Openstack and MySQL HA

Dimitri Vanoverbeke

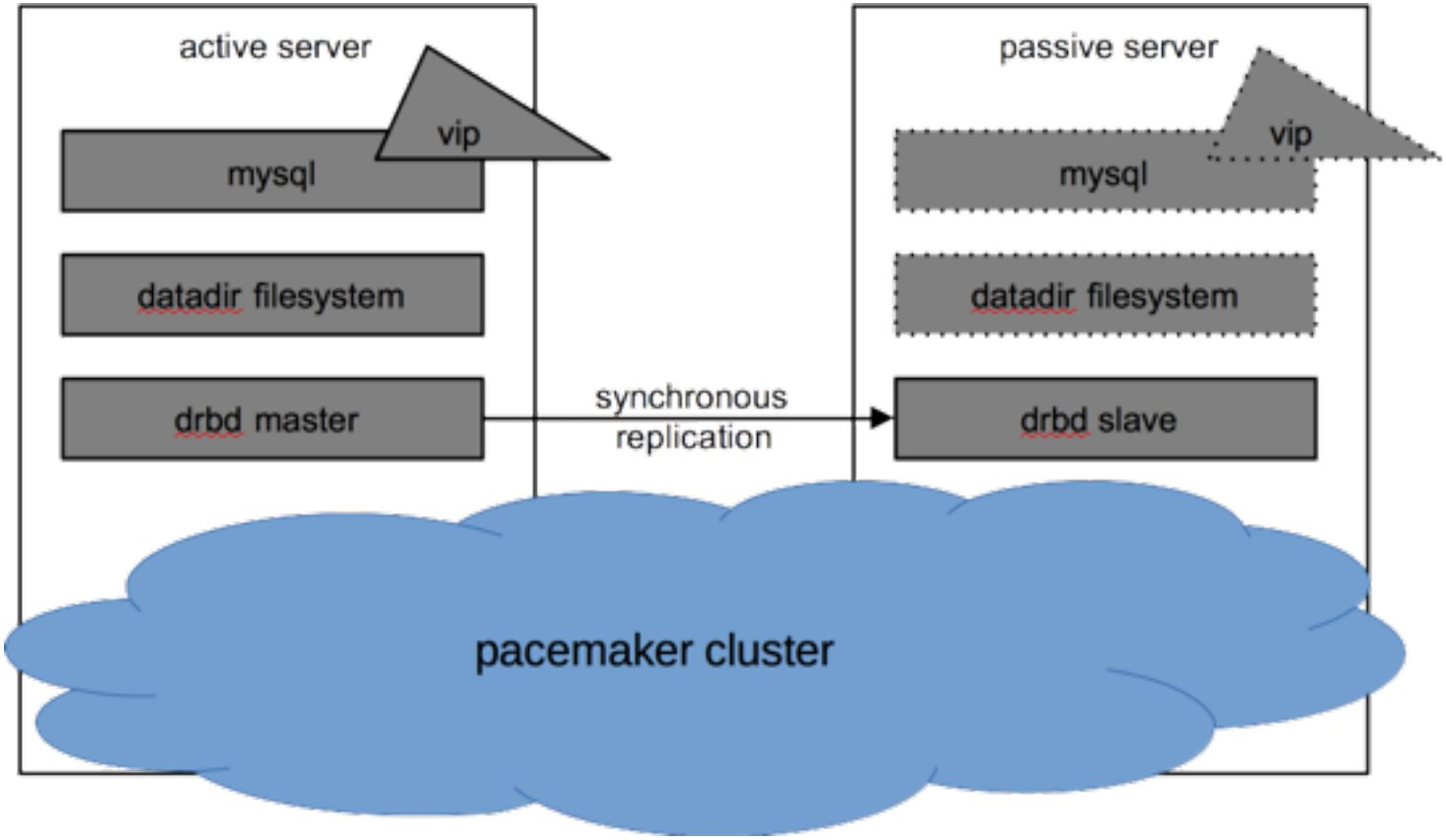
MySQL usage in OpenStack

- Metadata storage for Neutron, Nova and many other layers of the Open Stack solution.
- Not HA by design, single point of failure is MySQL.
- Data loss is problematic.
- Data volume is difficult to predict.

