PERCONA

Percona XtraDB Cluster
powered by Galera

Peter Zaitsev
CEO, Percona
Slide Credits: Vadim Tkachenko
Percona MySQL University,
Montevideo,UY Feb 5,2013

This talk

- High Availability
- Replication
- Cluster

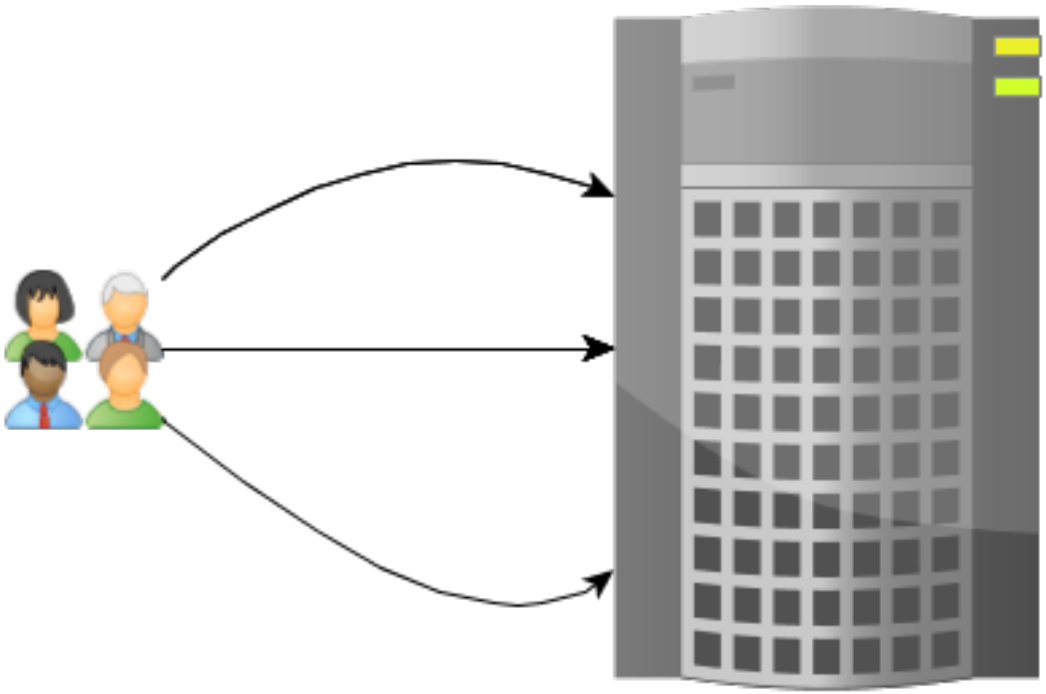
What is HA

Availability

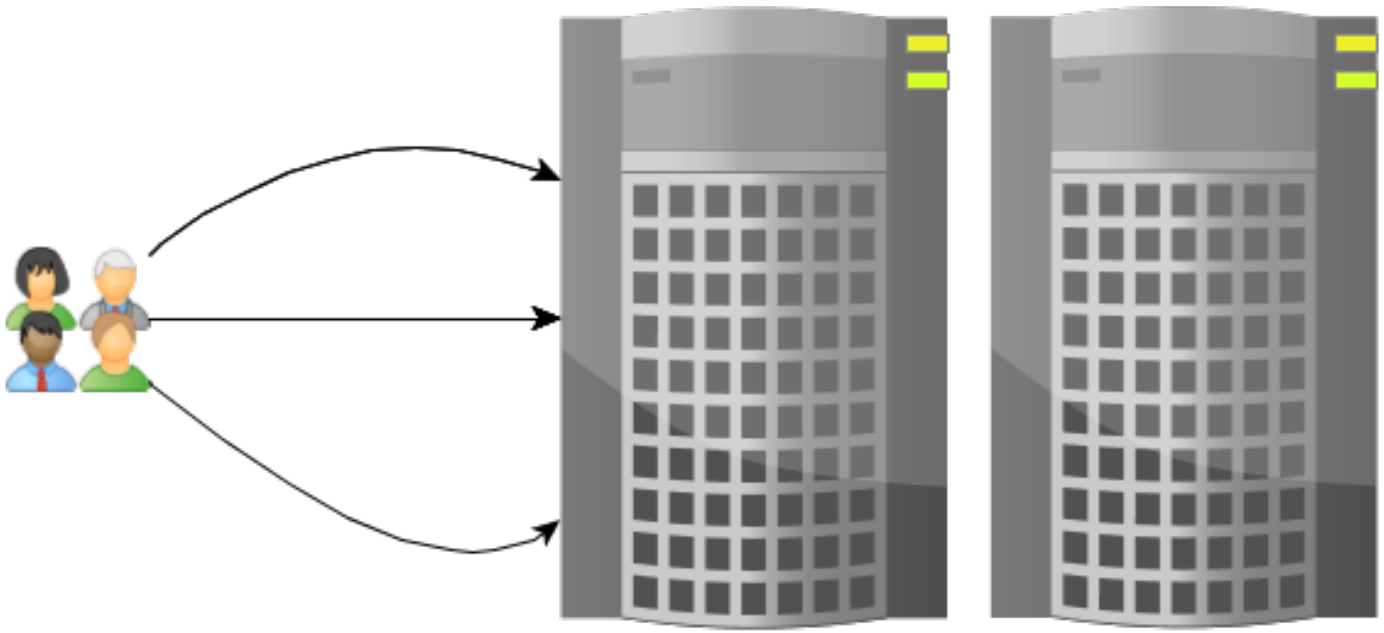
Avail ~ Ability

Ability to Avail

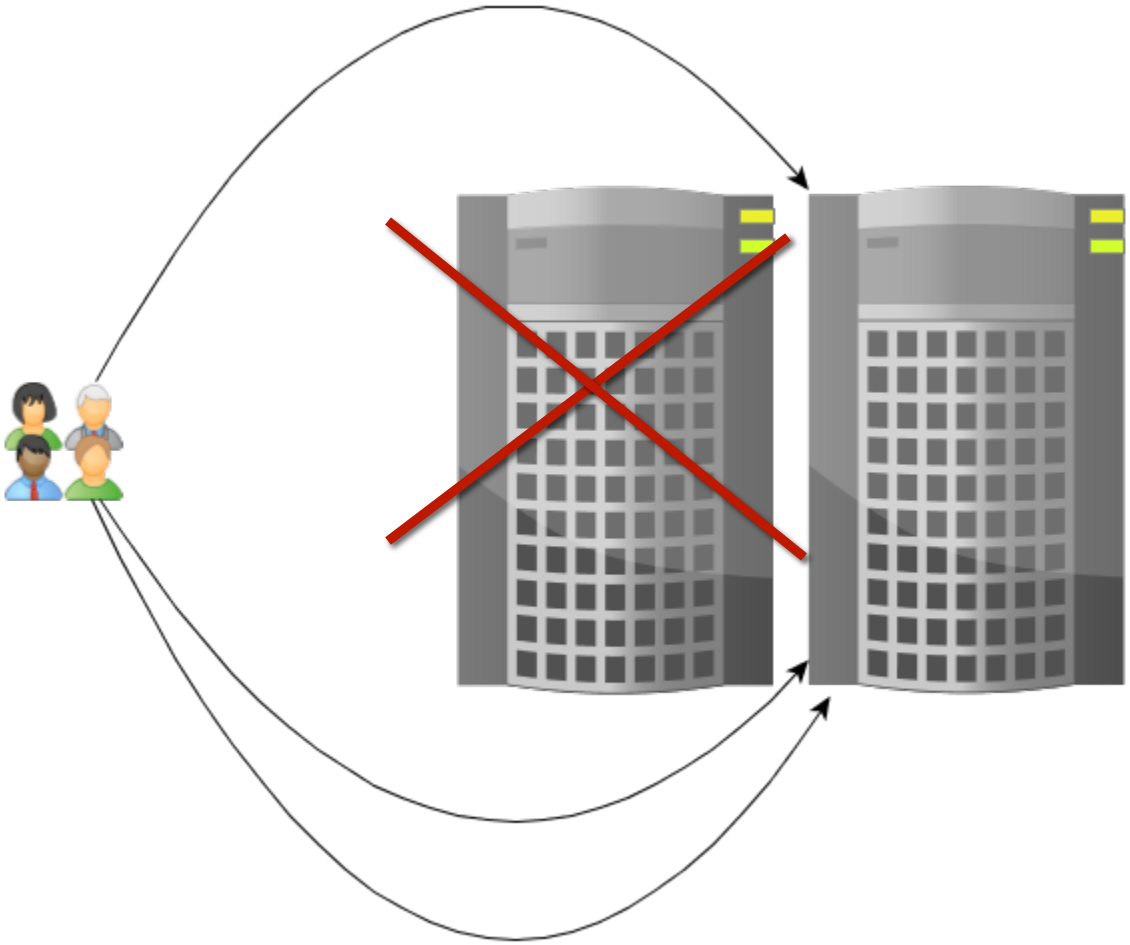
Availability by redundancy



Duplicate resources



Failover



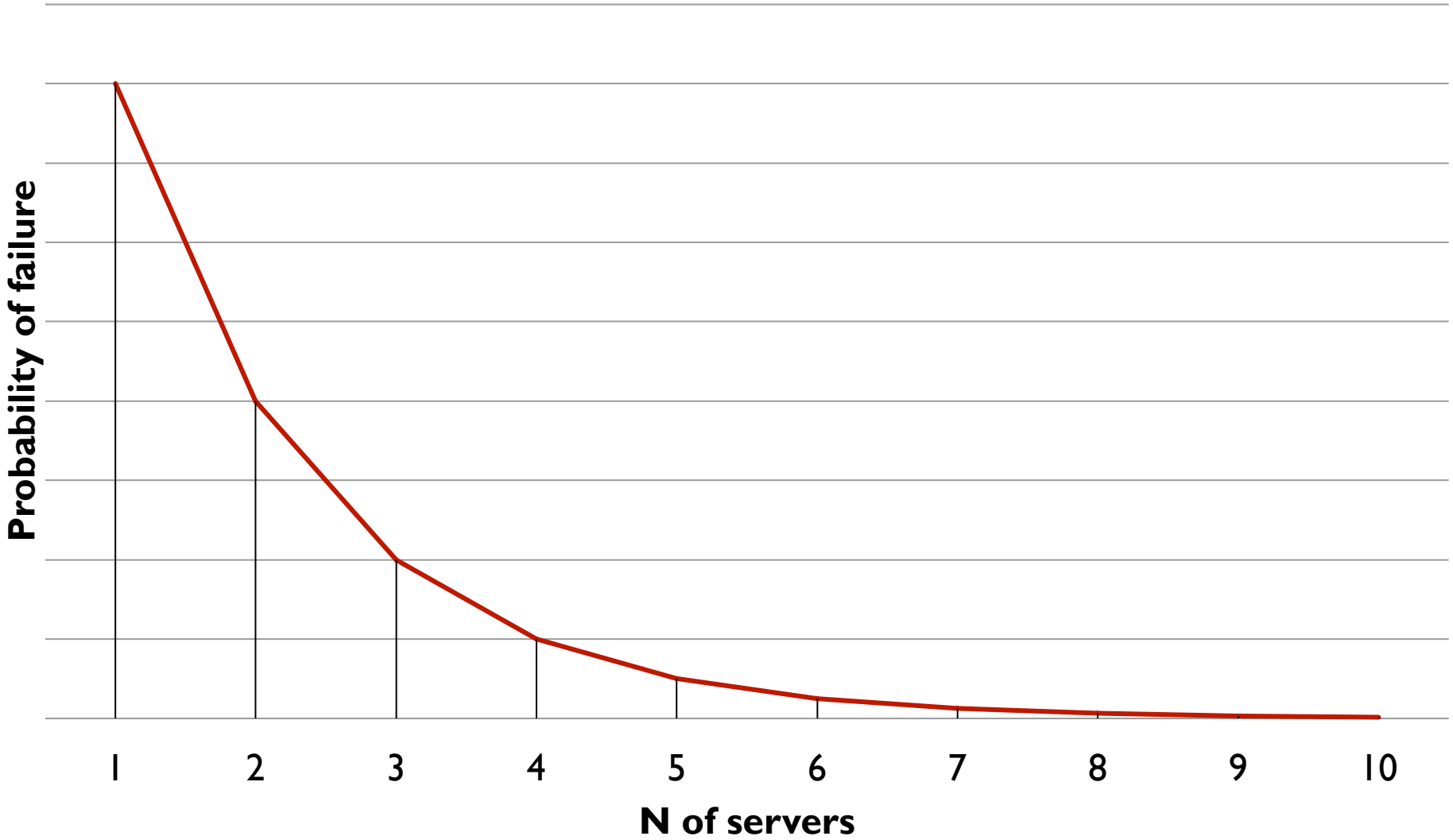
Probability of failure

Single server: P

Two servers: P/2

X servers: P/X

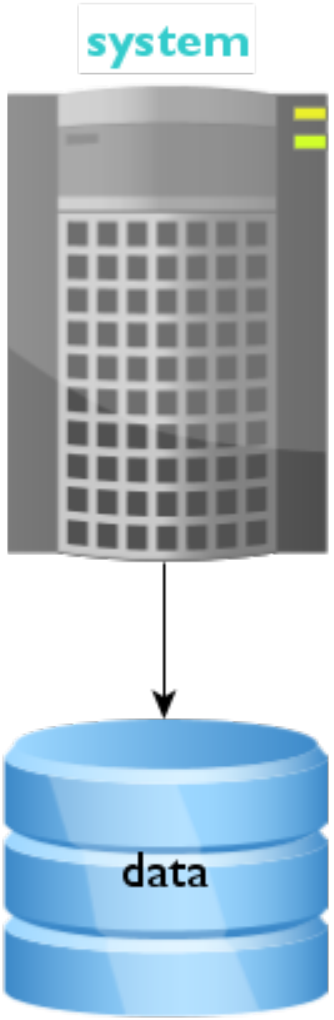
Probability of failure



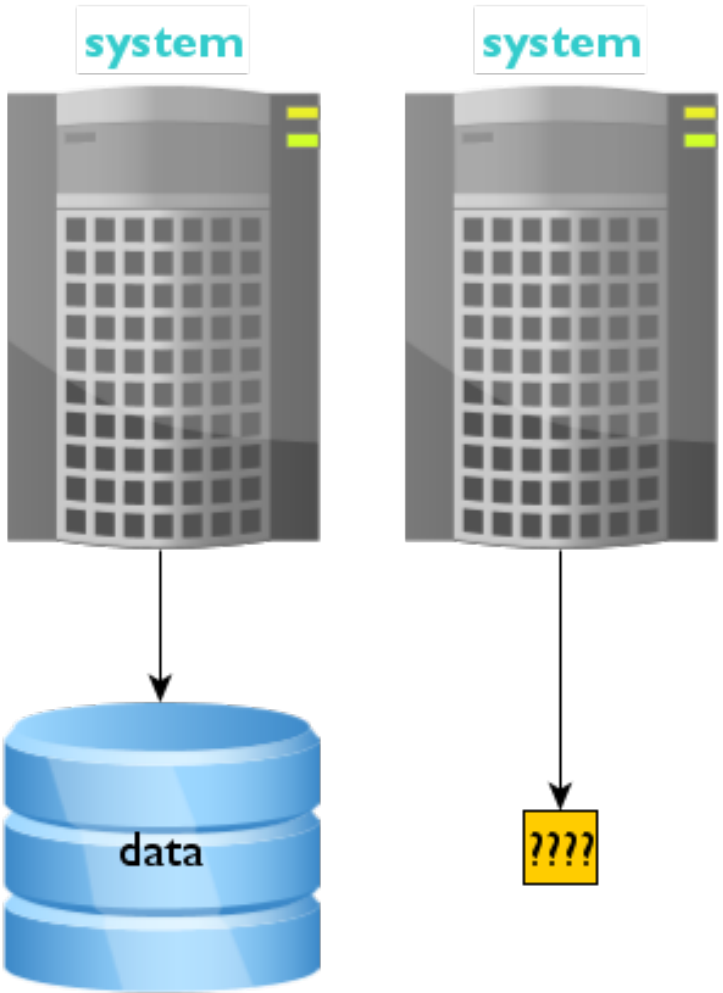
Easy ?

Not if we deal with databases

Database



Redundancy ?

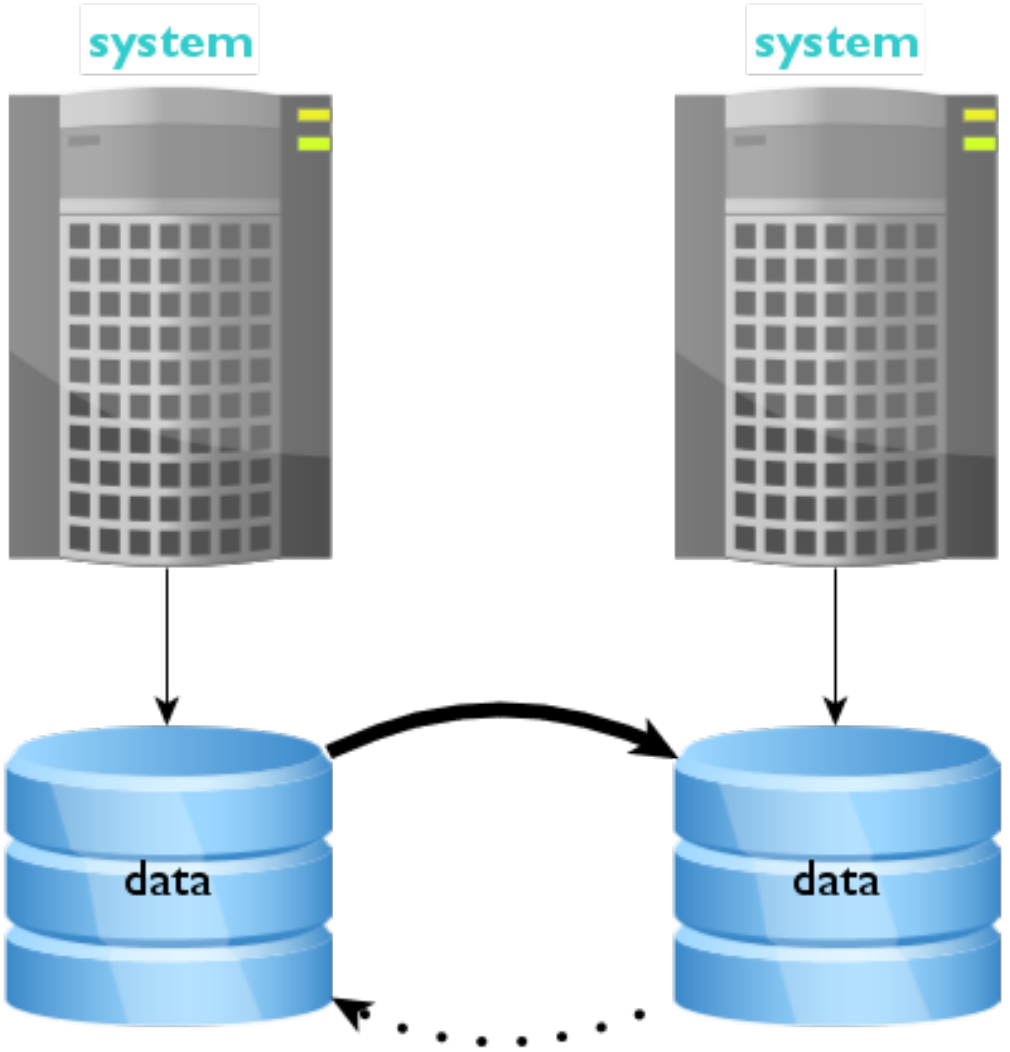


Database availability is hard

Service availability

Data availability

Replication



MySQL Replication

If your HA is based on MySQL Replication –
You are doing it wrong

What is wrong with MySQL replication ?

