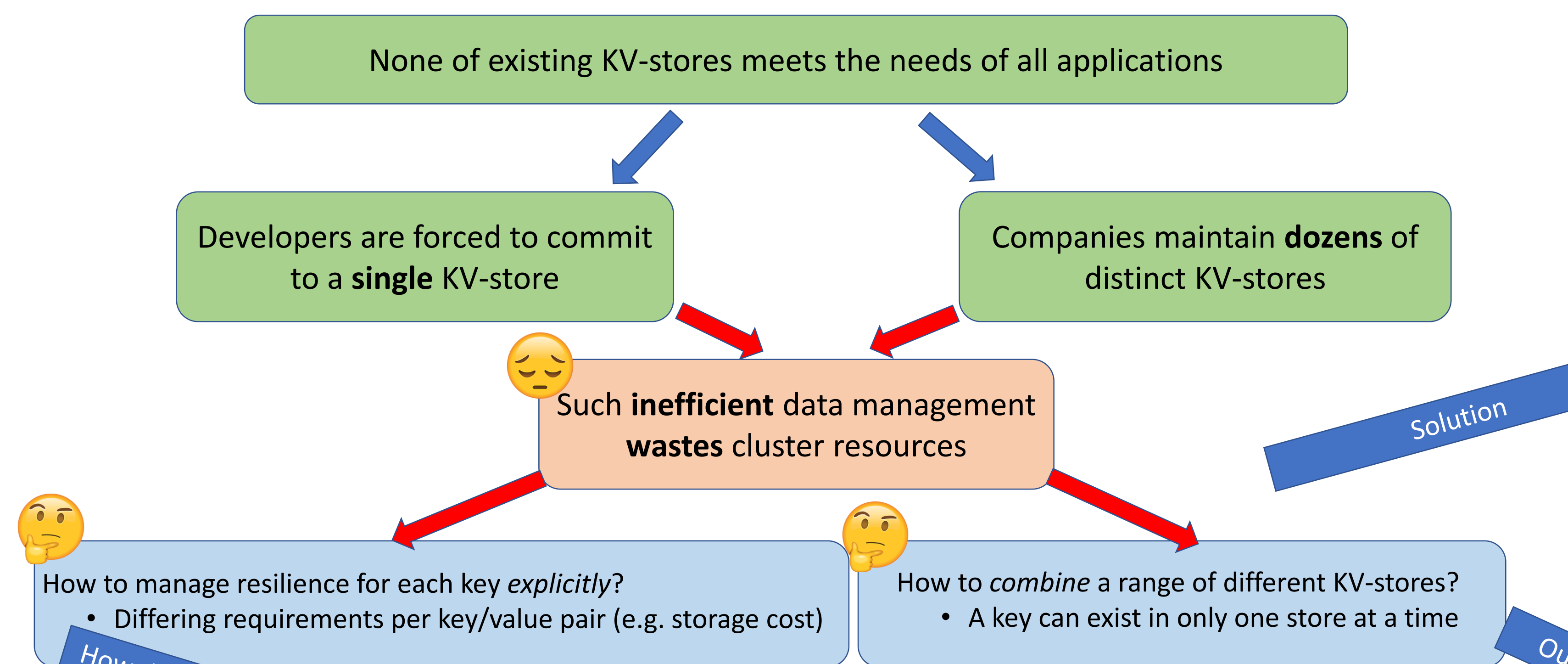




Konstantin Taranov, Torsten Hoefler

ETH Zurich, Scalable Parallel Computing Laboratory

## Problems & Motivation



## Ring KV-store

**Use cases**

RING is a multi-purpose KV-store

- One KV-store instead of a range of them

Resilience requirements to the data may alter over time

- For example, iterative algorithms such as PageRank

Multi-temperature data management

- Ring allows moving data from cold storage to hot one silently

Speedup heavy updates

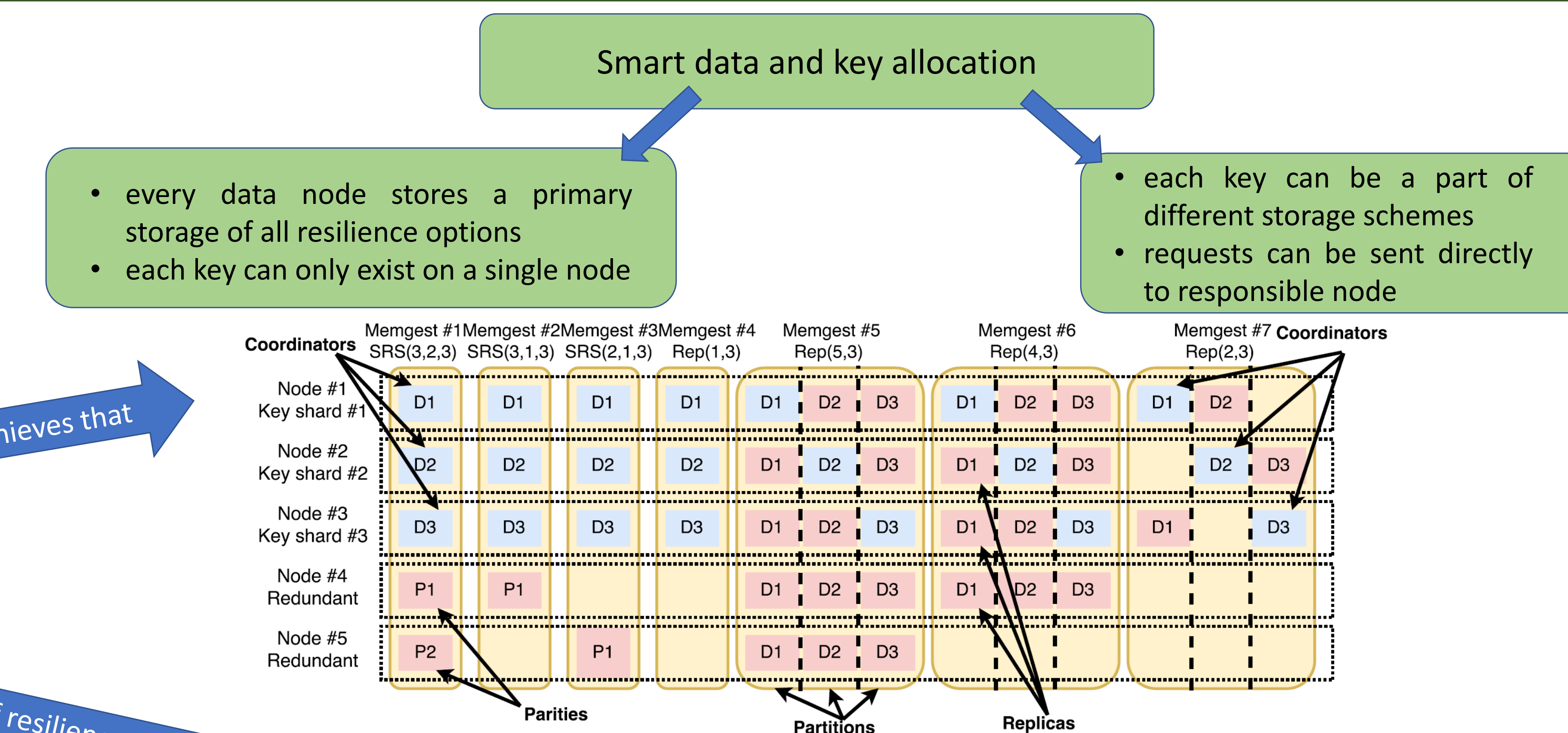
- Users may exploit temporal locality of frequently accessed keys

Conventional	Update resilience	Manage resilience
Get(key) Put(key,object) Delete(key)	Put(key,object,mID) Move(key,mID)	Create(mID,descriptor) Remove(mID) ReadDescriptor(mID)

**What are benefits?**

- It allows explicitly managing memory overhead of key-value pairs, which are stored in relatively expensive DRAM.
- Internal network traffic between storage nodes is strongly correlated to the used storage algorithm, thereby managing latency and throughput of clients' requests.
- Programmers can specify reliability and availability to ensure the required resilience of key-value pairs.