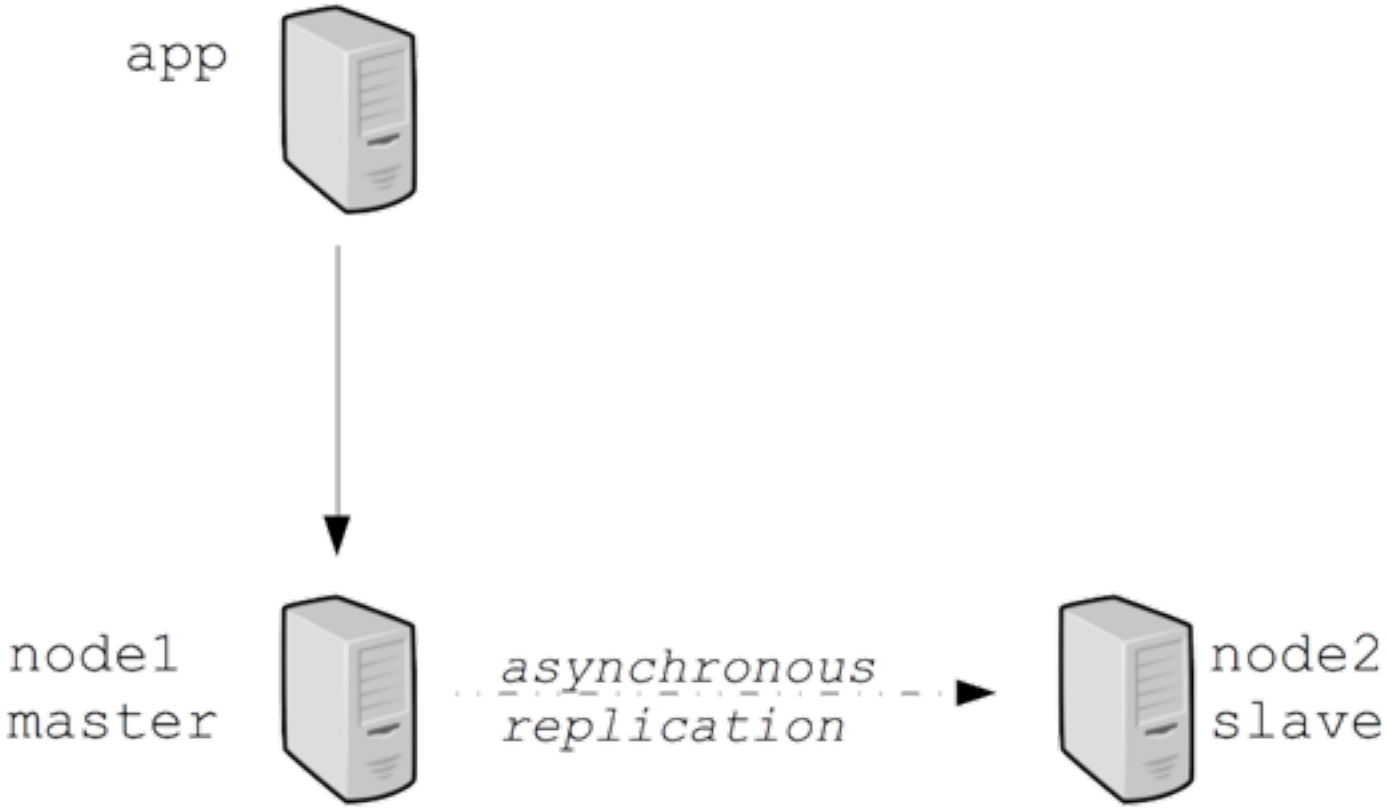
The principles of HA

- Redundancy of services/data
- Reducing Failure Impact
- Failover on a problem/maintenance
- Recovery from loss of redundancy