“a”

What is wrong with MySQL replication ?

"a" in async

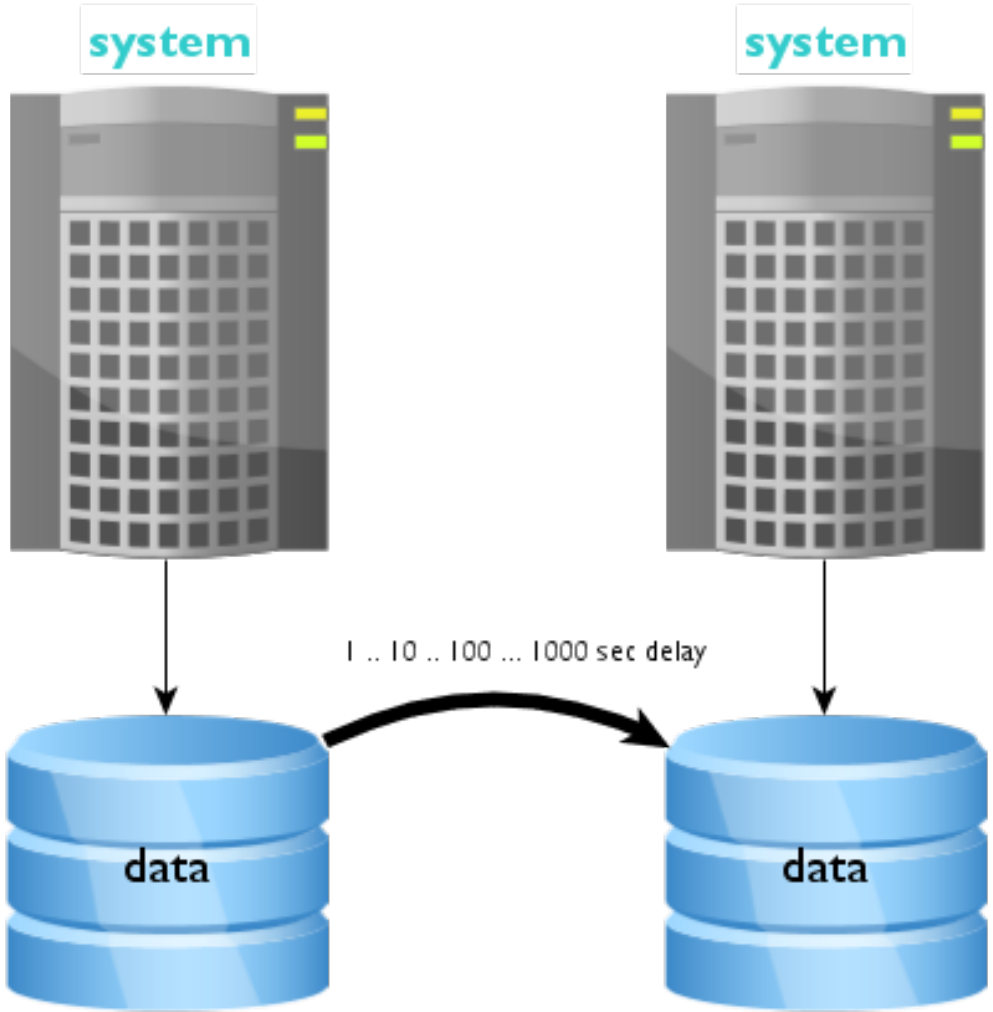
What is wrong with MySQL replication ?

"async"

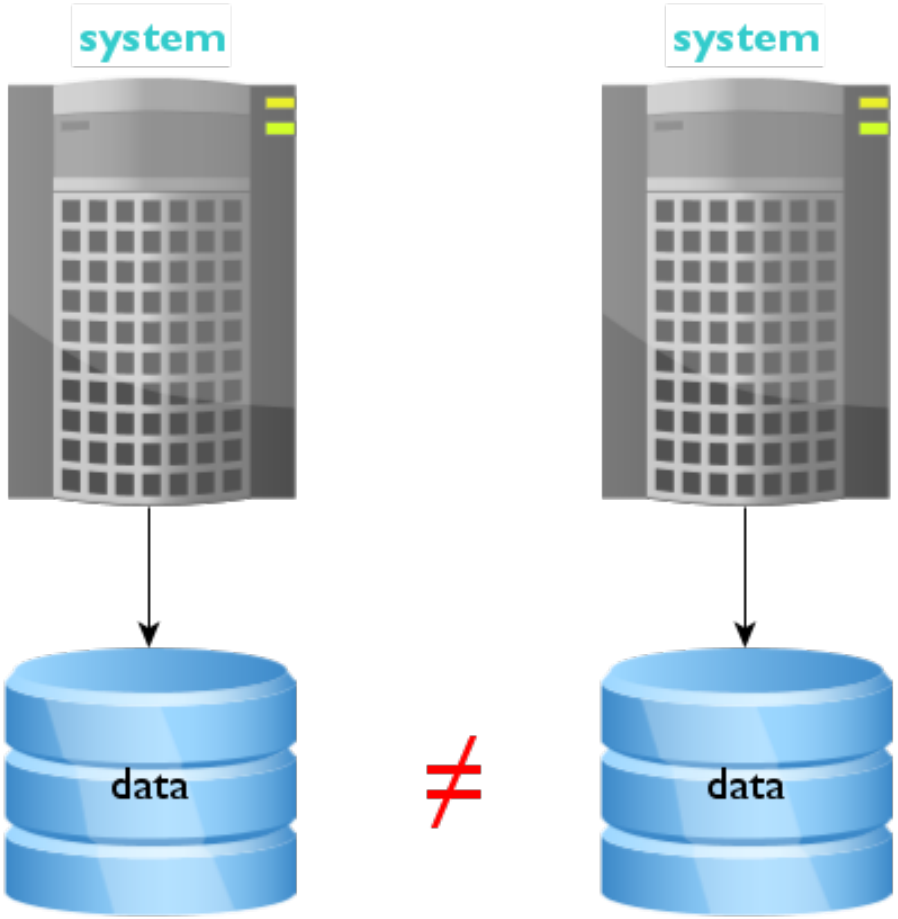
vs

"sync"

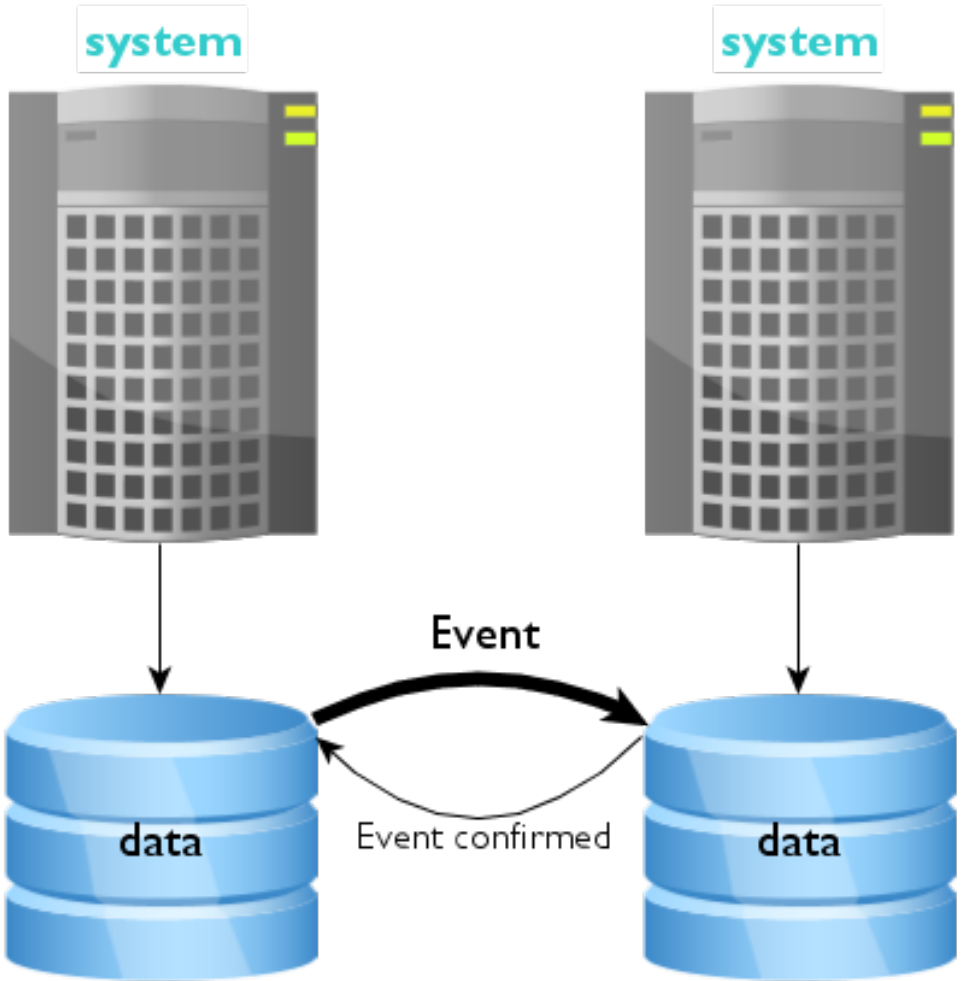
Async



Async

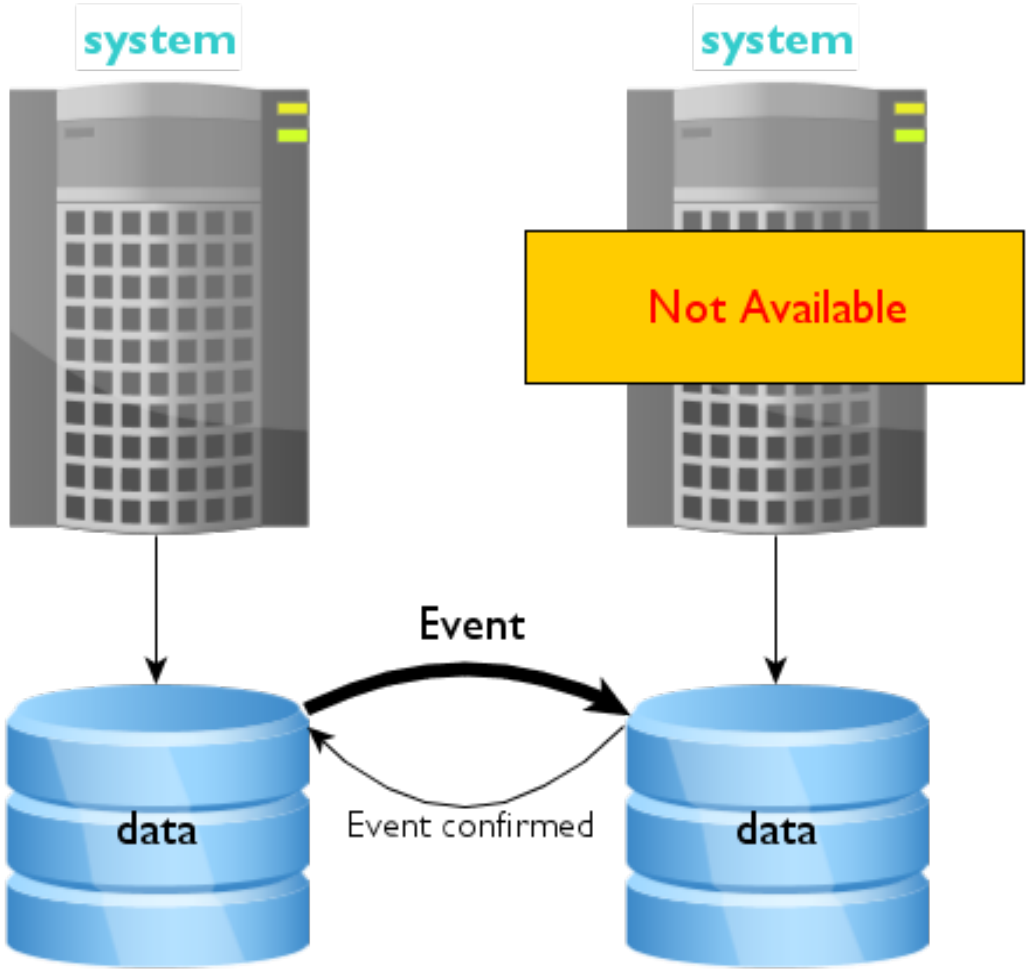


sync



Didn't we just reinvent DRBD ?