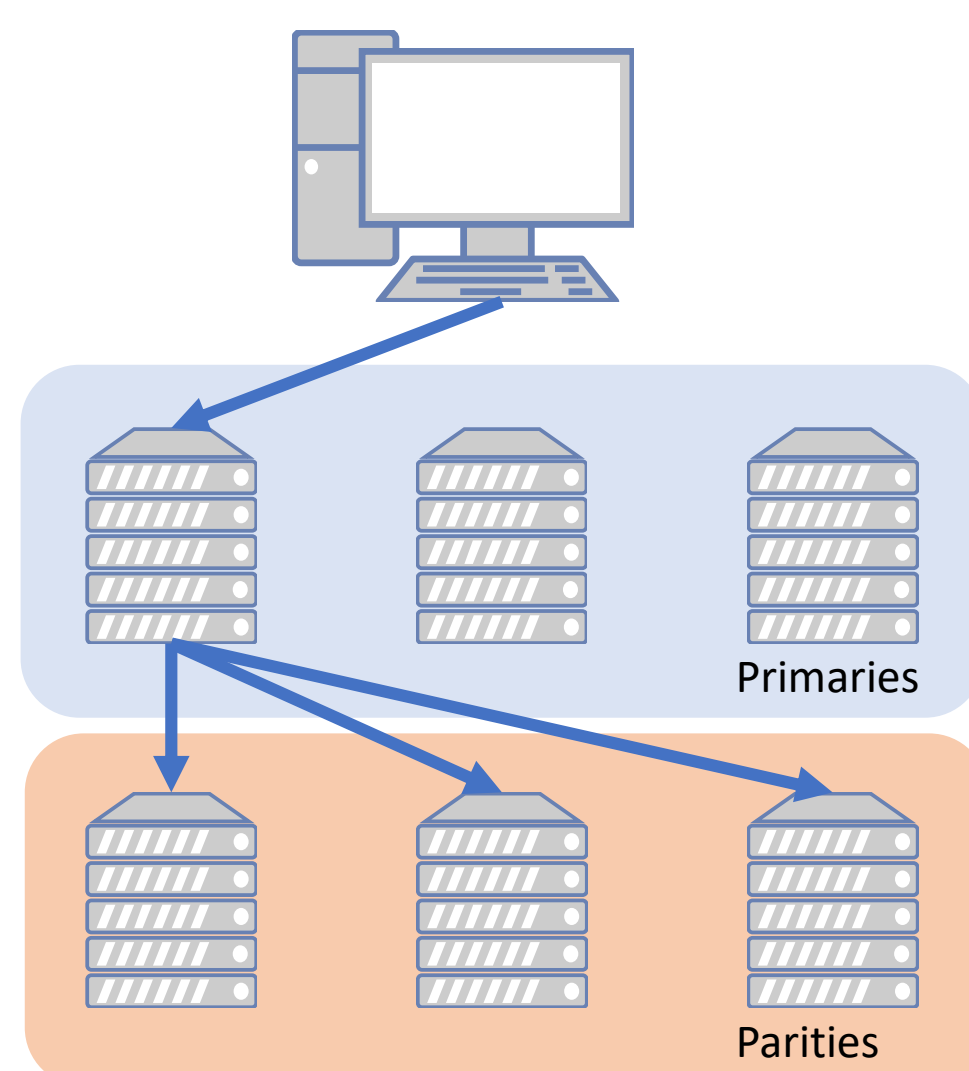
## System architecture



## Popular storage schemes

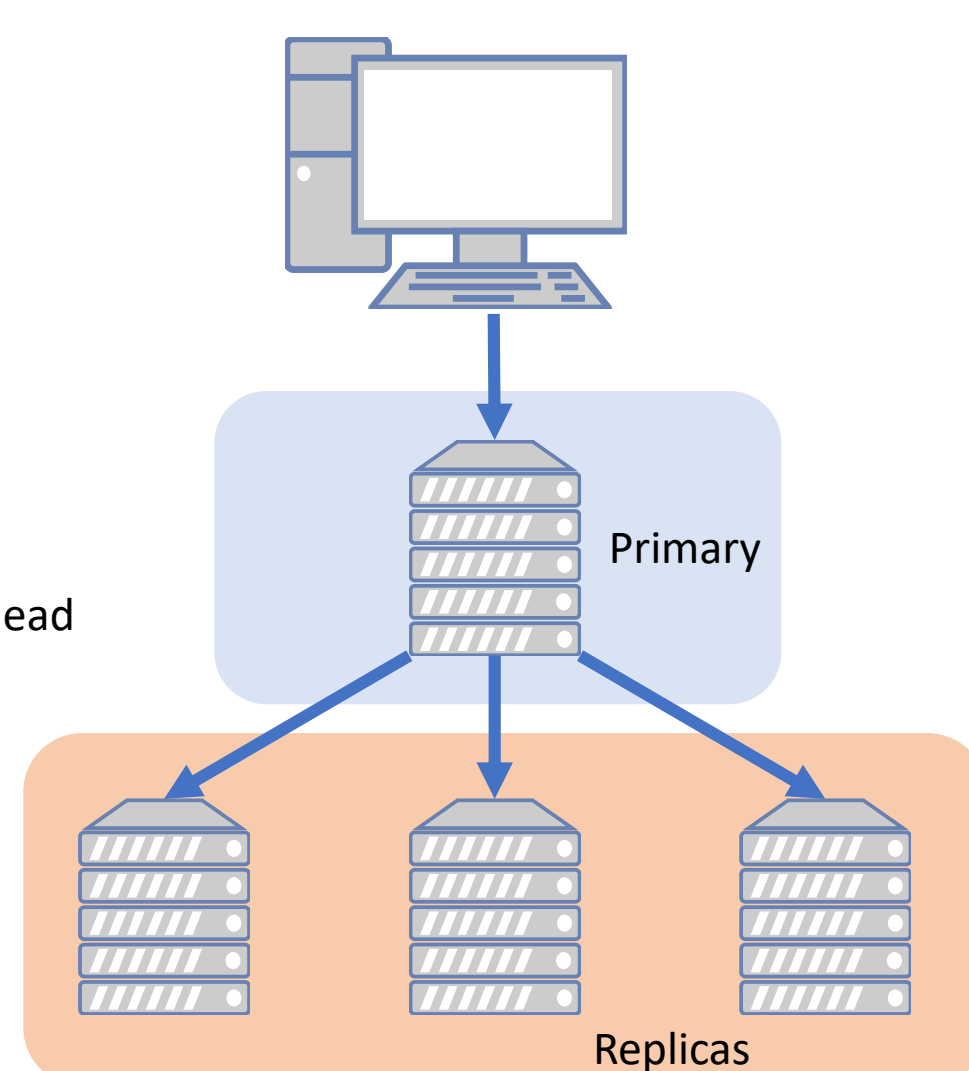
### Erasure coding

- ✓ Plenty of resilience options
- ✓ The lowest storage overhead
- ✓ High reliability
- ✗ Complex
- ✗ Slow recovery (Low availability)
- ✗ Additional computations

