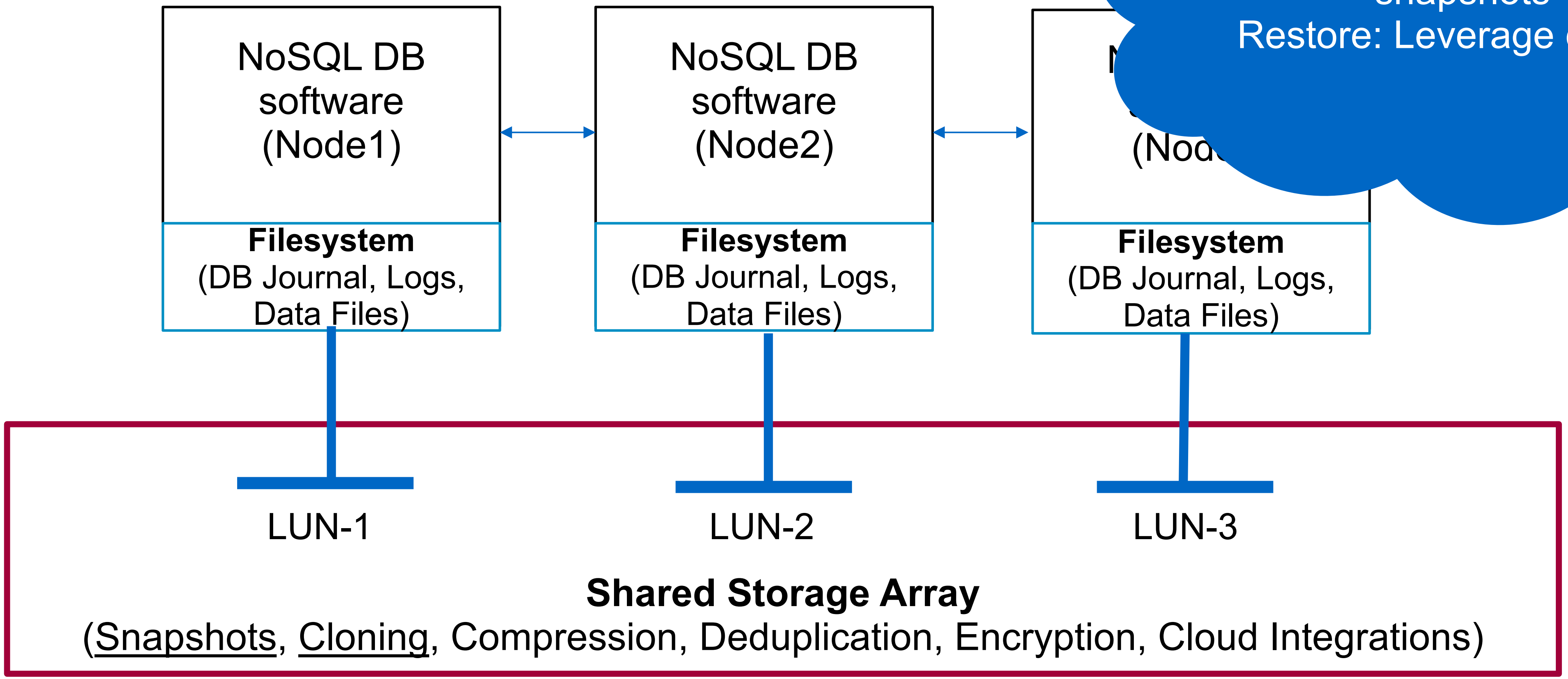
Image source: http://www.mongodb.com

NetApp

NoSQL DBs on Shared Storage

High-level, conceptual deployment architecture

Ideas:
Backup: Leverage storage snapshots
Restore: Leverage cloning



NetApp

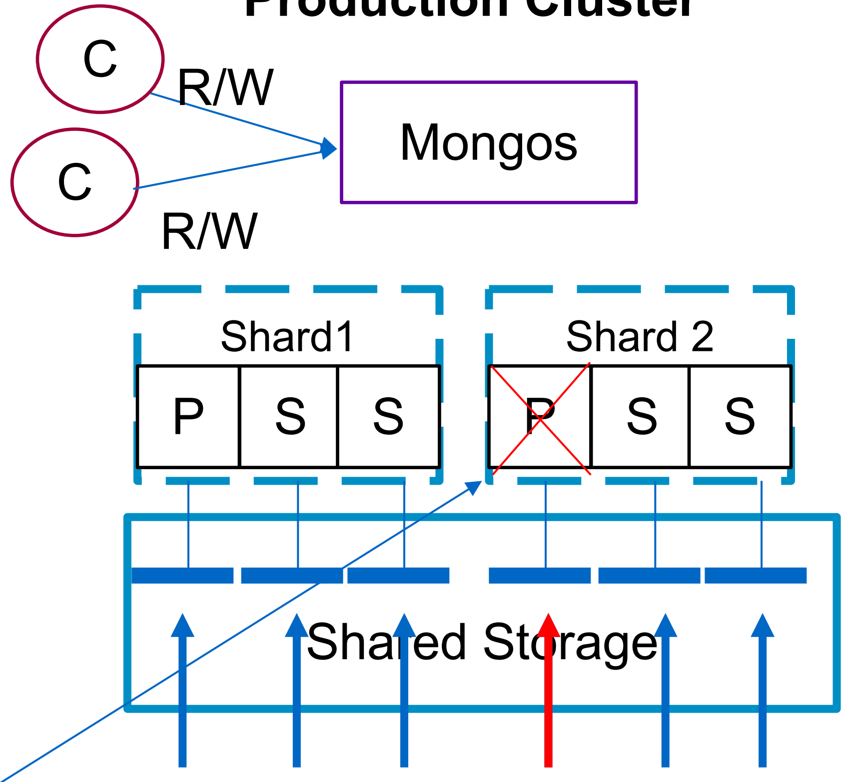
NoSQL Data Protection – Challenges

Master-Slave Databases

Challenges:

- Storage efficiency
  - Redundant data captured in backup (replicas)
  - **Replicas do not dedupe**
    - Unique ids per document per node
    - Compression, encryption
- Fault tolerance
  - Backup may capture unstable state of cluster
  - Can lead to higher RTO, due to new leader election during restore?