DRBD

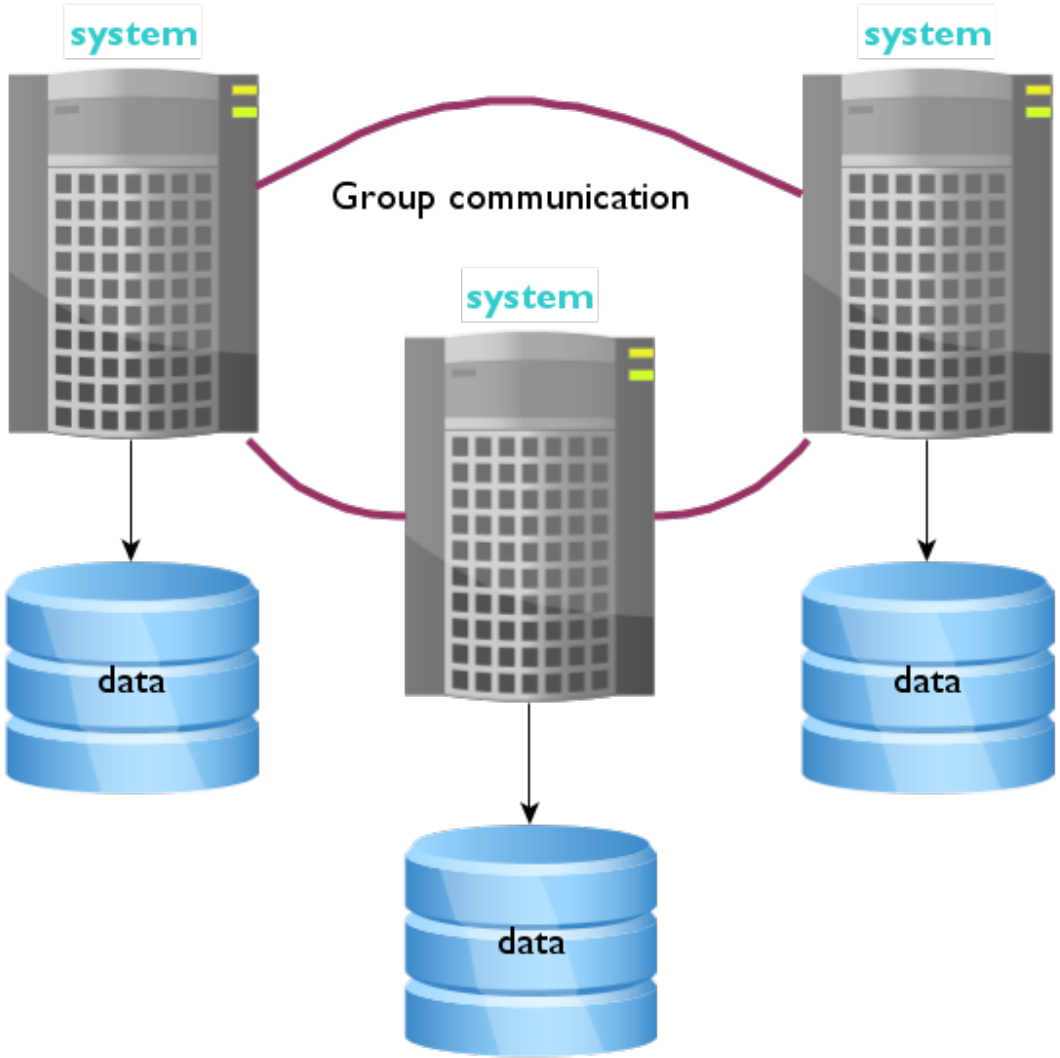

system
system
Not Available
Event
data
Event confirmed
data

Clustering

Percona XtraDB Cluster

Free and Open Source

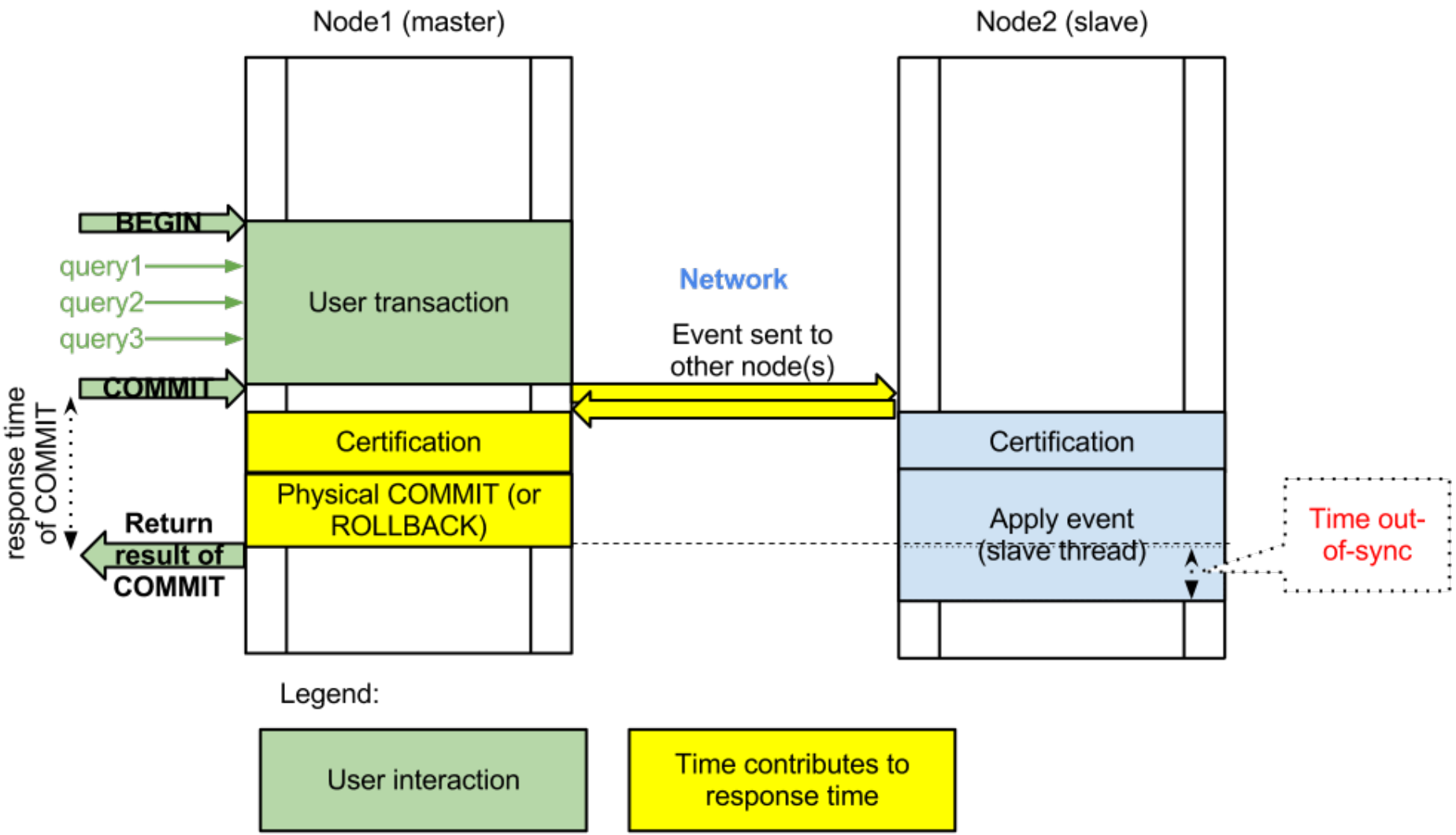
Percona XtraDB Cluster



Virtually synchronous

http://en.wikipedia.org/wiki/Virtual_synchrony

Virtually synchronous



synchronous replication

multi-master replication

parallel applying on slaves

data consistency

automatic node provisioning

synchronous replication

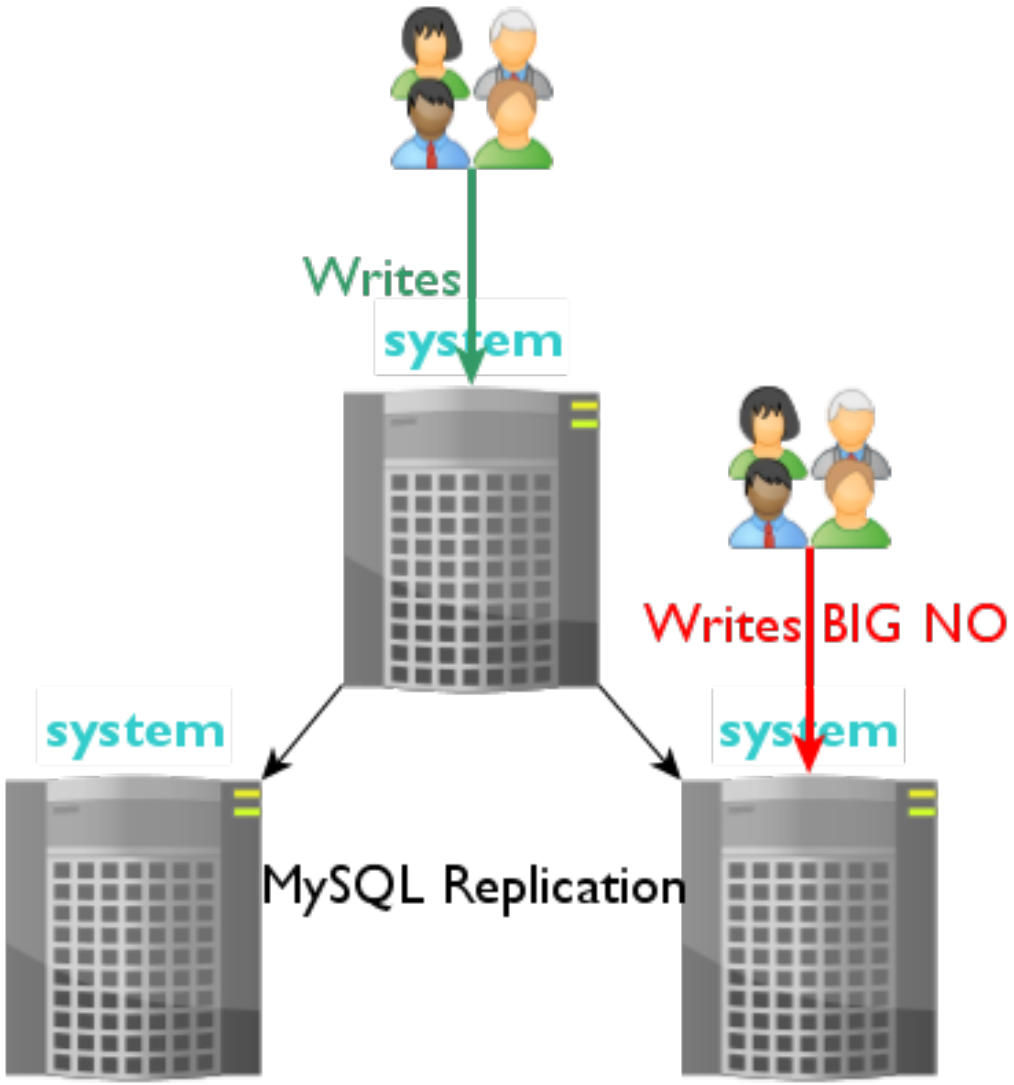
multi-master replication

parallel applying on slaves

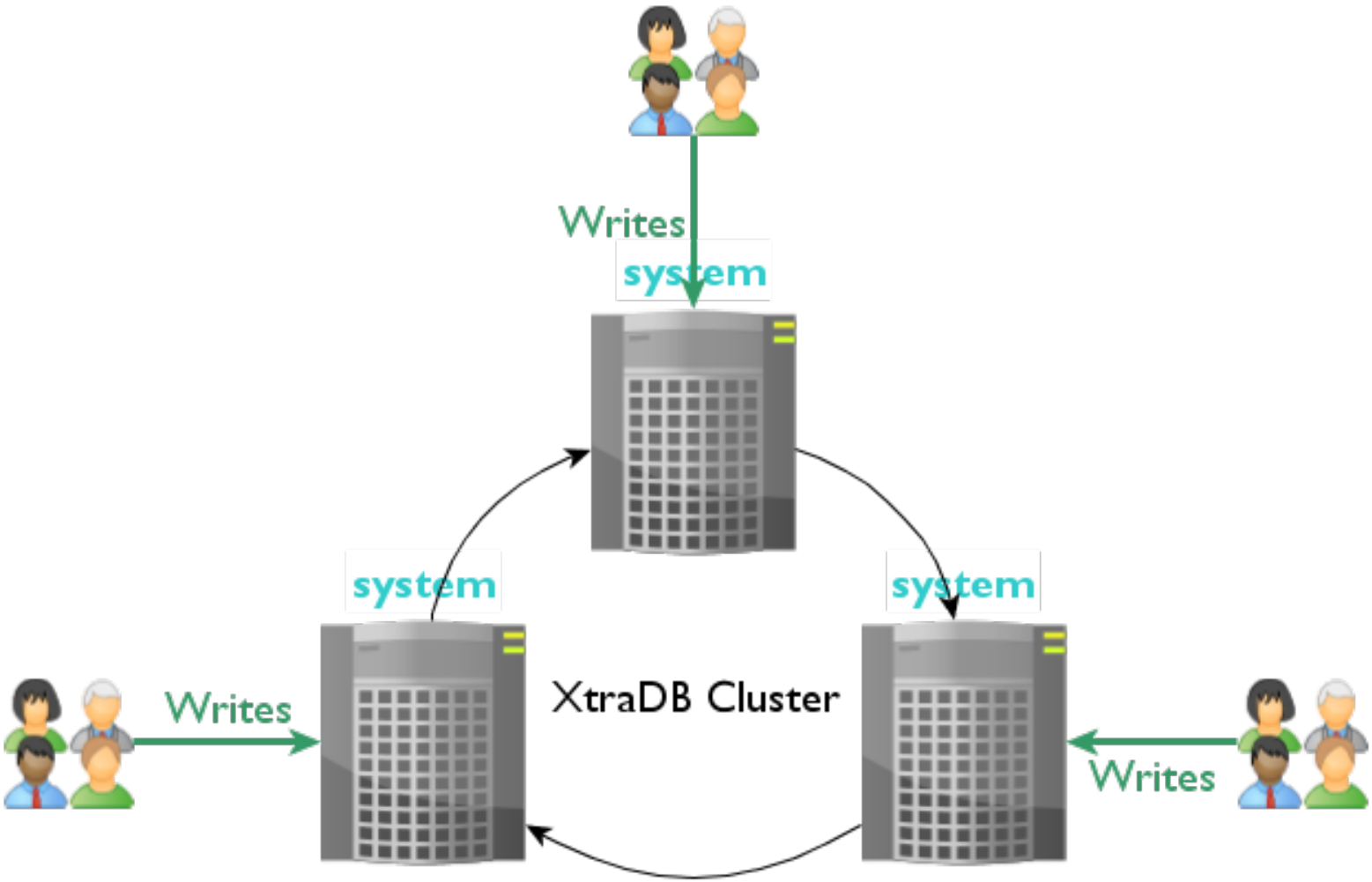
data consistency

automatic node provisioning

Multi-master: MySQL



Multi-master: XtraDB Cluster



synchronous replication

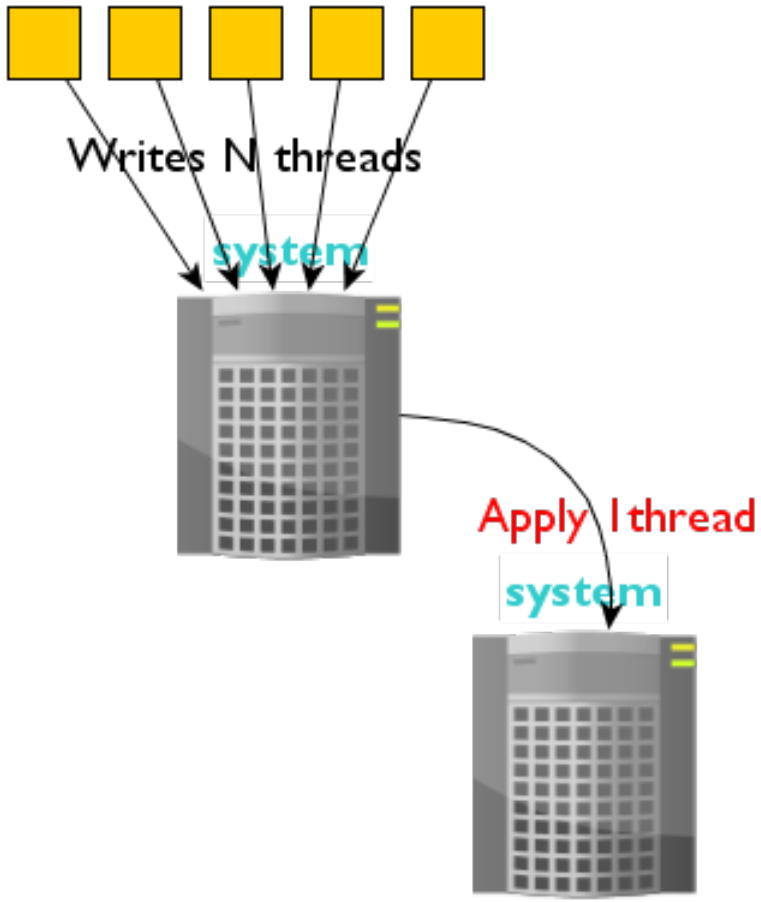
multi-master replication

parallel applying on slaves

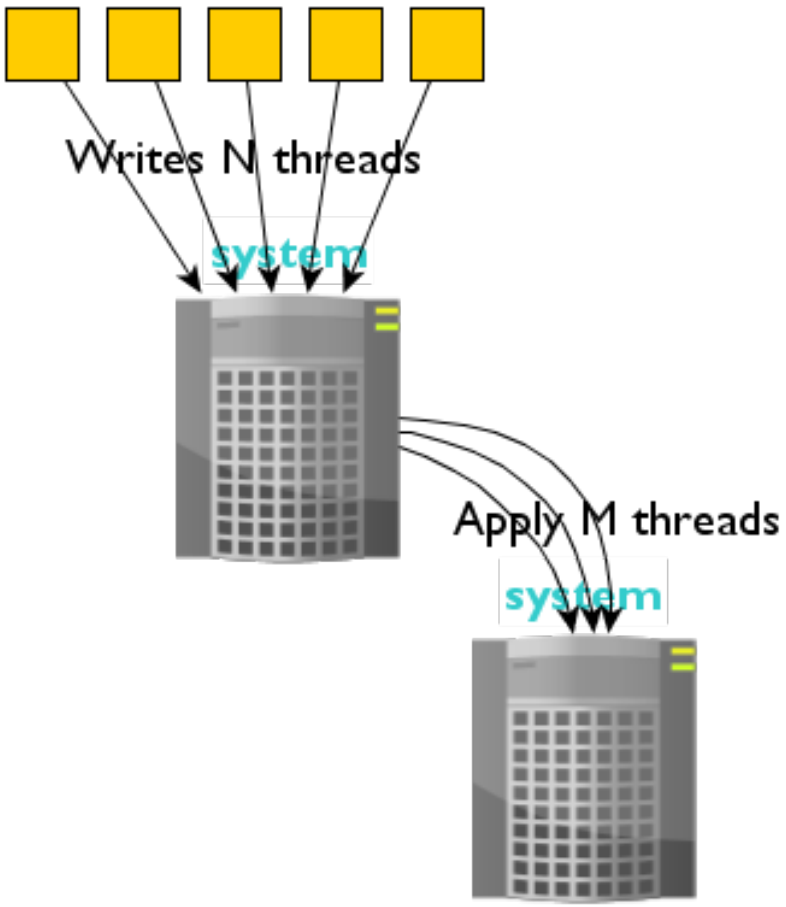
data consistency

automatic node provisioning

Parallel apply: MySQL



Parallel apply: XtraDB Cluster