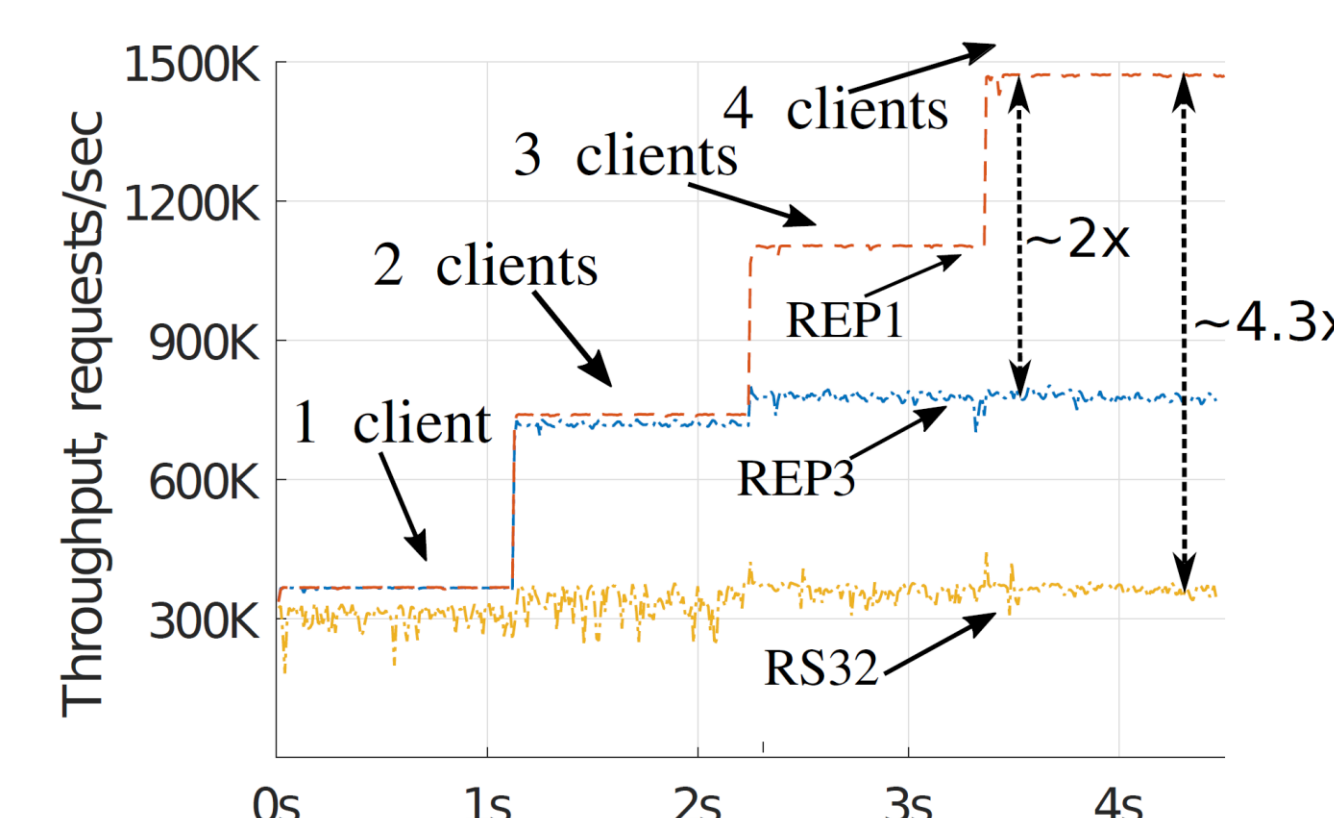


### Replication

- ✓ Easy to implement
- ✓ High availability
- ✓ Basic reliability
- ✗ High storage overhead



The way how data is stored implies completely different trade-offs