MySQL DRBD / pacemaker



Standard asynchronous replication



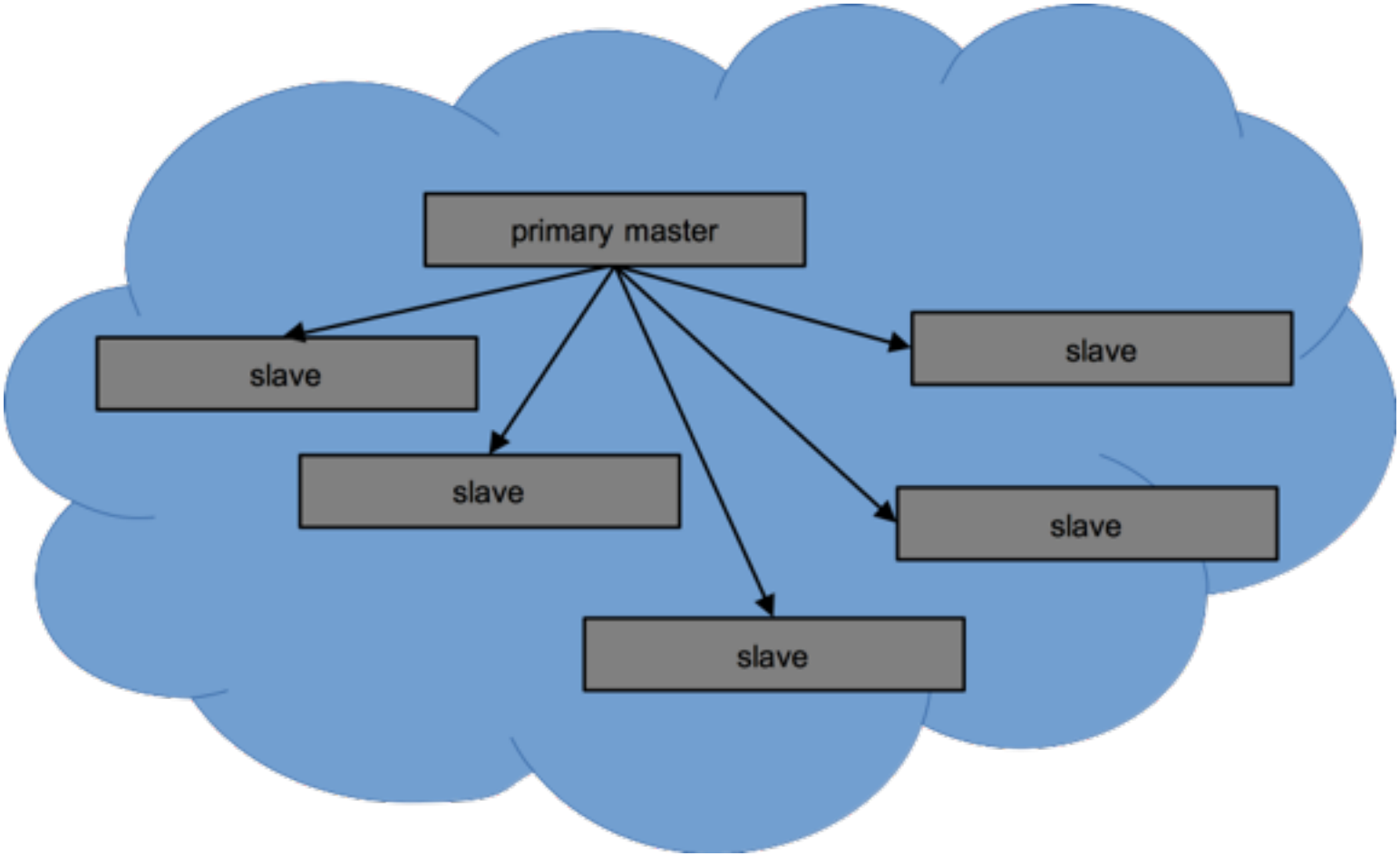
Asynchronous replication

- Slaves stream replication log from single master
- Slaves can be lagged, no delivery guarantees
- except semi-sync
- Provisioning of slaves is manual
- Consistency of slaves is assumed, manual
- Complex topologies can be built
- Circular replication possible
- Writes safe only from a single global master