synchronous replication

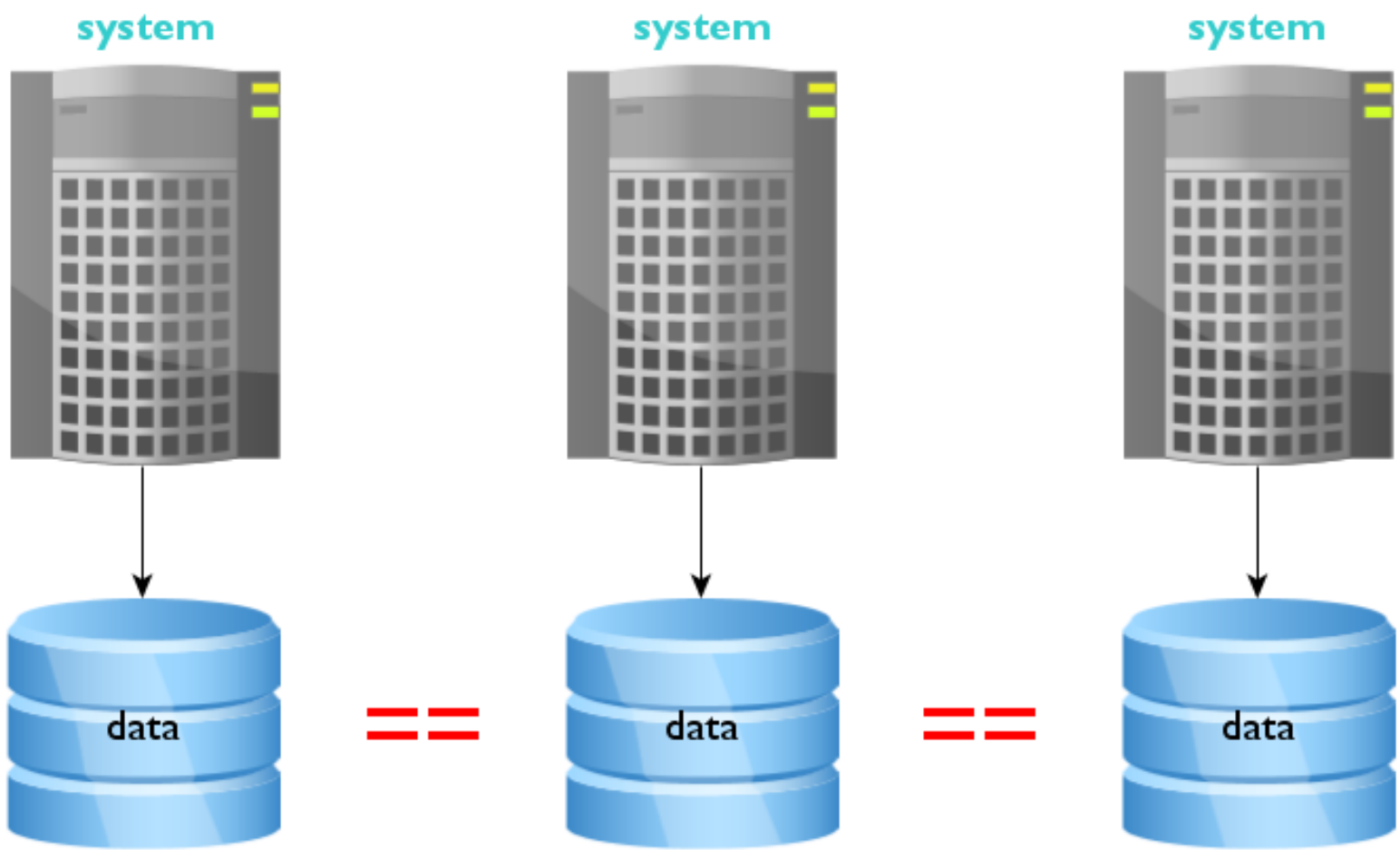
multi-master replication

parallel applying on slaves

data consistency

automatic node provisioning

XtraDB Cluster data consistency



synchronous replication

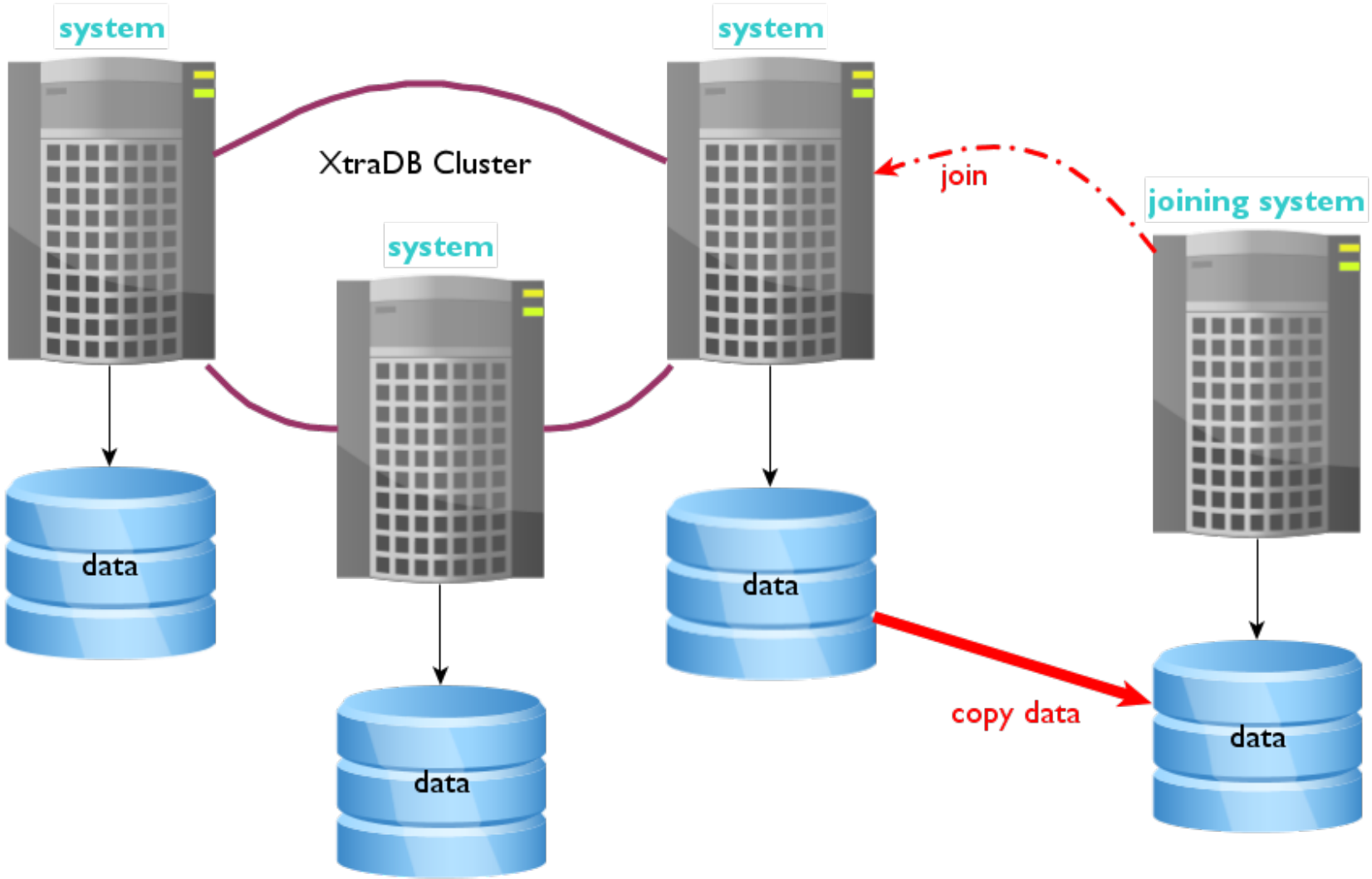
multi-master replication

parallel applying on slaves

data consistency

automatic node provisioning

Node provisioning



CAP theorem

http://en.wikipedia.org/wiki/CAP_theorem

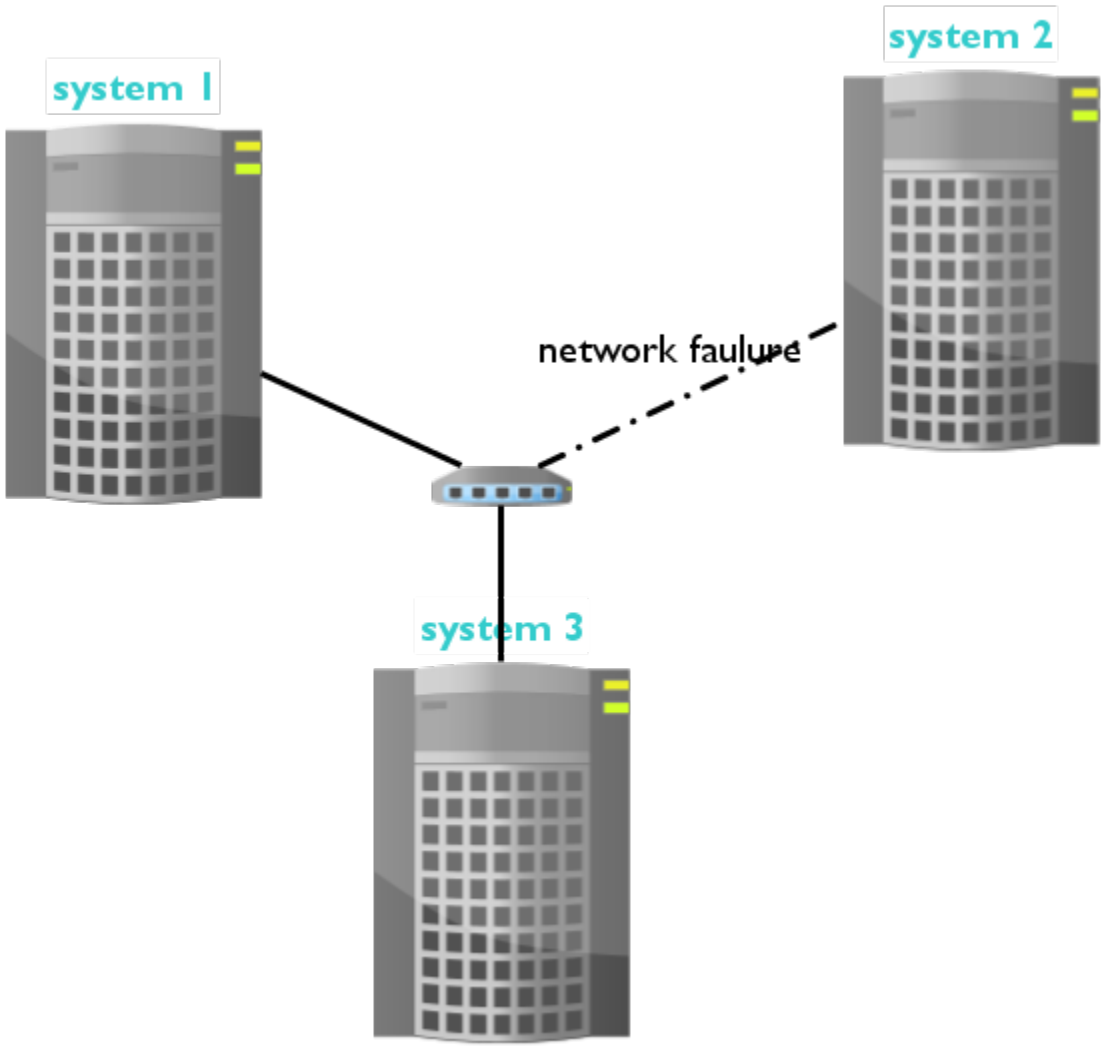
Pick only TWO

Consistency

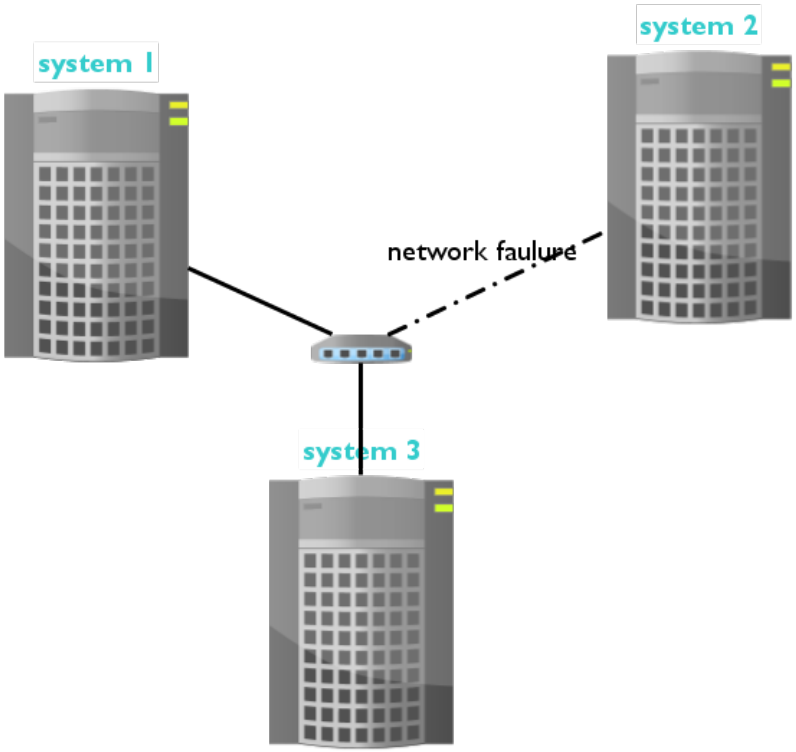
Node availability

Partition Tolerance

Network failure



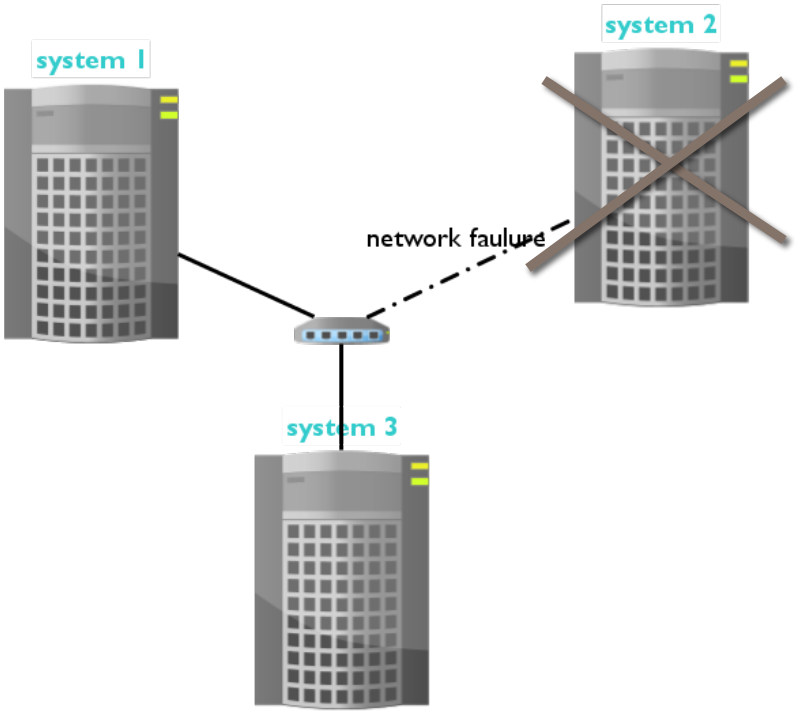
MySQL Replication



Access to all systems - YES

Data consistency - NO

XtraDB Cluster

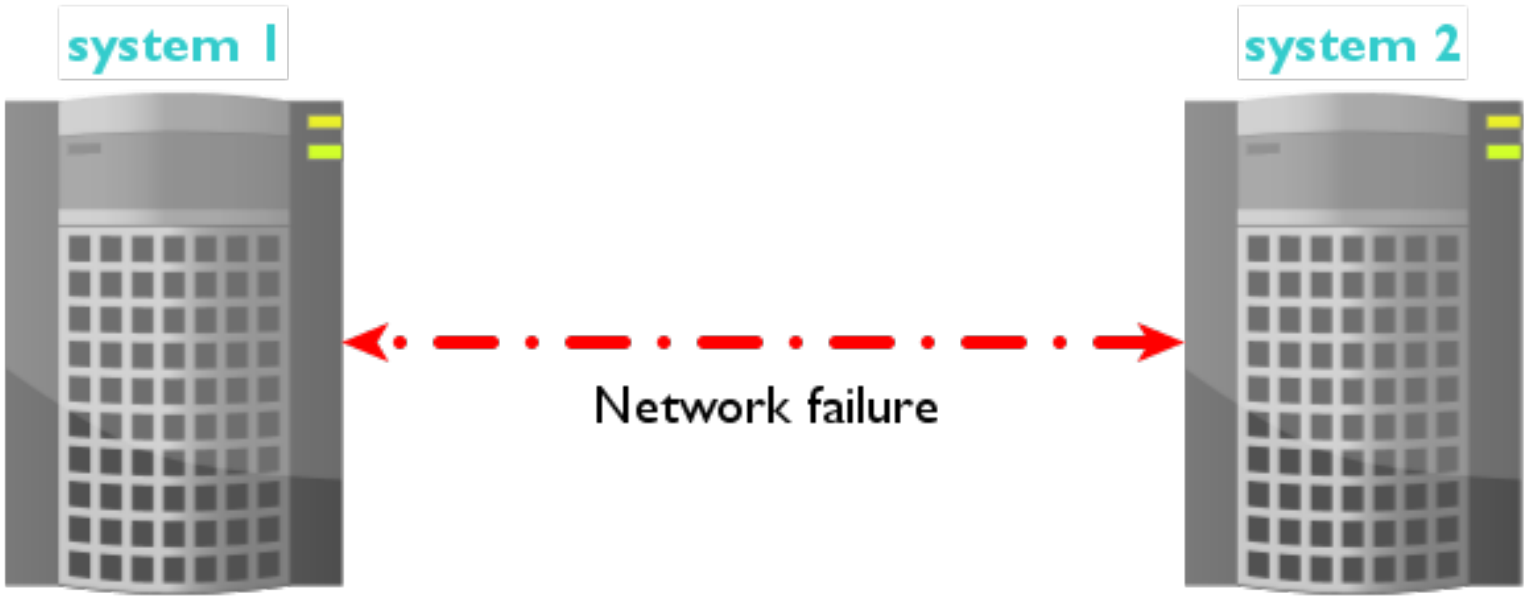


Access to all systems - NO

Data consistency - YES

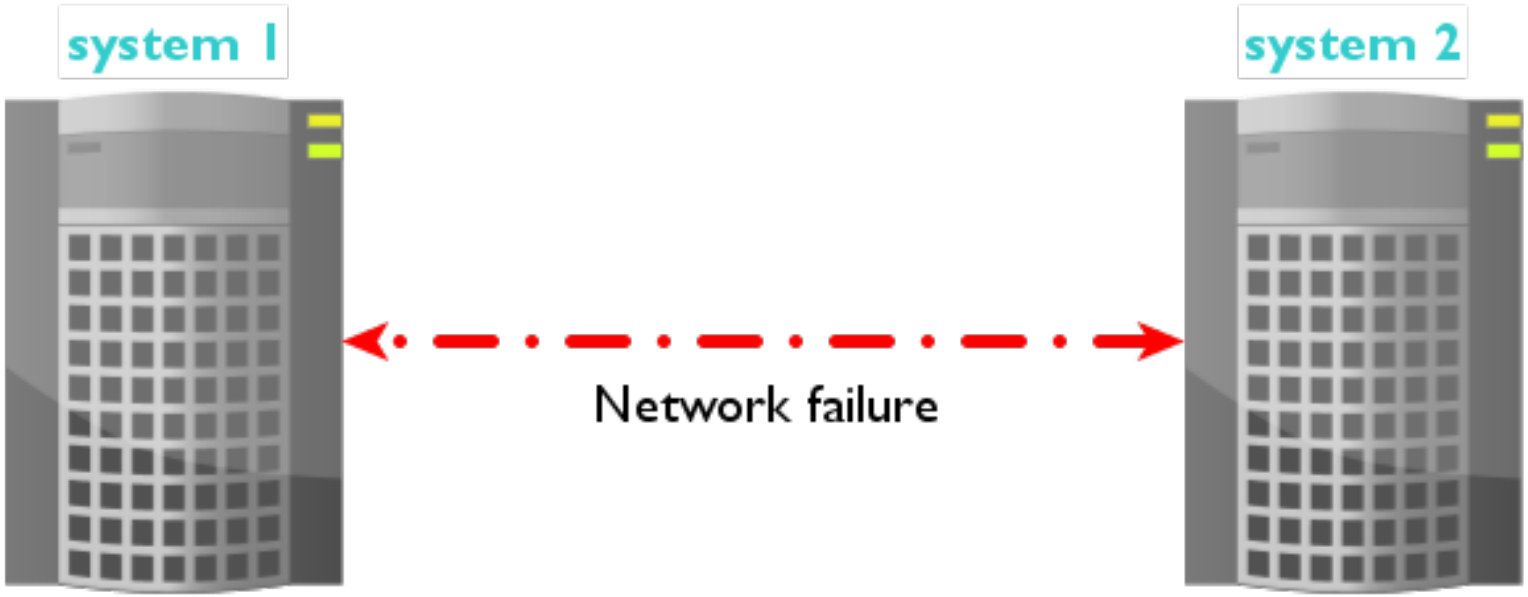
3 nodes is the minimal recommended configuration

Split brain



Which system to make available ?