between cluster resources such as *memory usage, network load, latency, availability and many others.*



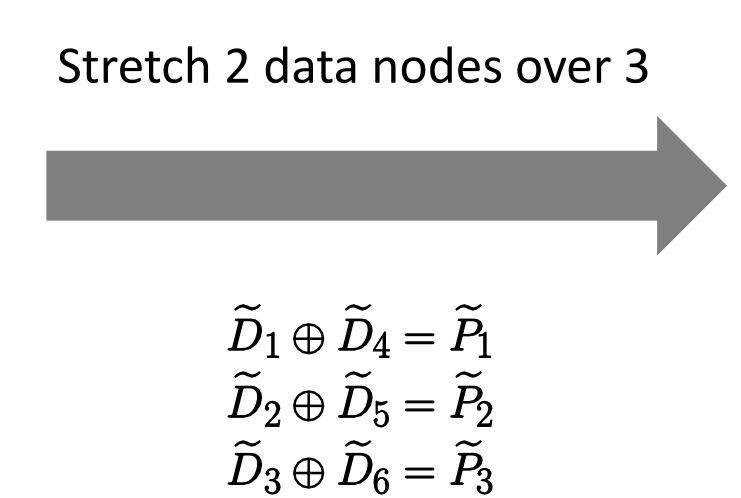
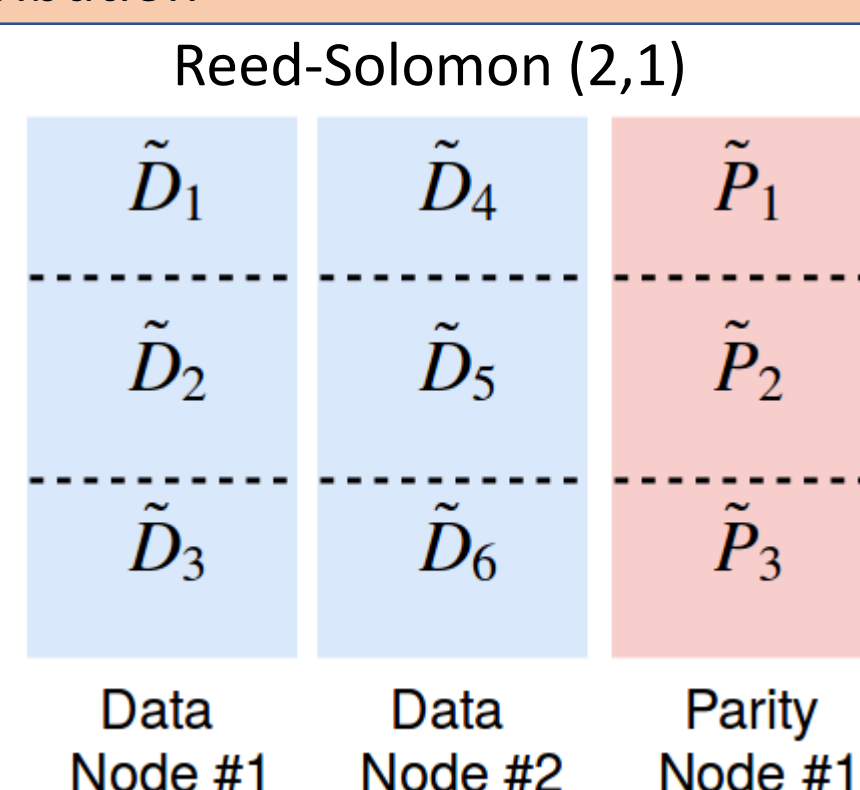
## Stretched Reed-Solomon codes