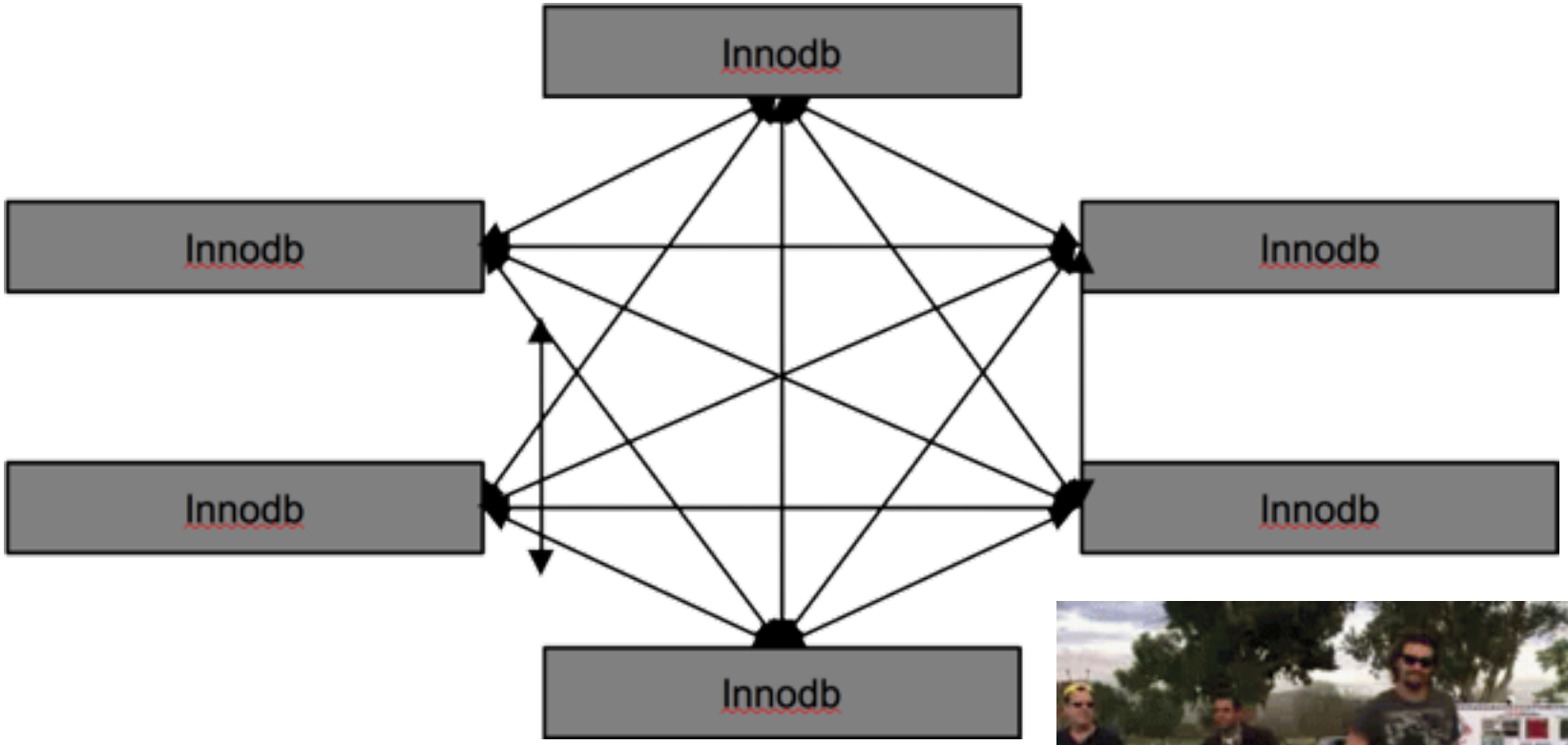
Replication based HA solutions

- Lots of tech in this area
- MHA and PRM are probably the lead contenders for a generic system, MMM can be considered as obsolete.
- MHA
  - Simple, centralized controller
  - Well tested slave resync code, new 5.6 GTID support
  - SSH + Perl duct tape
- PRM
  - Distributed quorum-based system (Pacemaker)
  - New slave resync code, 5.6 GTID support
  - Shell-based Pacemaker resource agent

PRM (Percona Replication Manager)



Percona XtraDB cluster



Percona Server

This is a free open source solution, Percona Server is a MySQL alternative which offers breakthrough performance, scalability, features, and instrumentation. Self-tuning algorithms and support for extremely high-performance hardware make it the clear choice for organizations that demand excellent performance and reliability from their MySQL database server.

PERCONA
SERVER

Replication layer

WSREP API is a project which goal was to develop a generic replication plugin interface for databases

Galera is wsrep provider that implements true multi-master virtually synchronous replication

Xtrabackup

- Backup solution for performing hot binary backups
- Incremental backup support
- No Locking backups on InnoDB
- Point in time recovery
- Streaming backup to SSH
- Streaming backups o Swift

Features of Percona XtraDB cluster

Full Compatibility with Existing Systems
Minimal efforts to **migrate**
Minimal efforts to **return back**