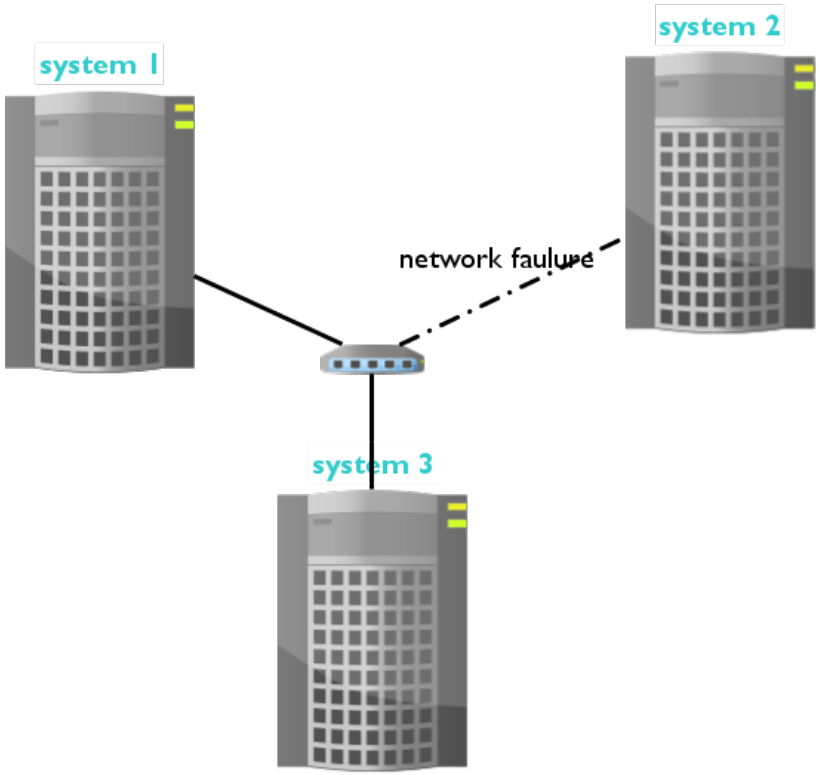
Split brain



You still can have this setup

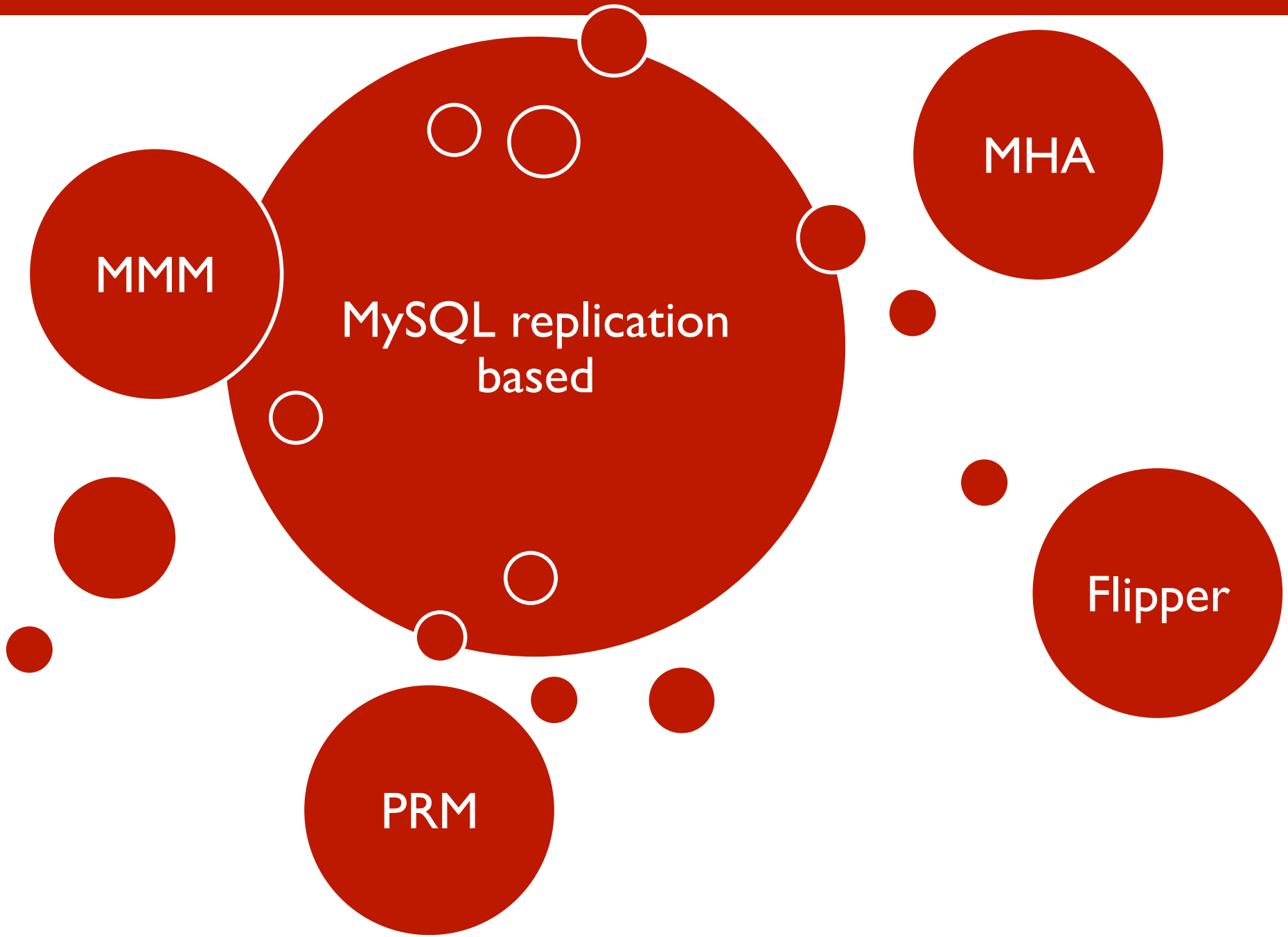
But you deal with consequences

Choice



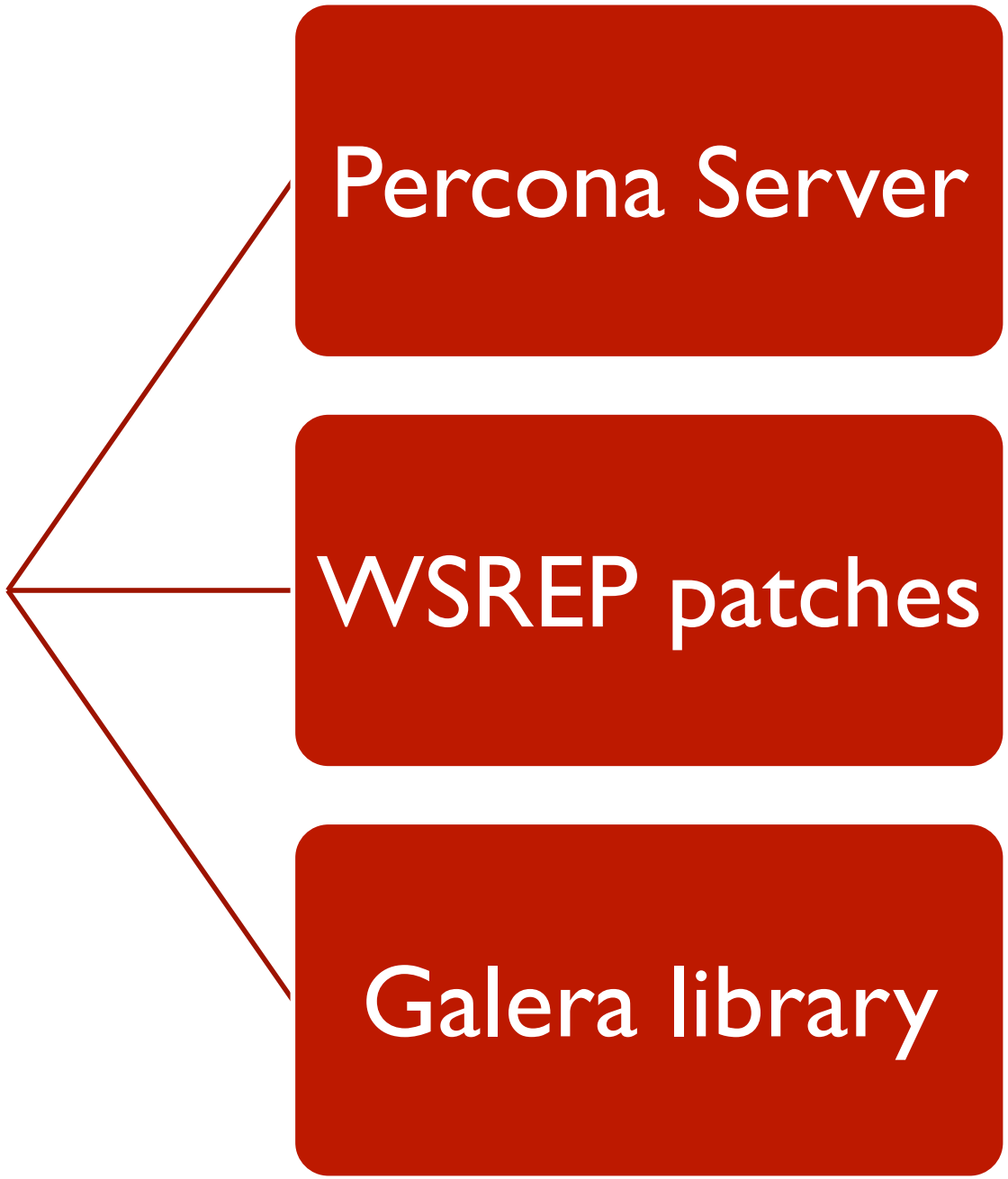
MySQL Replication:
Access to all systems

XtraDB Cluster:
Data consistency

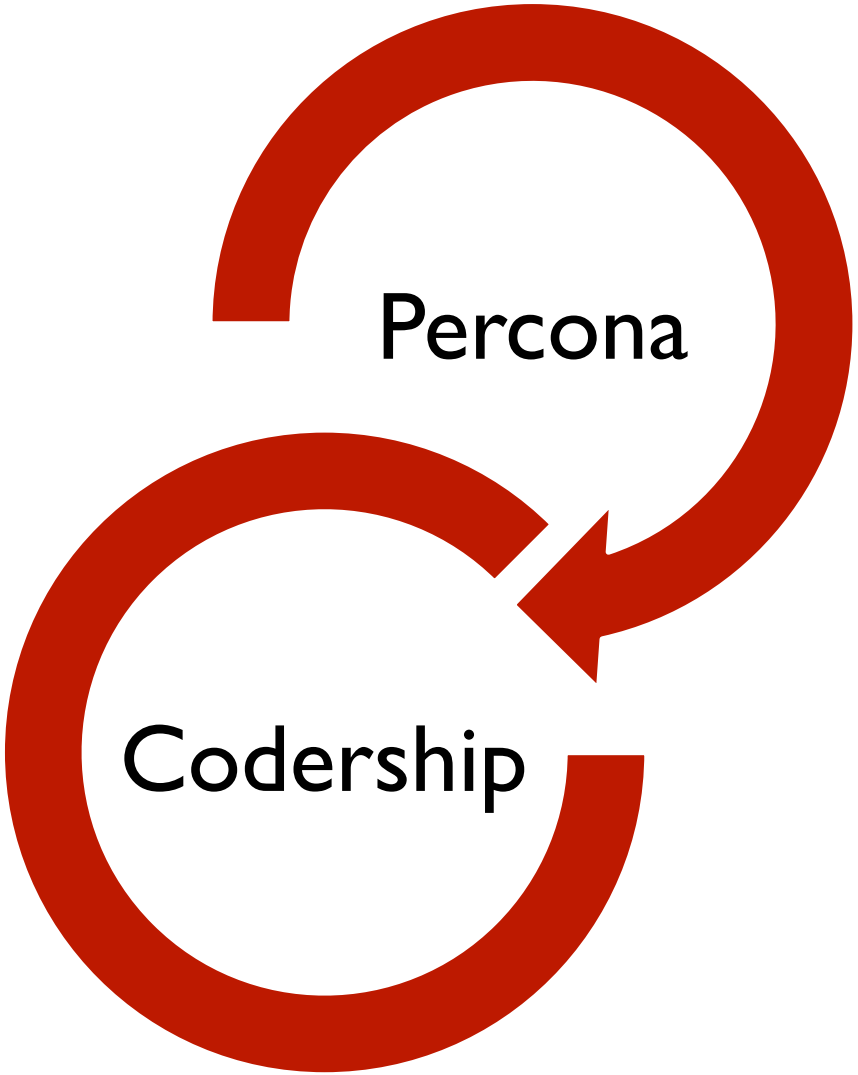

MMM
MySQL replication based
MHA
Flipper
PRM

Percona XtraDB Cluster details

Percona
XtraDB Cluster



Partnership



Full compatibility with existing systems

Minimal efforts to migrate

Minimal efforts to return back to MySQL

So, is this a perfect solution?

Limitations

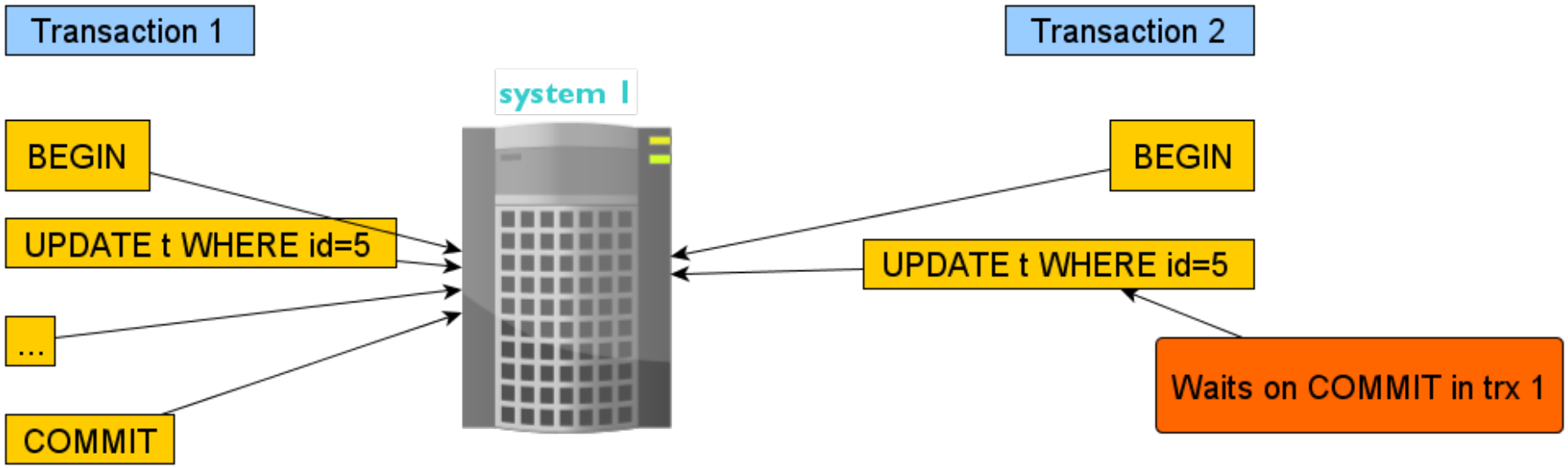
~~Only InnoDB tables are fully supported~~