**Problem**

- Different Reed-Solomon schemes can have different hash key distributions
- Changing resilience must not change hash key distribution

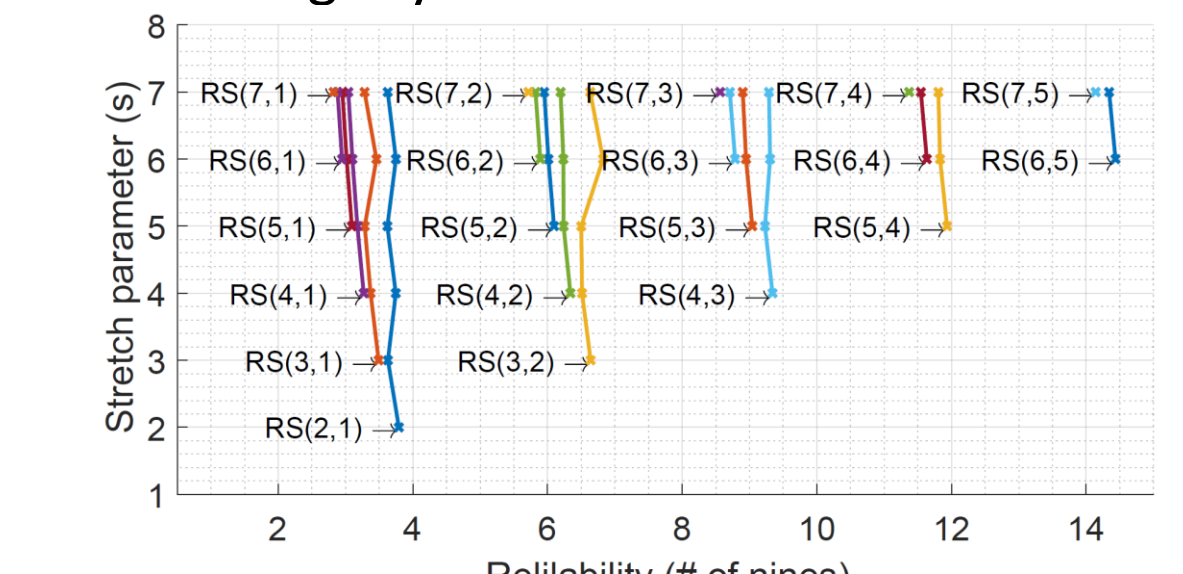
### Solution

- Retain original coding scheme
- Rearrange data blocks to preserve hash key distribution



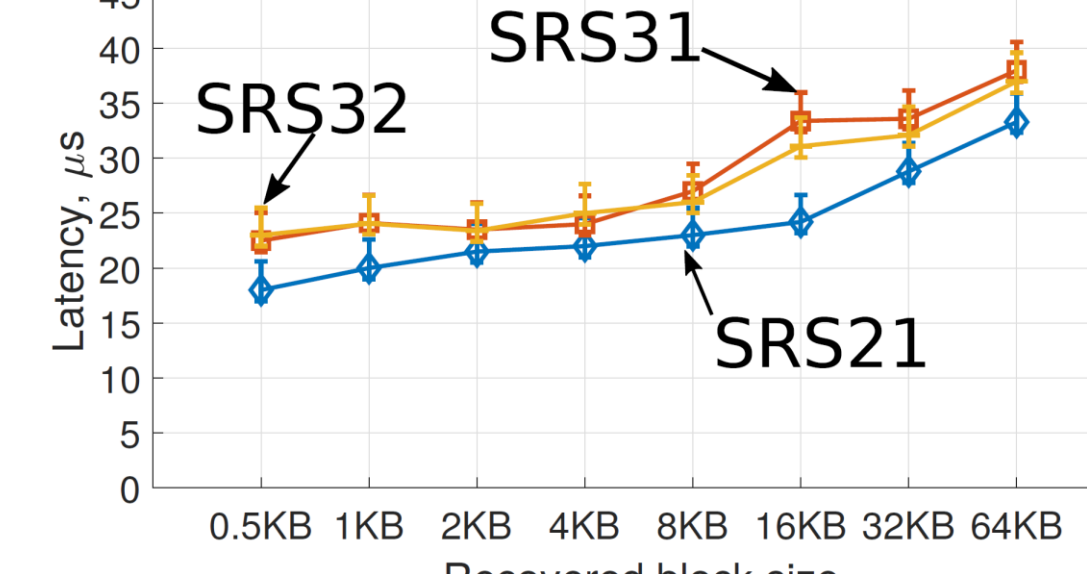
**How does stretching affect reliability?**