Synchronous Replication

Parallel Applying on all slaves

Multi Master Replication

Data Consistency

Automatic node provisioning

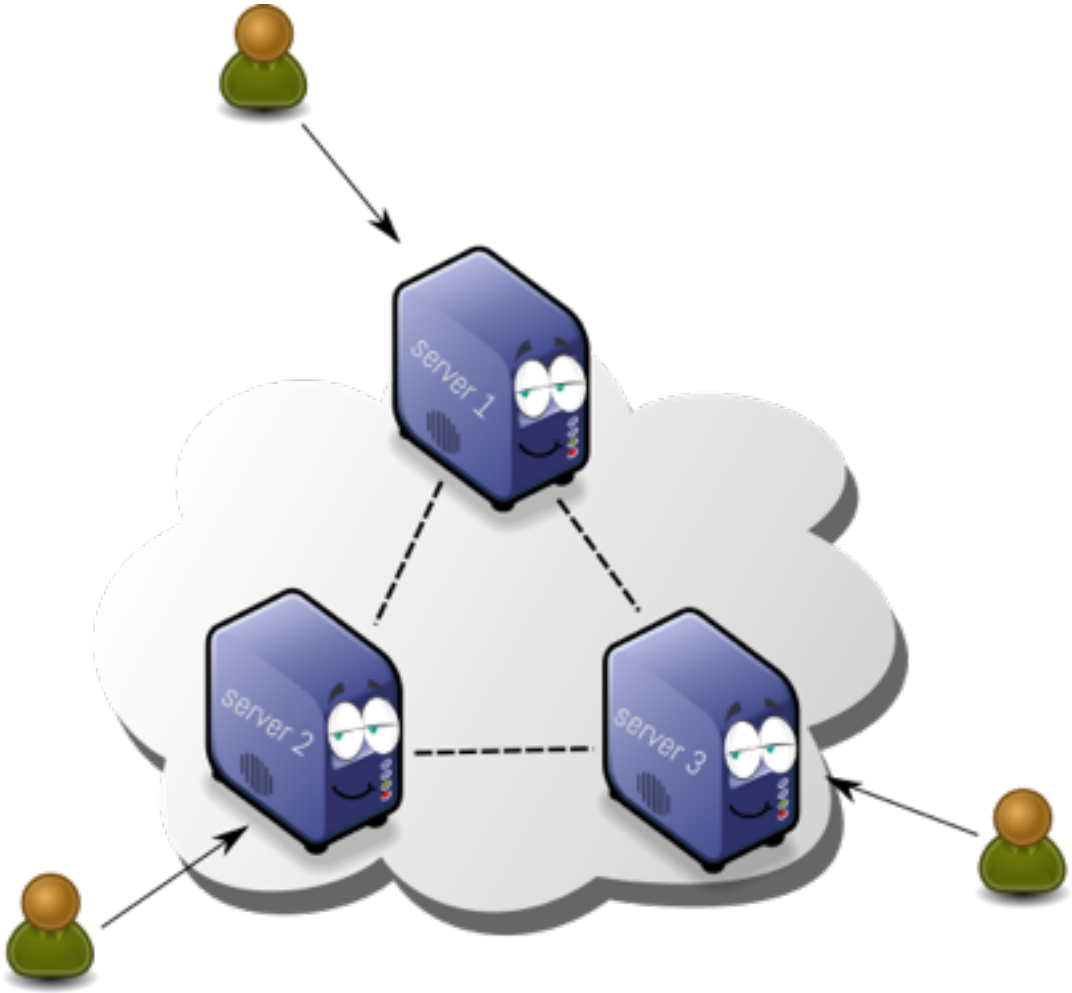
Galera overview

- Every node has a full copy
  - no sharding/scaling
- Cluster maintains its own state with Quorum-based rules
- Replication is different
  - Every node can read and write
  - Conflicts are pushed to application layer
  - Nodes are not generally permitted to fall behind
- Cluster provisions nodes automatically
- No direct VIP/Load balancing/etc management