Production Cluster



Take snapshot storage snapshots,
Clone them to secondary storage

1. When primary fails, new leader is elected
2. Non-quorum data in failed primary is rolled back

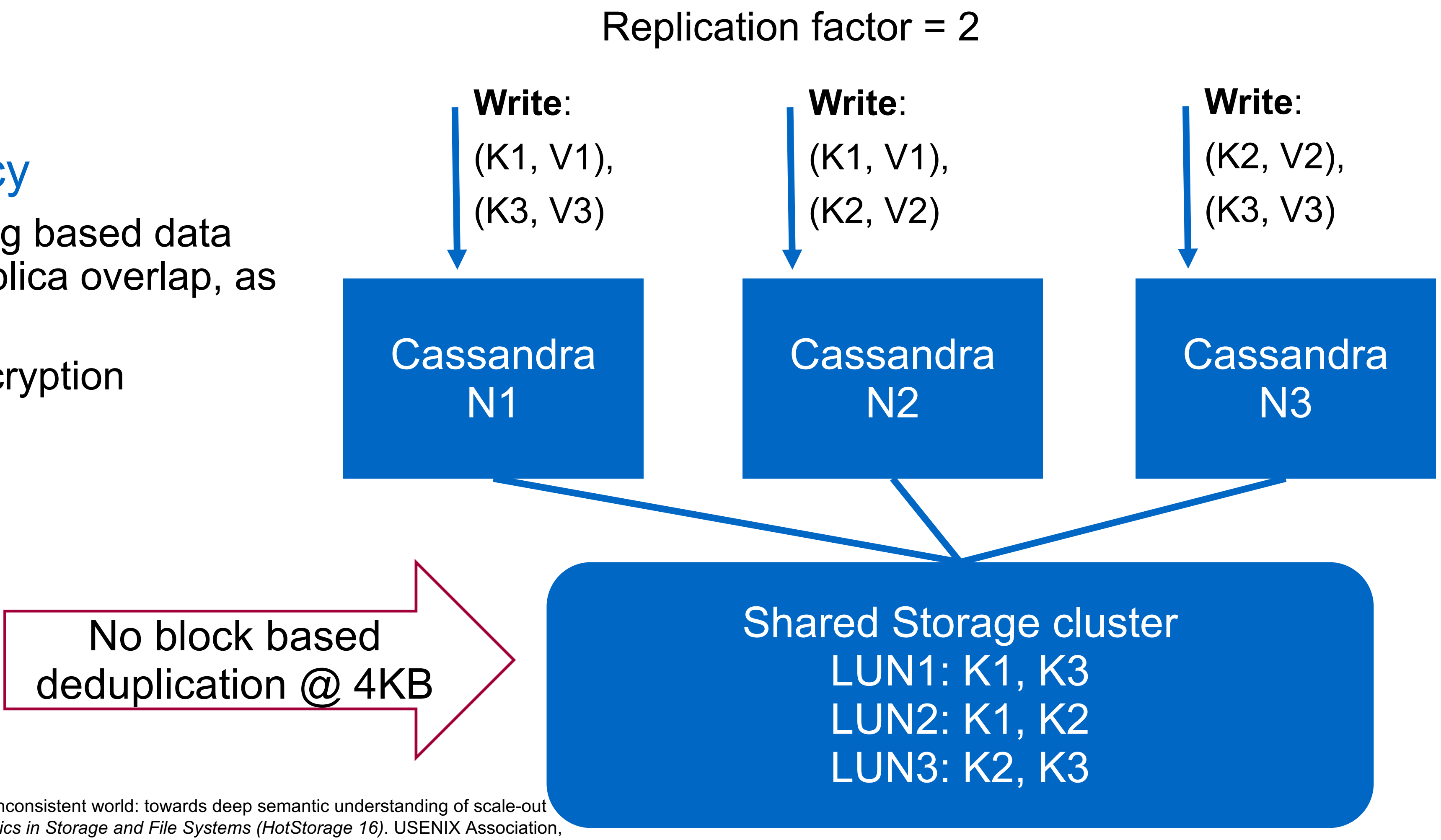
NetApp

NoSQL Data Protection – Challenges

Master-Less Databases

Challenges:

- Storage efficiency
  - Consistent hashing based data layout leads to replica overlap, as shown
  - Compression, encryption



Refer: Carvalho, Neville, et al. "Finding consistency in an inconsistent world: towards deep semantic understanding of scale-out distributed databases." *8th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 16)*. USENIX Association, 2016.