MyISAM support is limited

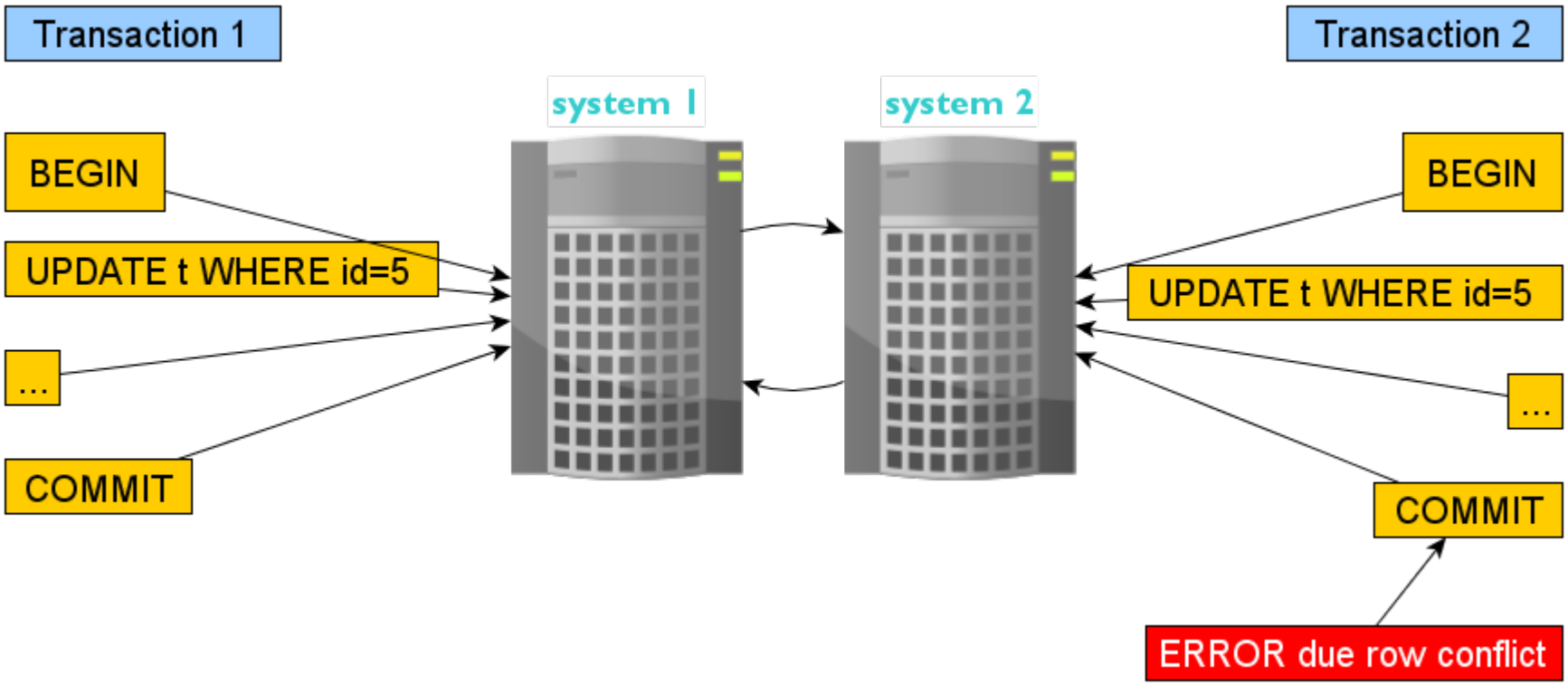
OPTIMISTIC locking for transactions on different servers

http://en.wikipedia.org/wiki/Optimistic_concurrency_control

Traditional locking

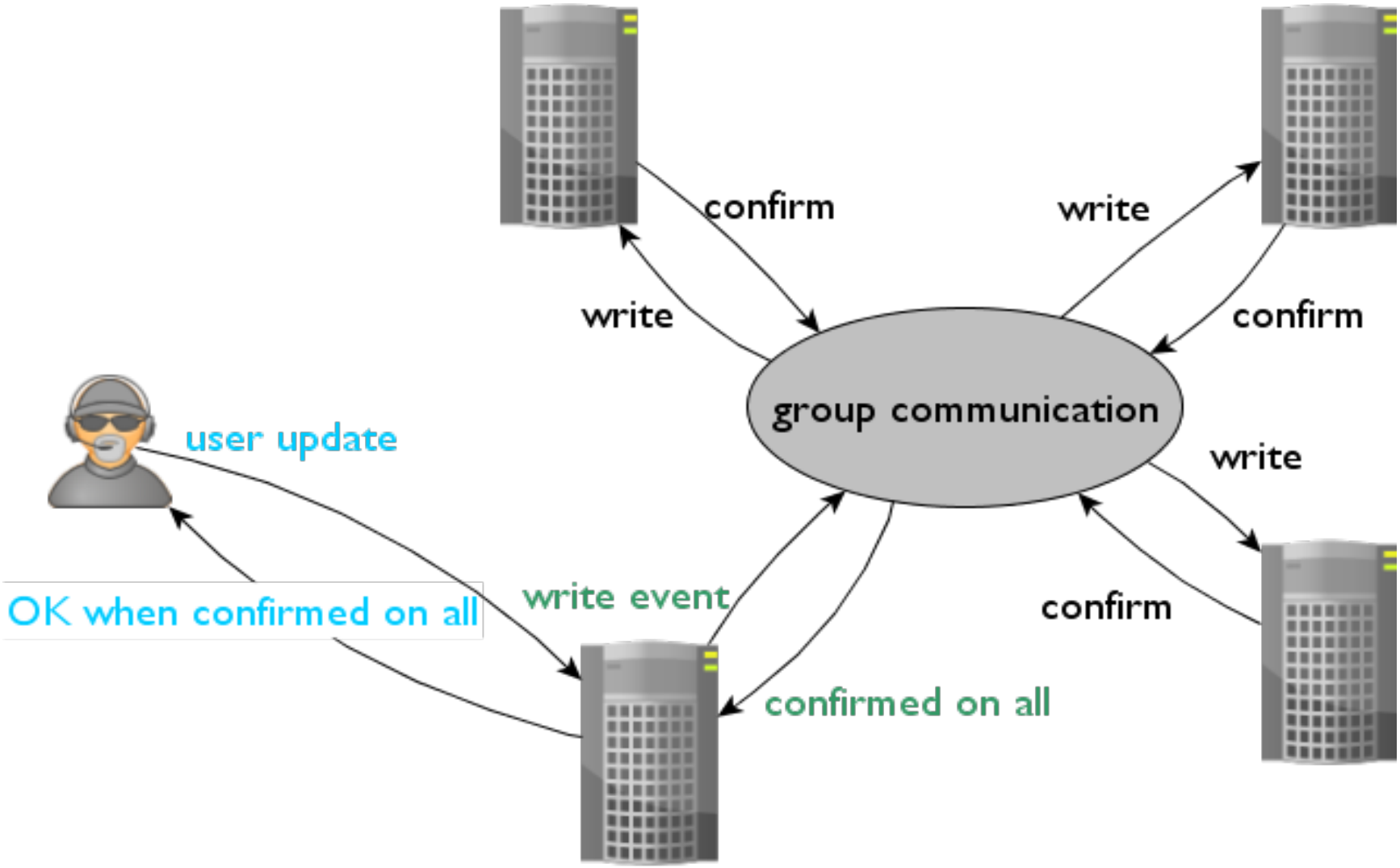


Optimistic locking



The write performance is limited by weakest node

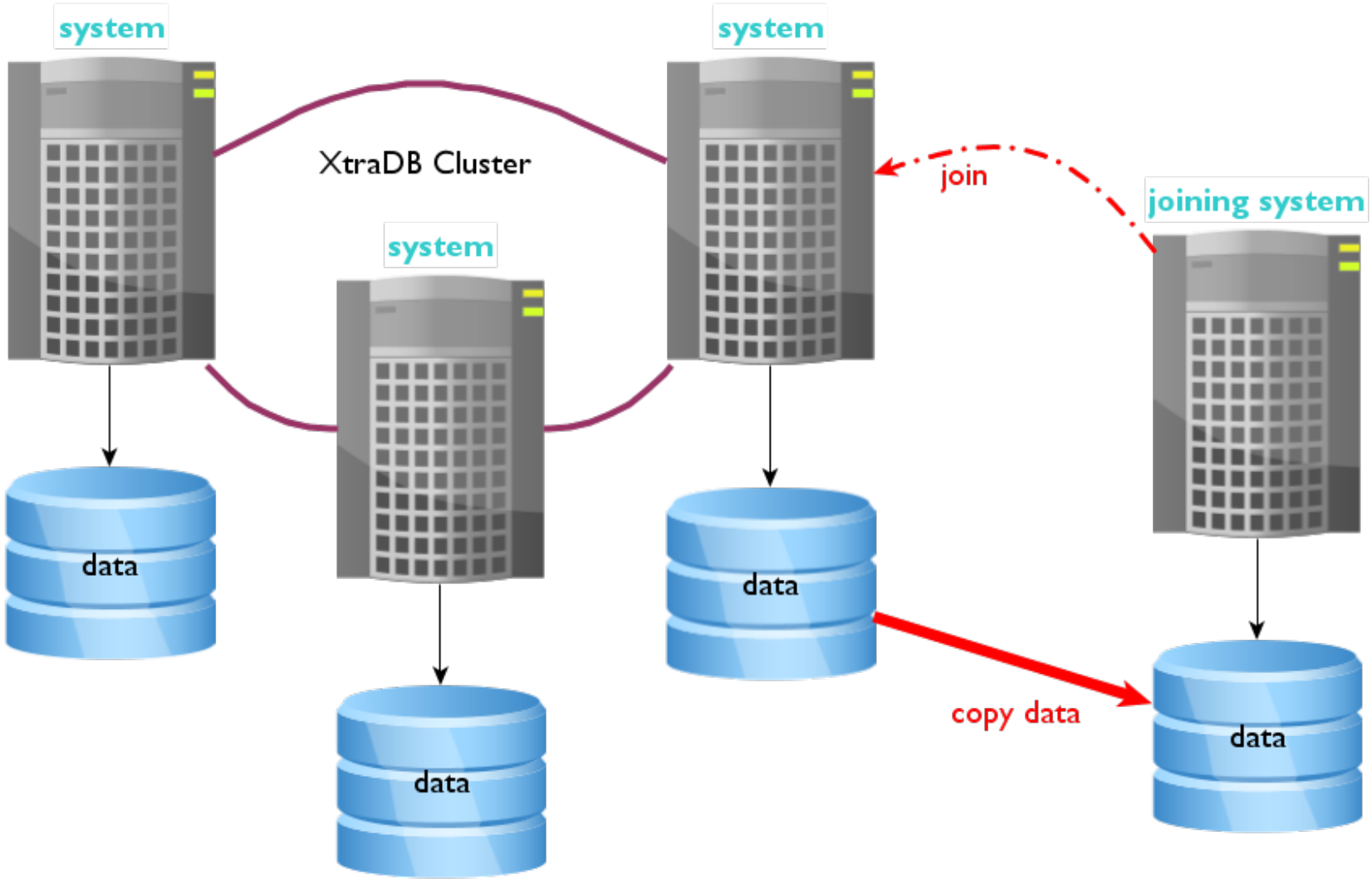
Write performance



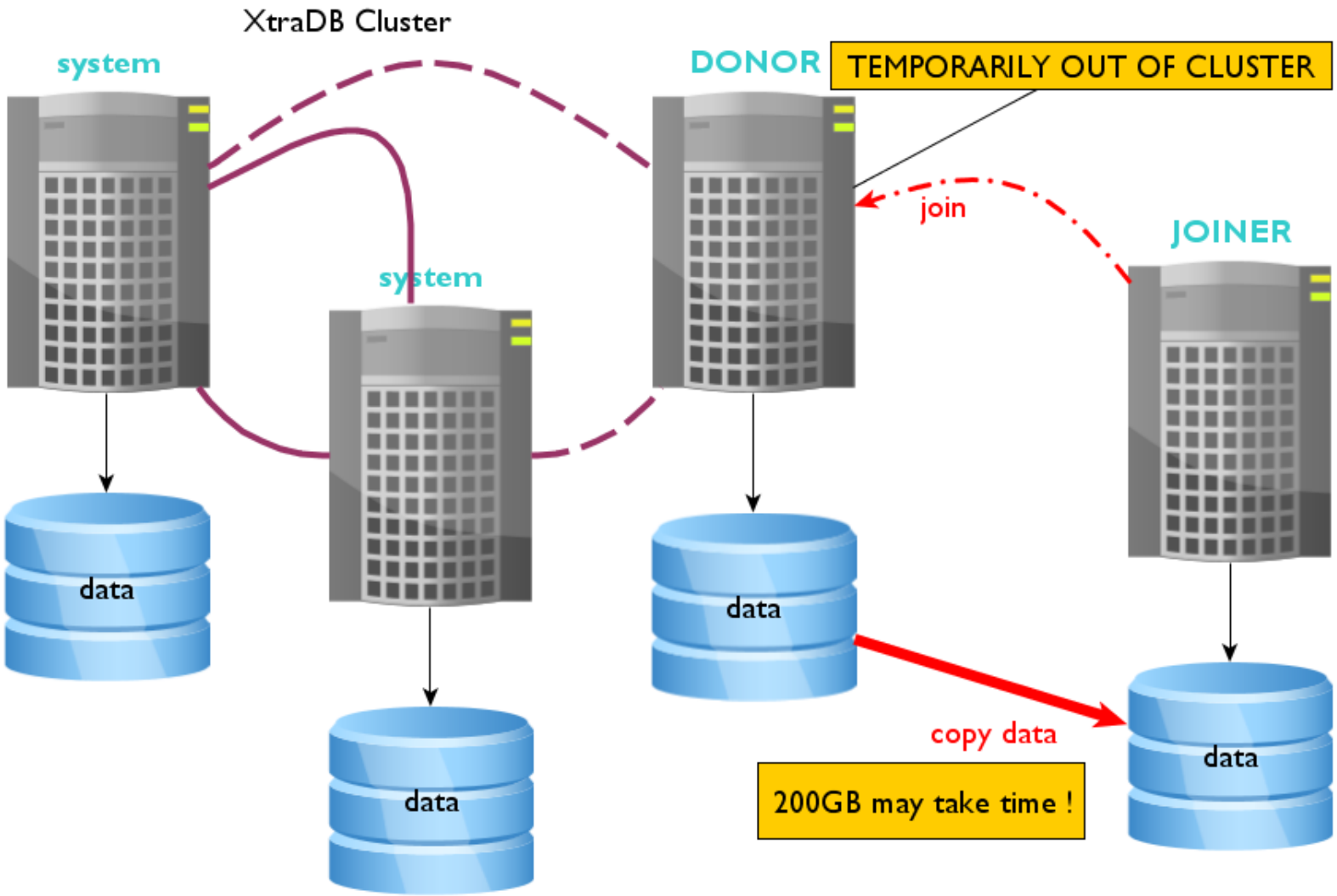
For write intensive applications there could be datasize limit per node