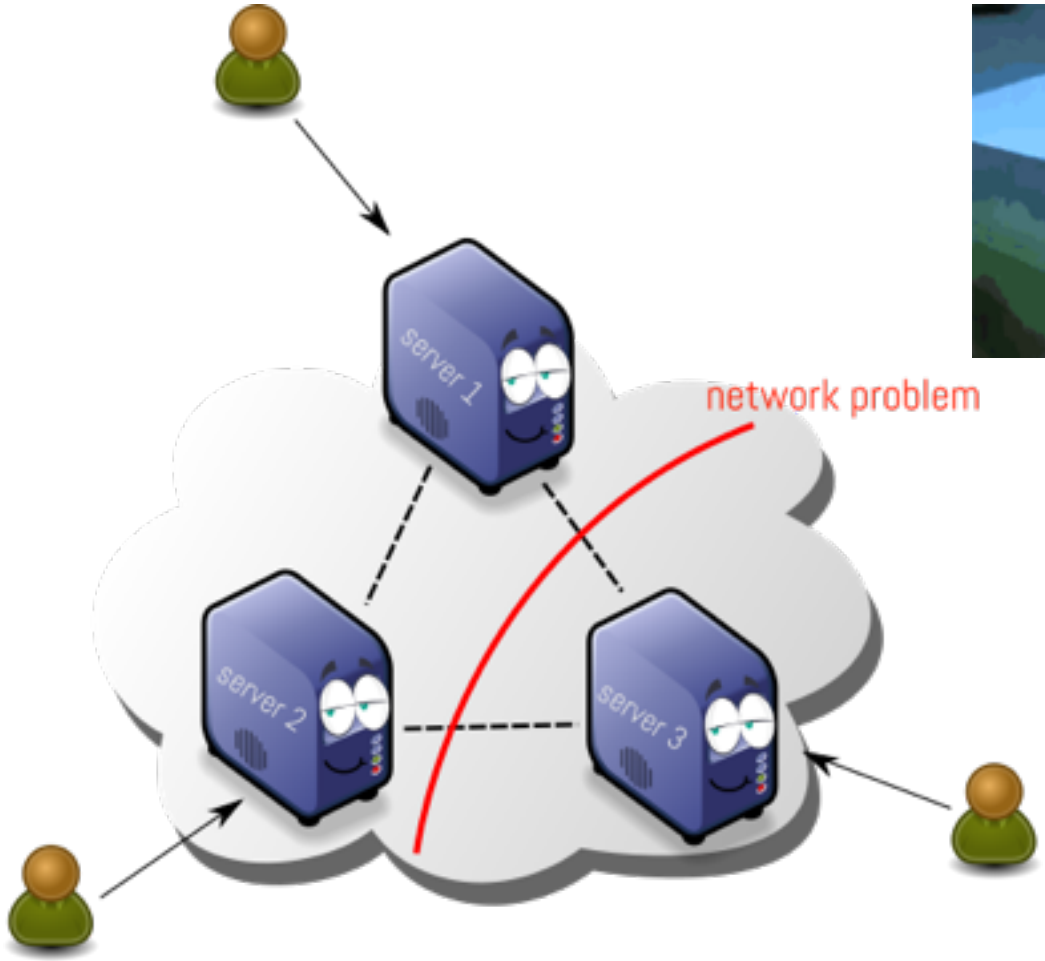
Node types

- Data nodes
  - mysqld
  - galera (shared library)
- Arbitrator nodes
  - just Galera
  - Replication
- No controllers, adding or confirmation nodes