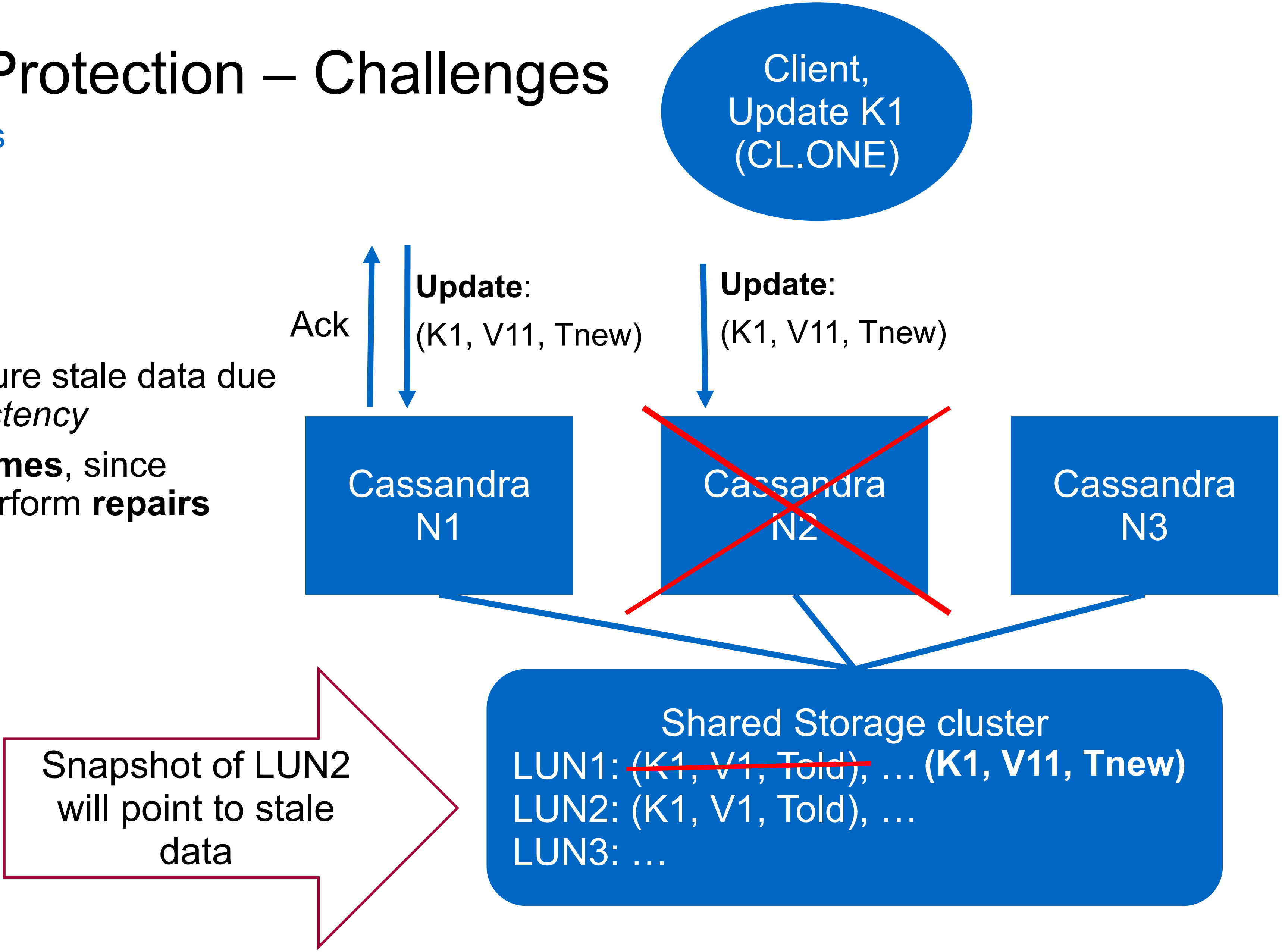
NetApp

NoSQL Data Protection – Challenges

Master-Less Databases

Challenges:

- Fault tolerance
  - Backup may capture stale data due to *eventual consistency*
  - **Higher restore times**, since Cassandra will perform **repairs during restore**



NoSQL DB Protection – Challenge

NoSQL backup/restore challenges

- Flexible topology restores
  - Production cluster topology may have changed since backup
    - Commodity components may fail, cluster might be re-scaled
  - May need to restore to smaller/larger cluster for test/dev or analytics
- Challenge: Storage *needs* context of NoSQL DB cluster topology changes

Production Cluster
(100 nodes)

Clone to lower cardinality?

Test/Dev Target Cluster
(10 nodes)

NetApp

Data Protection Summary and Solution Directions

Potential Solution directions

- **Cluster-consistency at scale**
  - Need to tolerate faults during backup and combat eventual consistency
  - Potential solution directions:
    - Take crash consistent snapshots
    - Post process crash consistent snapshots (in a sand-box) using NoSQL DB stack to reach an cluster-consistent state [4]
- **Space Efficiency**
  - Replica set data copies do not de-duplicate. Moreover data could be encrypted.
  - Potential solution direction: Remove replicas logically (application aware backup)
- **Topology Changes**
  - Cluster topology may change across backup and restore schedules
  - Storage snapshots do not have context about cluster topology
  - Use cases may require restore to a test/dev cluster of different cardinality
  - Solution direction: Save Cluster topology and storage mapping as part of backup

More details in to-be-published USENIX, HotStorage paper:

[4] Atish Kathpal, Priya Sehgal, **BARNS: Towards Building Backup and Recovery for NoSQL Databases**, 9th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 17), July 2017, Santa Clara, CA

URL: https://www.usenix.org/conference/hotstorage17/program/presentation/kathpal

NetApp

Conclusions

- Shared storage has relevance as backend storage for NoSQL DBs
  - Independent scaling of compute and storage
  - Storage consolidation => easier administration of resources
  - Flash based networked storage can meet challenges of performance, scalability and consolidation

- Data protection
  - Existing solutions have several inefficiencies like copy-based backup and have poor integrations with shared storage
  - Opportunity for shared storage to provide differentiation through efficient snapshots and clones
  - Need to address challenges of cluster consistent and storage efficient backup and flexible topology restores

Thank You.

atish-[at]-netapp.com

Acknowledgements: Priya Sehgal, Gaurav Makkar, Parag Deshmukh

NetApp