Not physical but logical

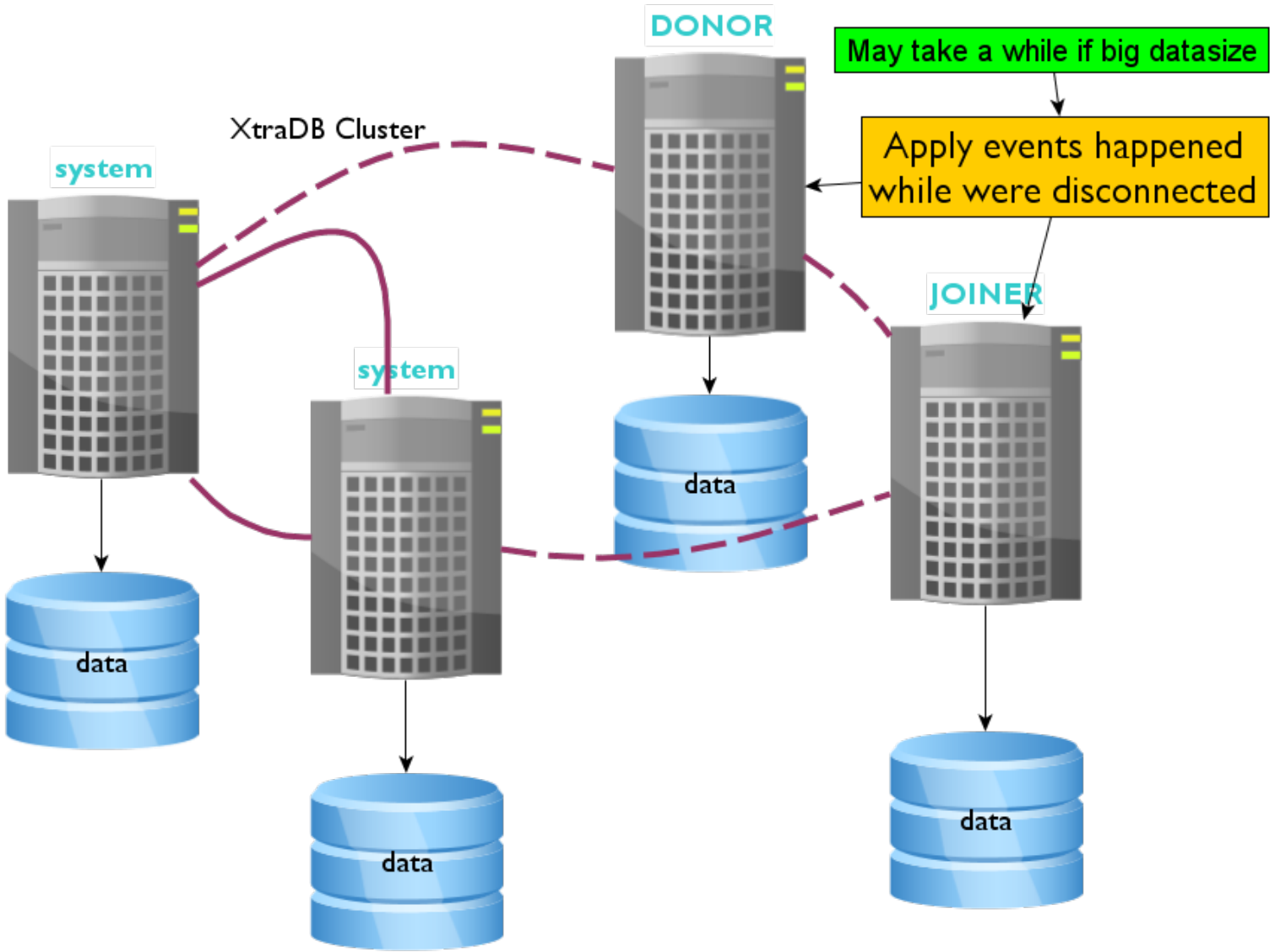
Join process. Step 1



Join process. Step 2

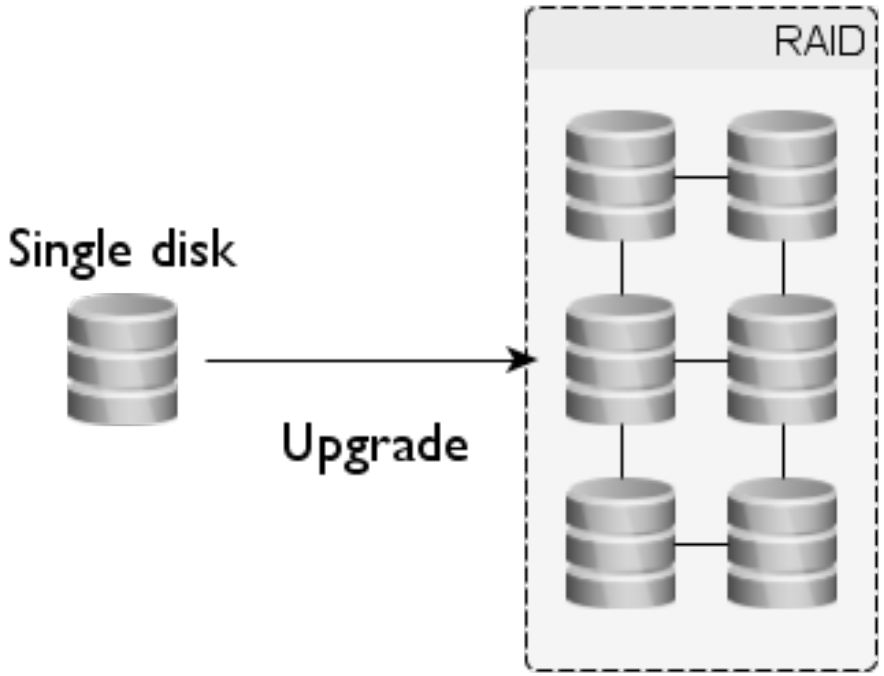


Join process: step 3

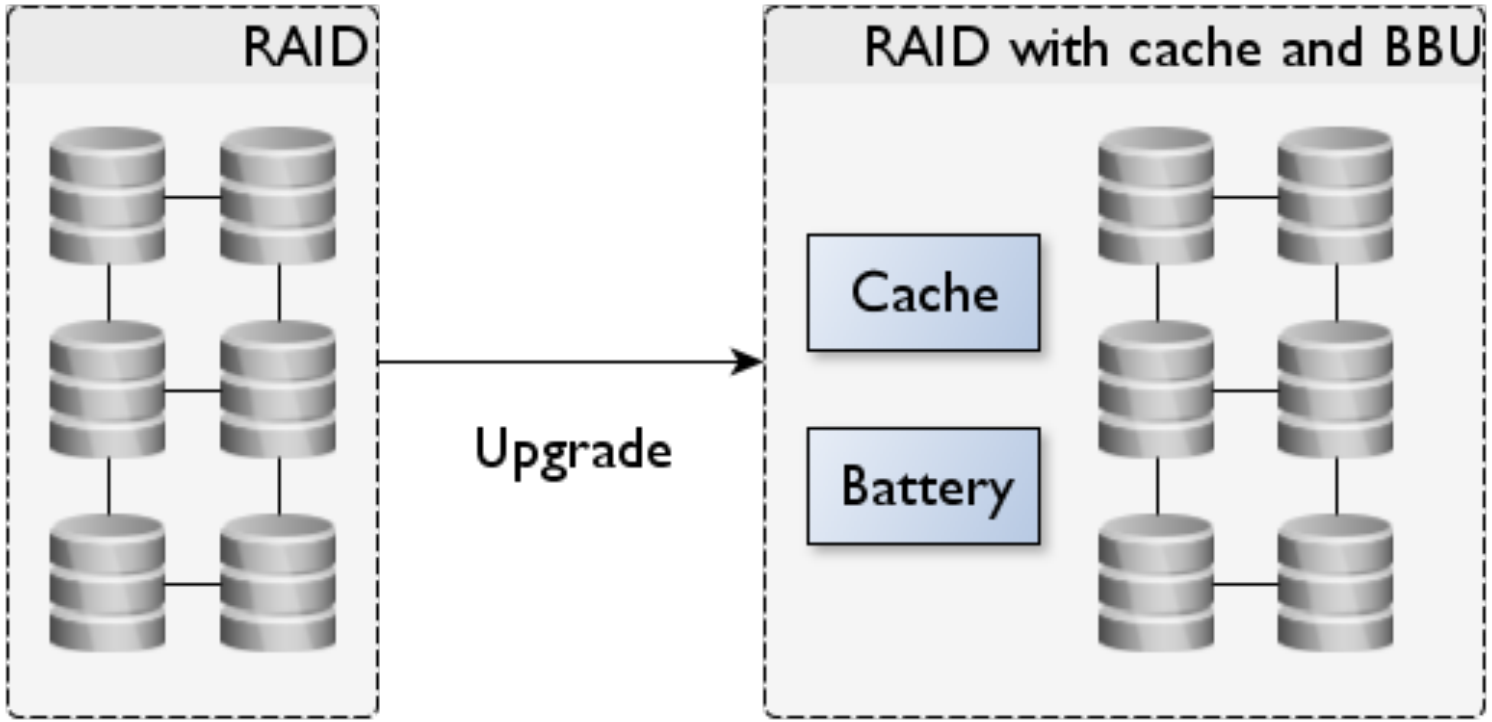


This is software + hardware solution

InnoDB write performance



InnoDB performance + ACID



Cluster performance

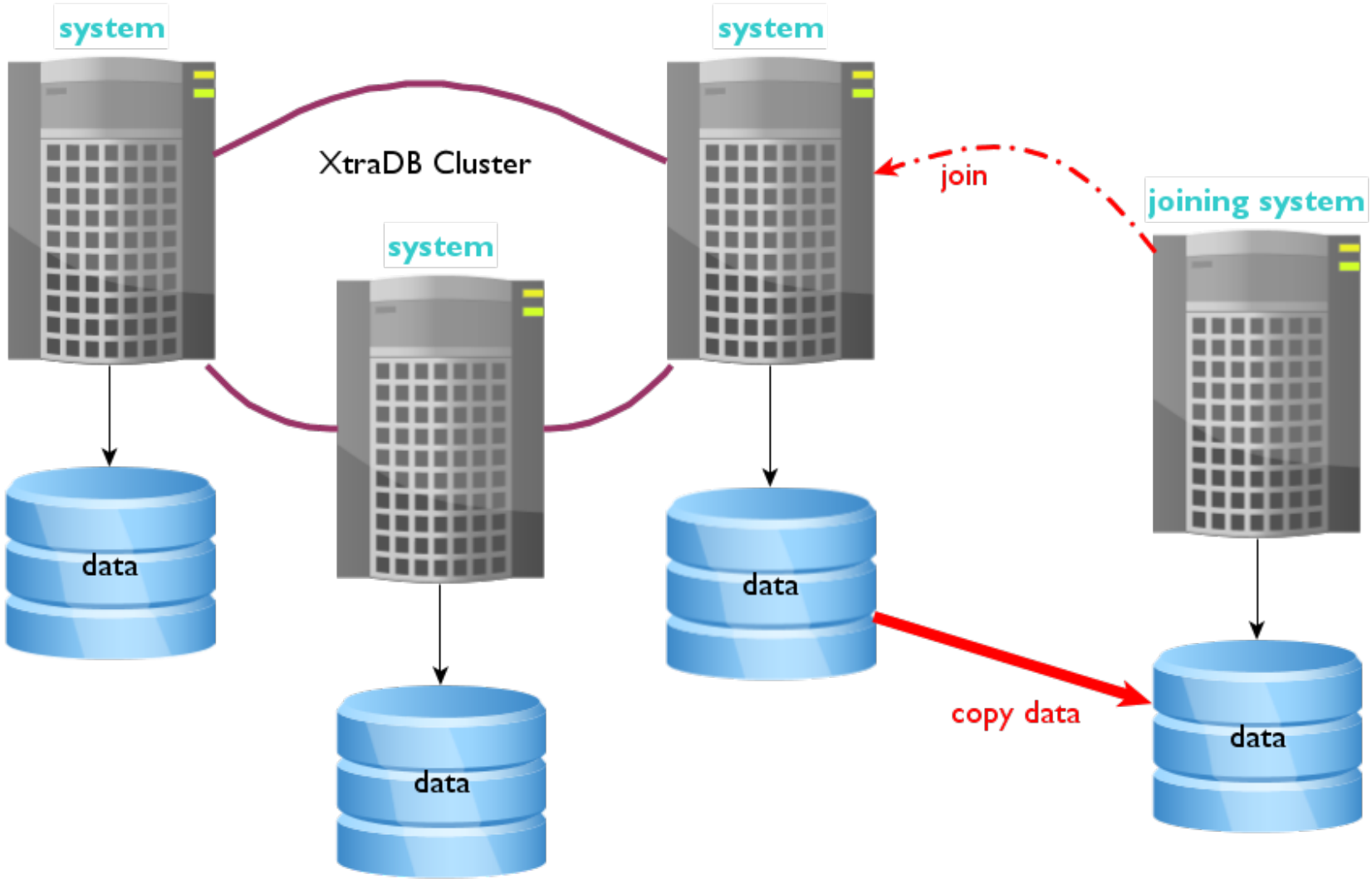
Network

- 10 GigE
- Infiniband

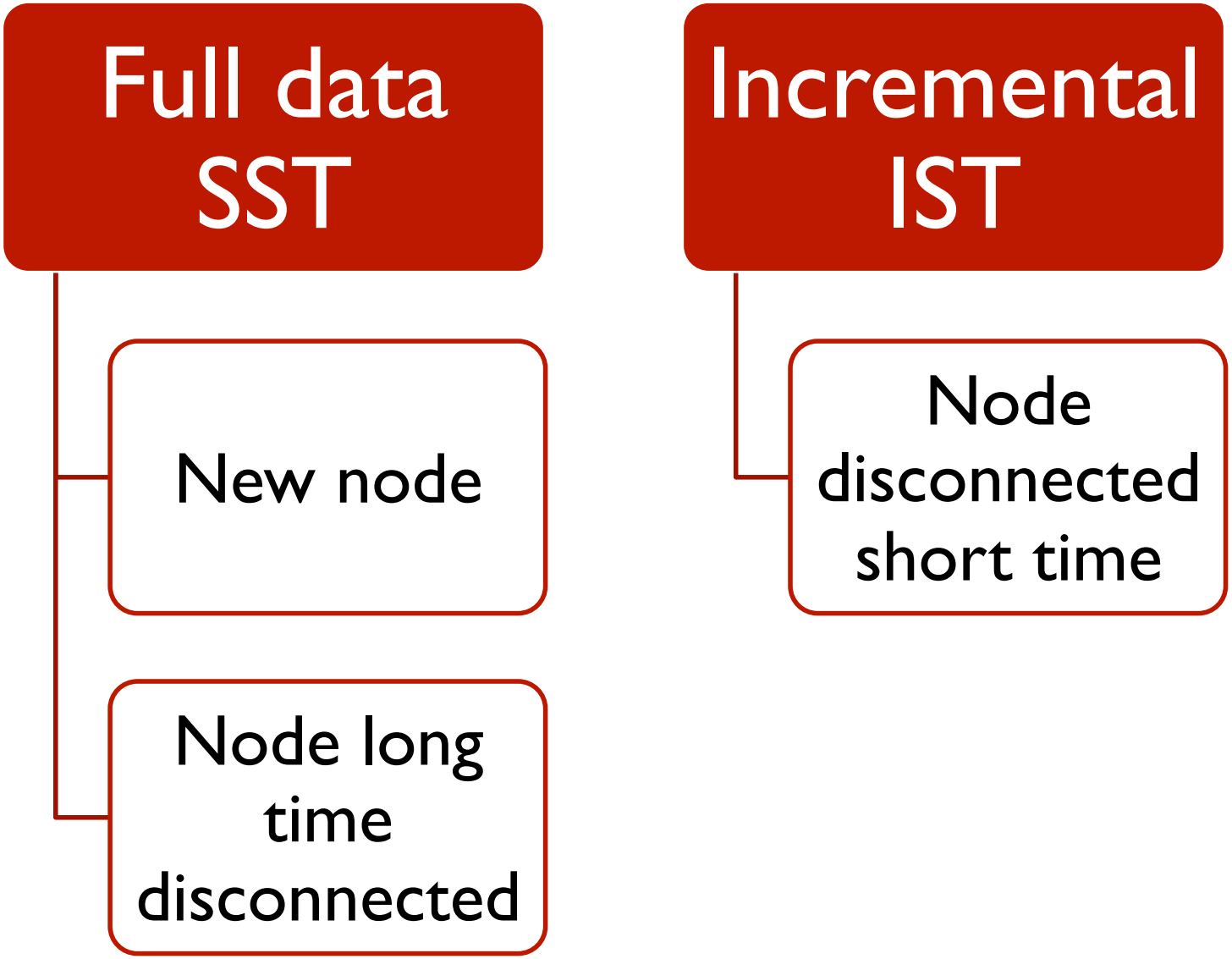
Storage

- SSD
- PCI-e Flash

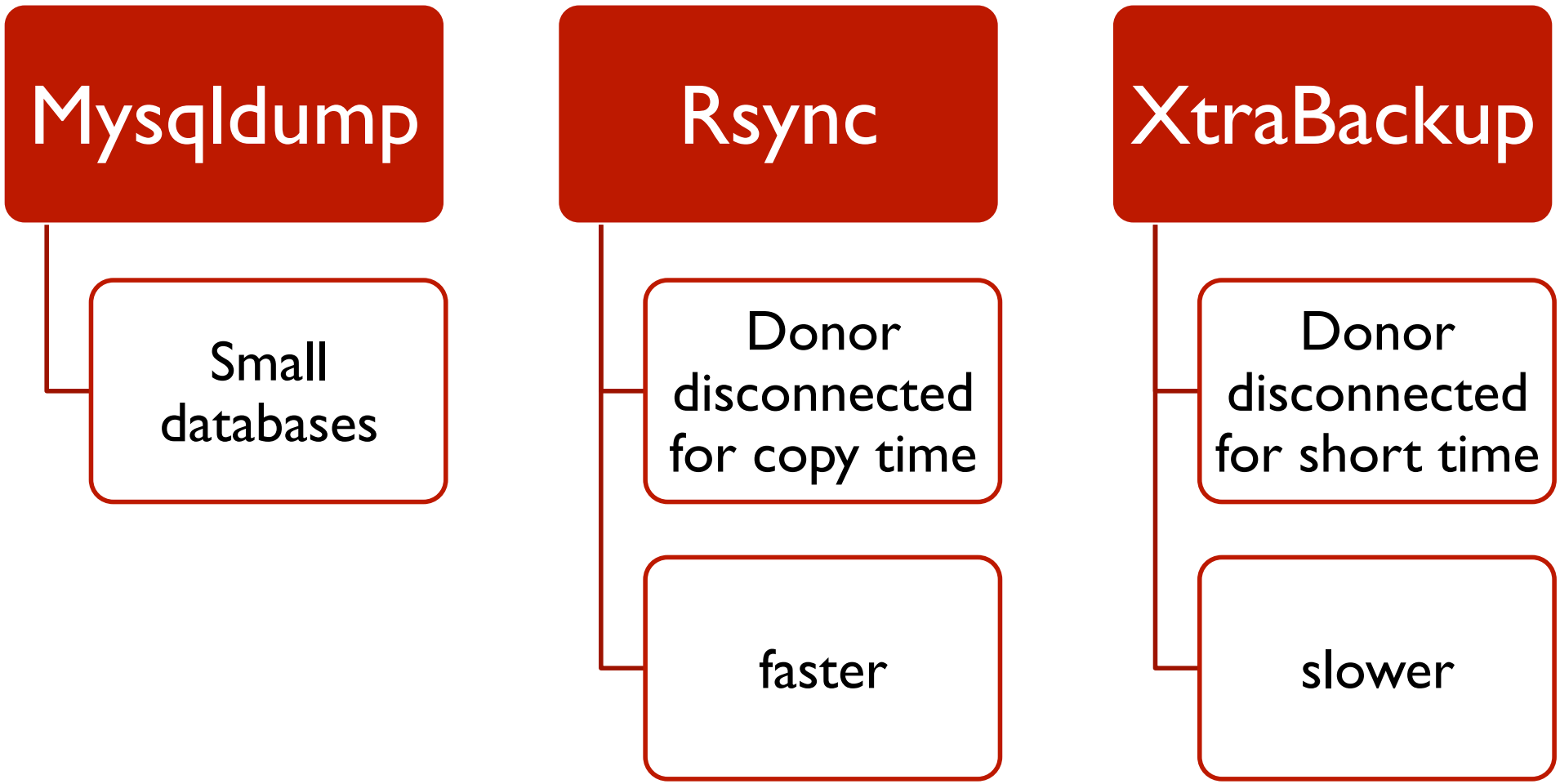
Join process



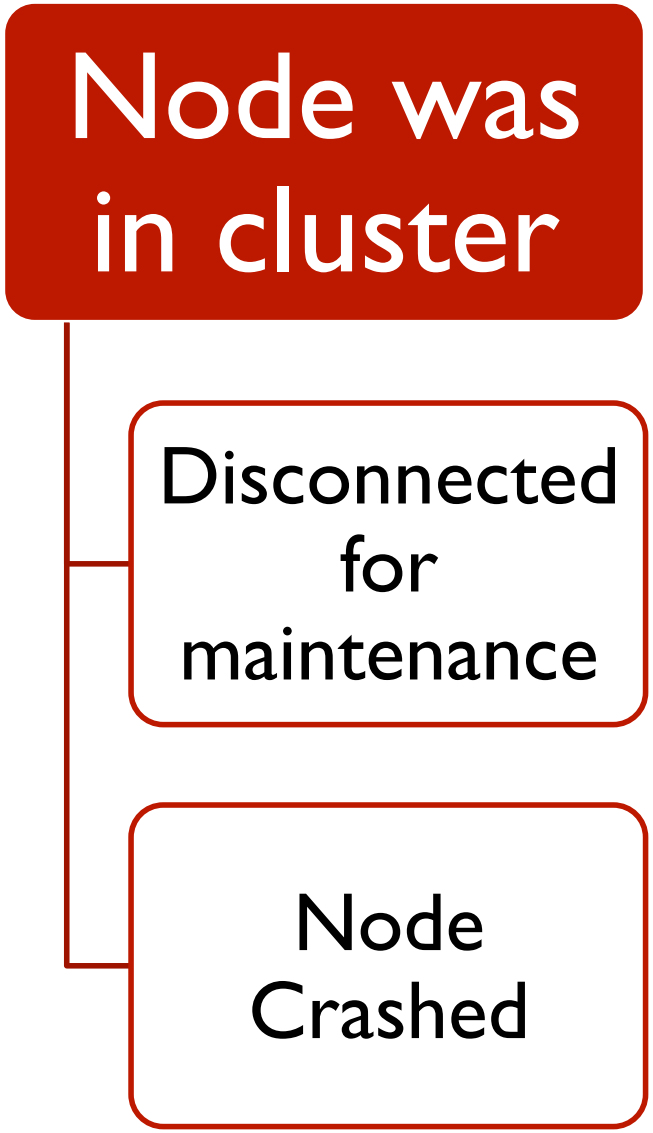
State Transfer



Snapshot State Transfer



Incremental State Transfer



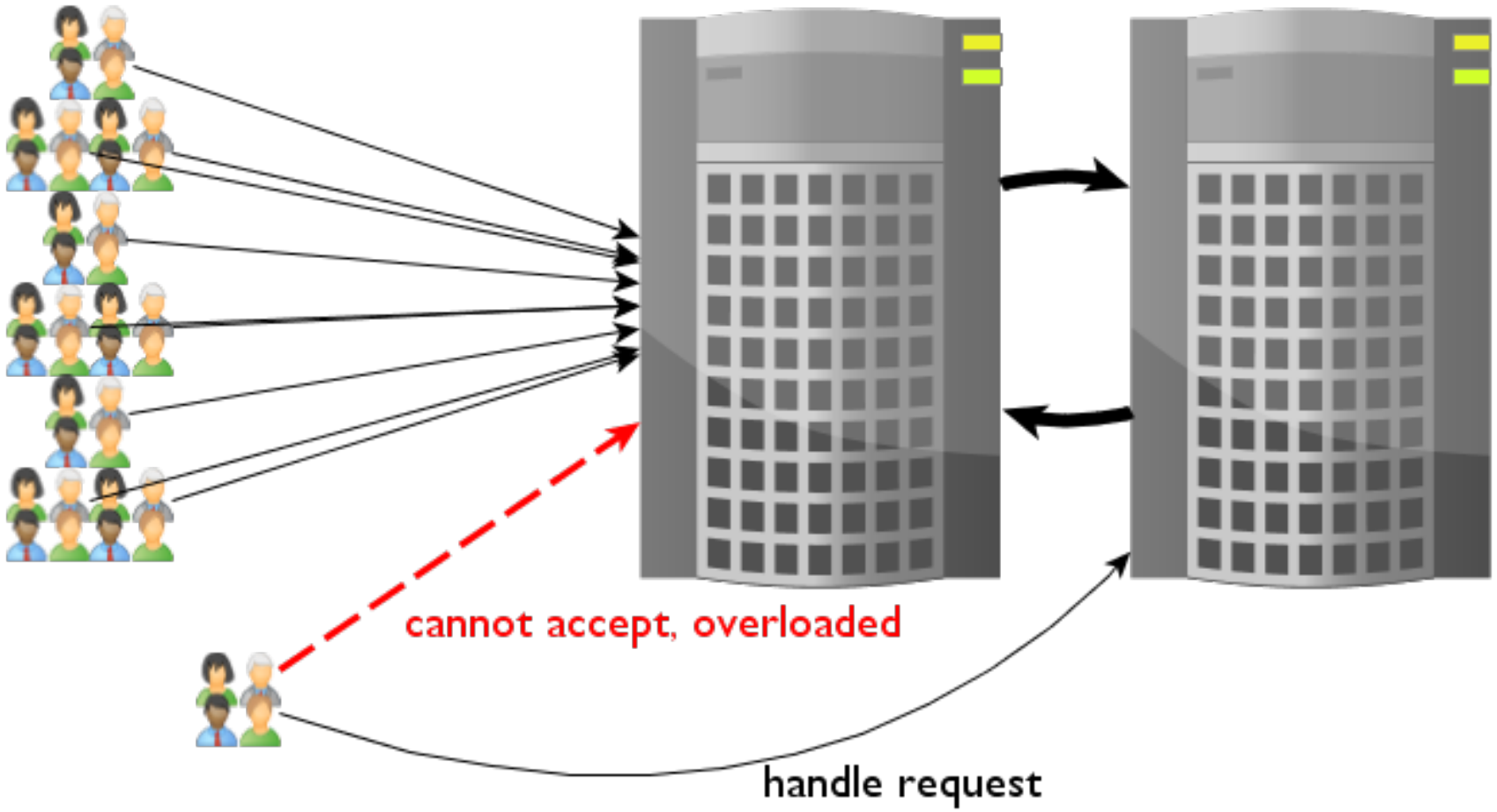
Scaleability

Scaleability

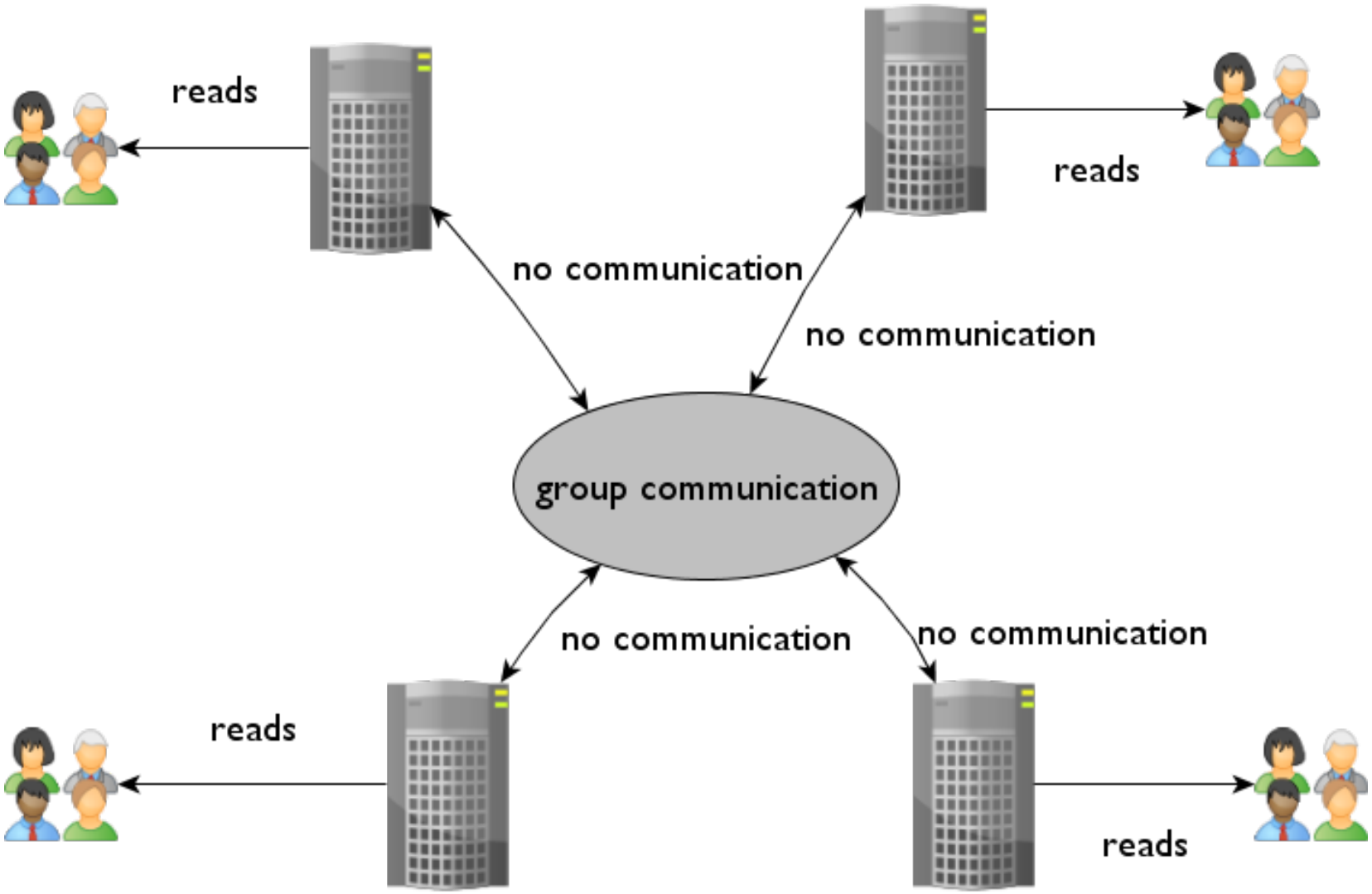
Scale ~ Ability

Ability to Scale

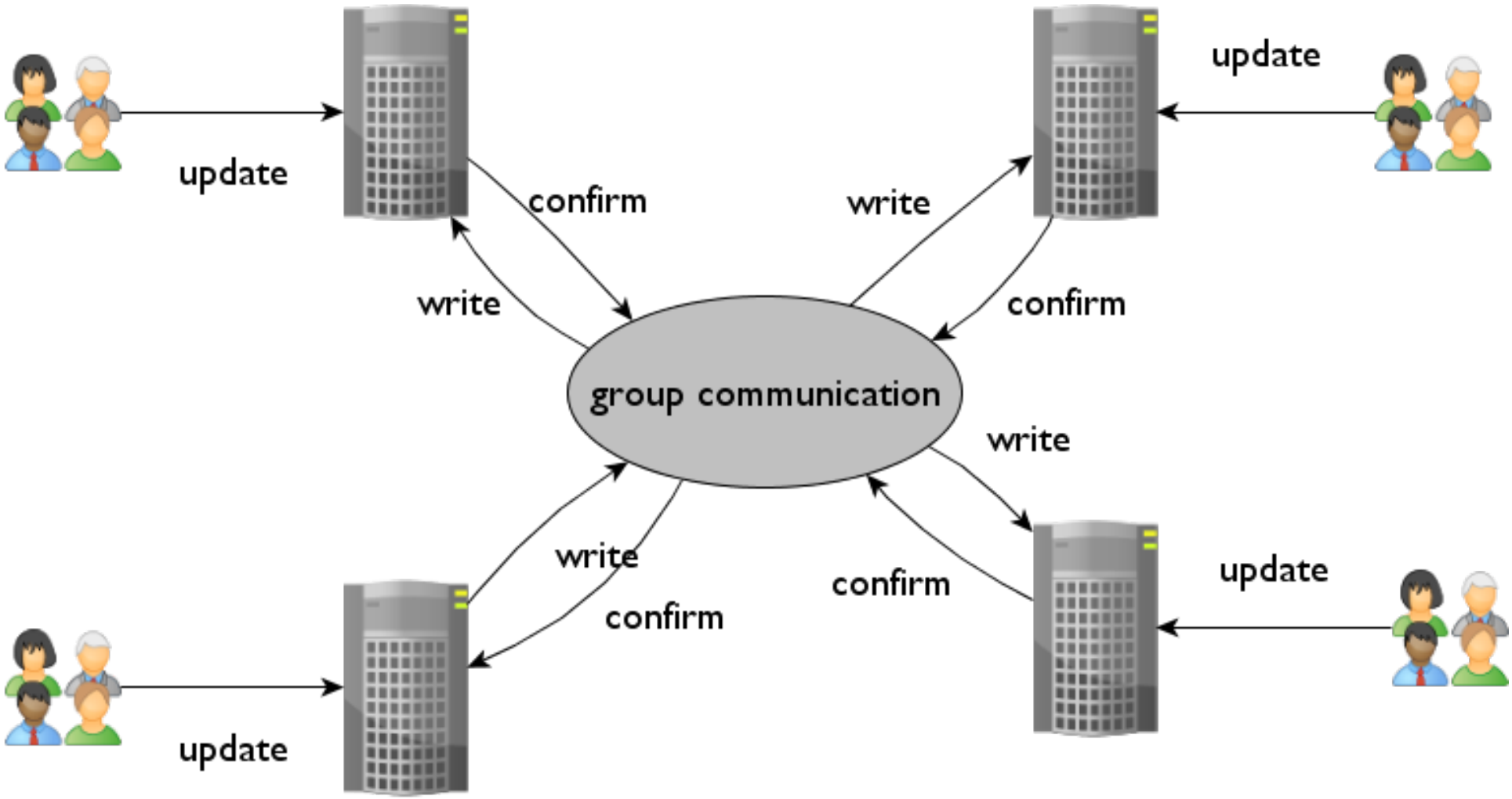
Scaleability is similar to availability



XtraDB Cluster: Reads scalability is easy



Write scalability is complicated



N servers scale to :

100% reads	• N factor
…	• …
50/50	• N/2 factor
…	• …
100% writes	• 1 or const

10 servers scale to :

100% reads

- 1 server: 100 q/s
- 10 servers: 1000 q/s

50/50

- 1 server: 100 q/s
- 10 servers: 500 q/s

100% writes

- 1 server: 100 q/s
- 10 servers: 100 q/s (can be more)

FAQ

Questions I am asked

- What happens if one node temporary unreachable ?
- How ALTER tables are handled
- What happens if someone runs update of 1000000 rows ?
- Show numbers on latency and throughput
- How connect node to a cluster ? Just show an example
- How cluster decides what nodes to keep in cluster and what to throw away
- Can I select a specific node as DONOR
- Load balancing ?
- XtraBackup SST – locking for short period