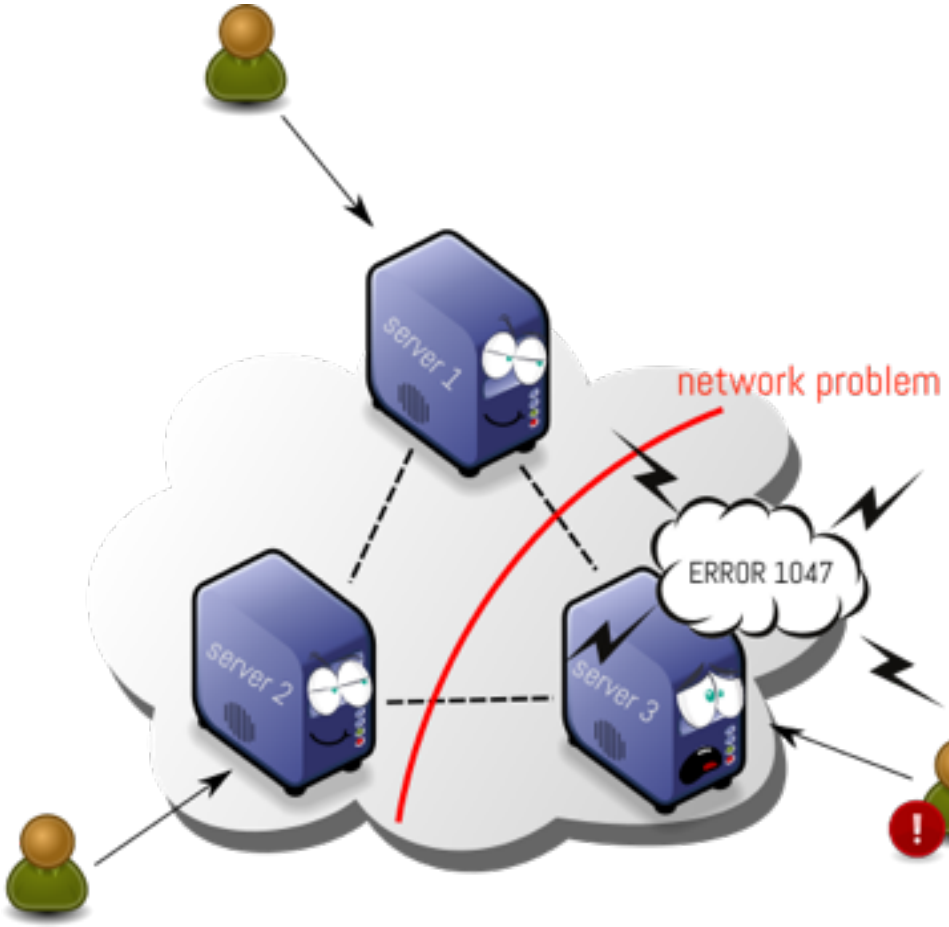
Majority HA