- Reliability stays almost the same
- It can slightly increase and decrease reliability



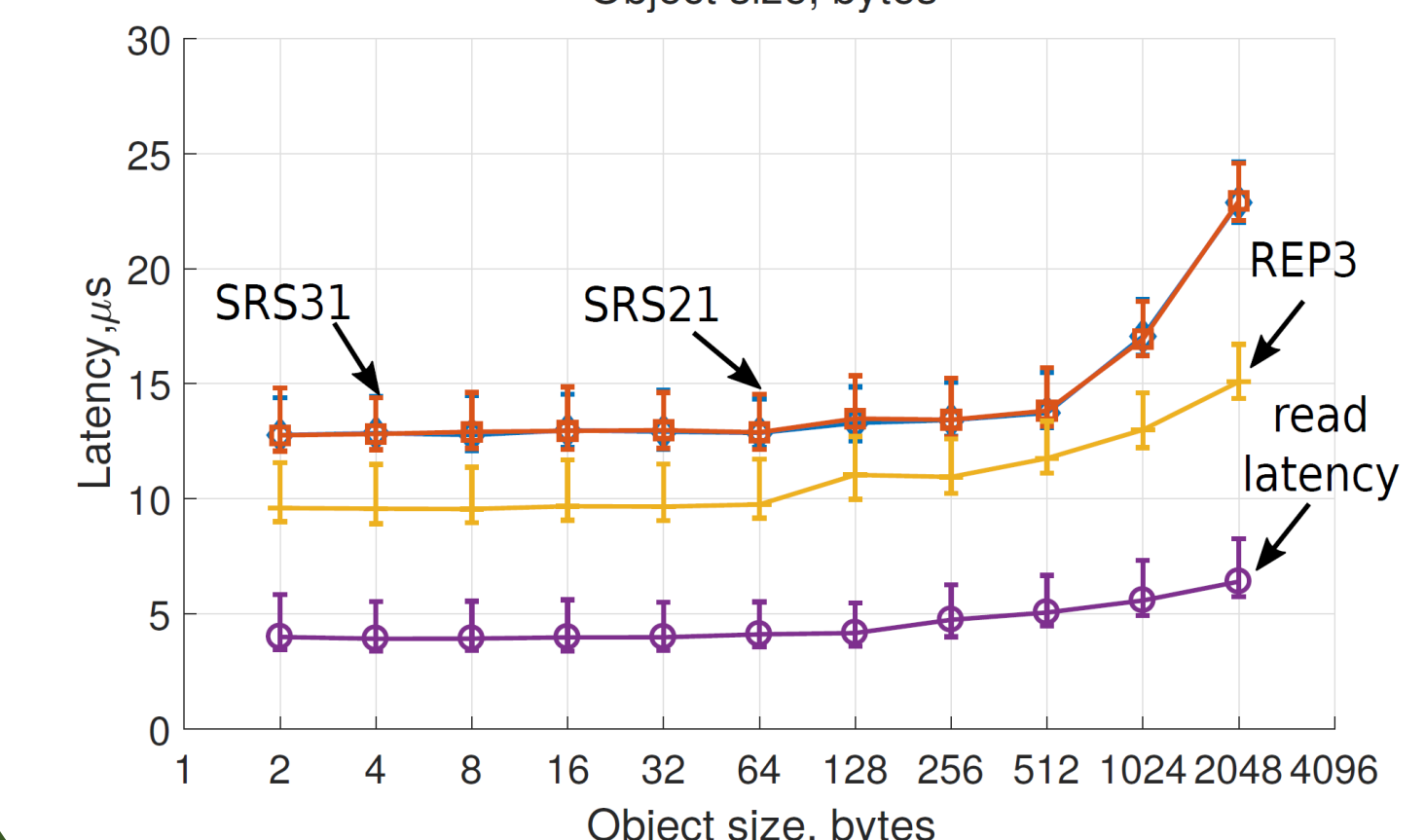
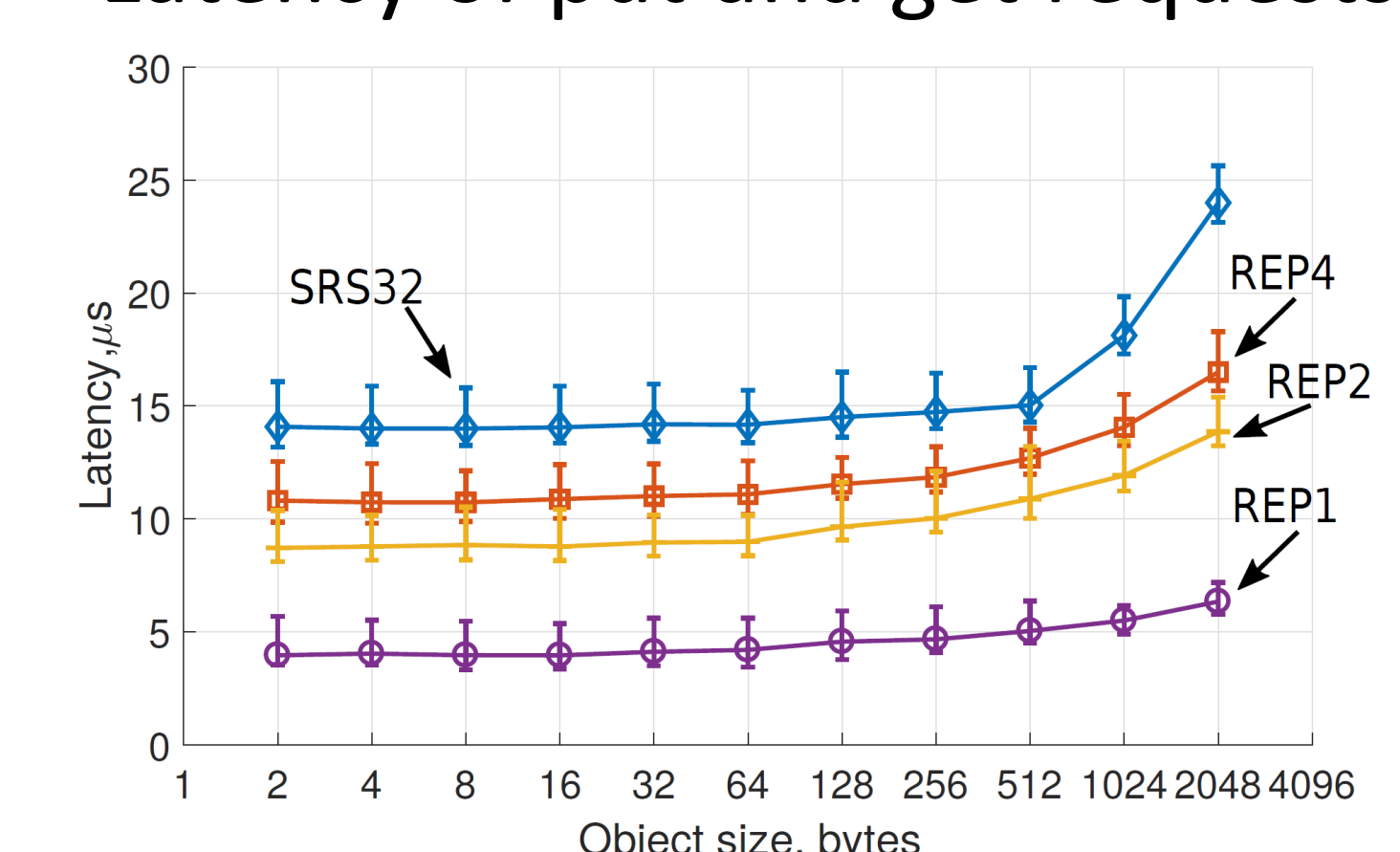
Are SRS(2,1,3) and SRS(3,1,3) the same?

NO! They have different recovery time!



## Evaluations

### Latency of put and get requests



### Latency of move requests