- How auto_increment is handled ?
- What is use case for XtraDB Cluster ?

It looks so easy. Why did not you implement it earlier?

It is not easy.

Computer science of **group communication** and **distributed transactions**.

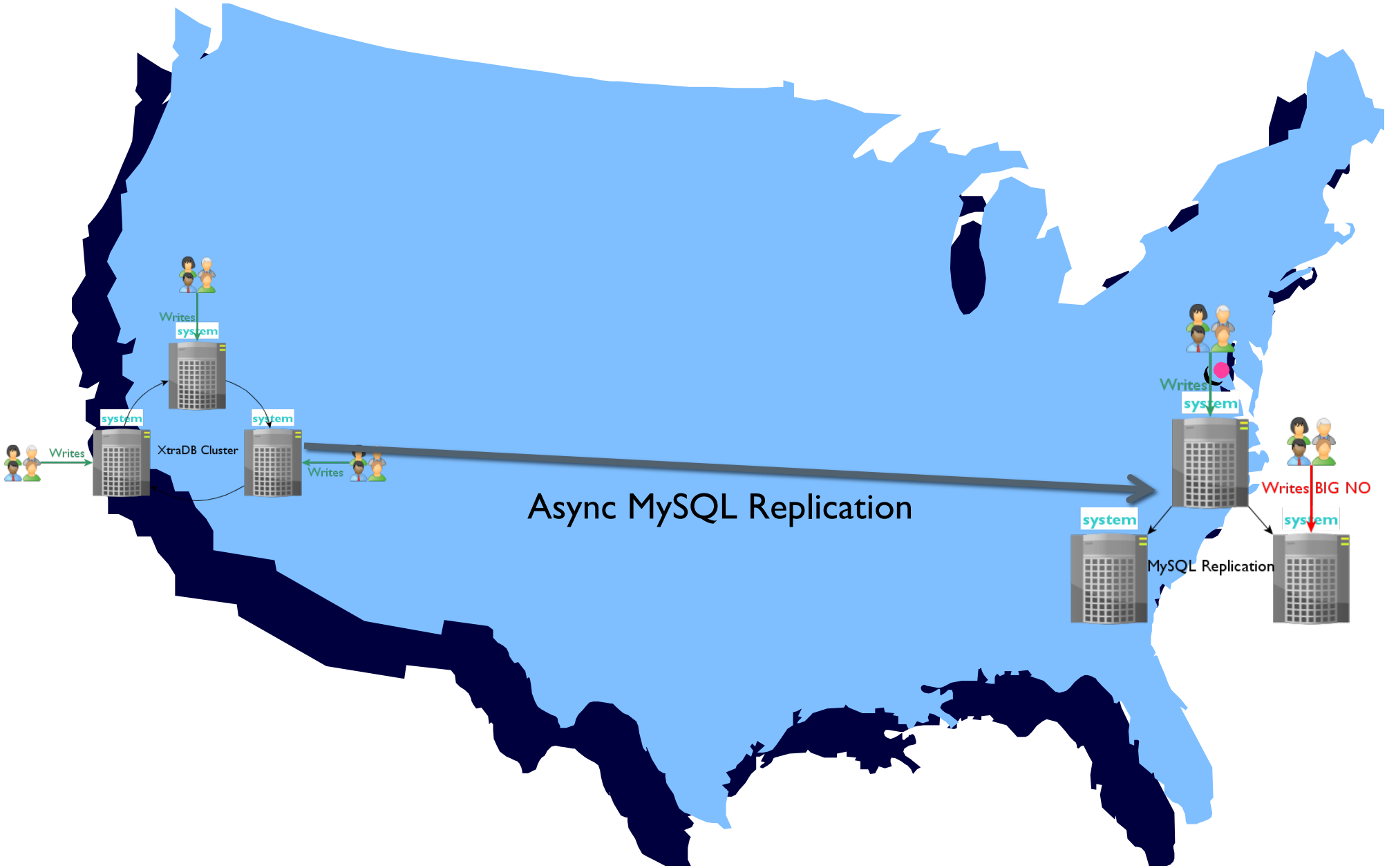
Credits to Codership Oy

How fast is it?

Reasonably fast.

Can I replicate XtraDB Cluster to MySQL Replication?

Yes


Writes
system
system
Writes
XtraDB Cluster
system
Writes
Async MySQL Replication
Writes
system
Writes BIG NO
system
system
MySQL Replication

Would I install it on a production system?

gutefrage.net

Germany's No. 3 Web Destination

http://www.percona.com/about-us/pressreleases/gutefragenet-standardizes-percona-xtradb-cluster

95% reads

How it is compared to MySQL Cluster?

It is different

	XtraDB Cluster	MySQL Cluster
Easy to migrate	✓	
Easy to use	✓	
Cloud / EC2	✓	
Changes in an application		✓
Write scaling		✓
99.999%		✓

More questions

- What happens if one node temporary unreachable ?
- How ALTER tables are handled
- What happens if someone runs update of 1000000 rows ?
- How cluster decides what nodes to keep in cluster and what to throw away
- Can I select a specific node as DONOR
- Load balancing ?
- How auto_increment is handled ?

Resources

- http://www.percona.com/software/percona-xtradb-cluster/
- http://www.codership.com/wiki/doku.php
- Virtual synchrony
  - http://en.wikipedia.org/wiki/Virtual_synchrony
- CAP Theorem
  - http://en.wikipedia.org/wiki/CAP_theorem
- Optimistic locking
  - http://en.wikipedia.org/wiki/Optimistic_concurrency_control

Credits

- WSREP patches and Galera library is developed by Codership Oy

Thank you!

Questions ?

You can try Percona XtraDB Cluster today!