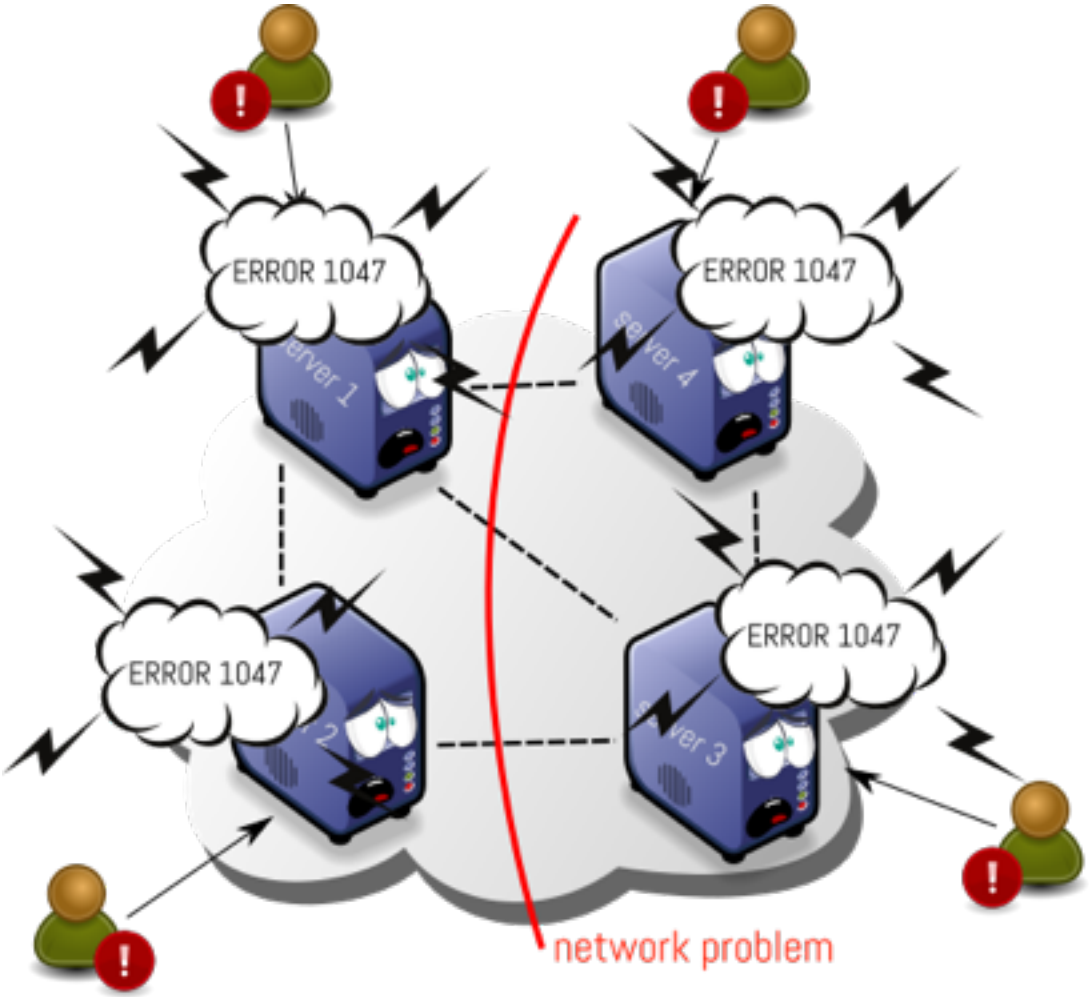
Loss of connectivity to 1 node



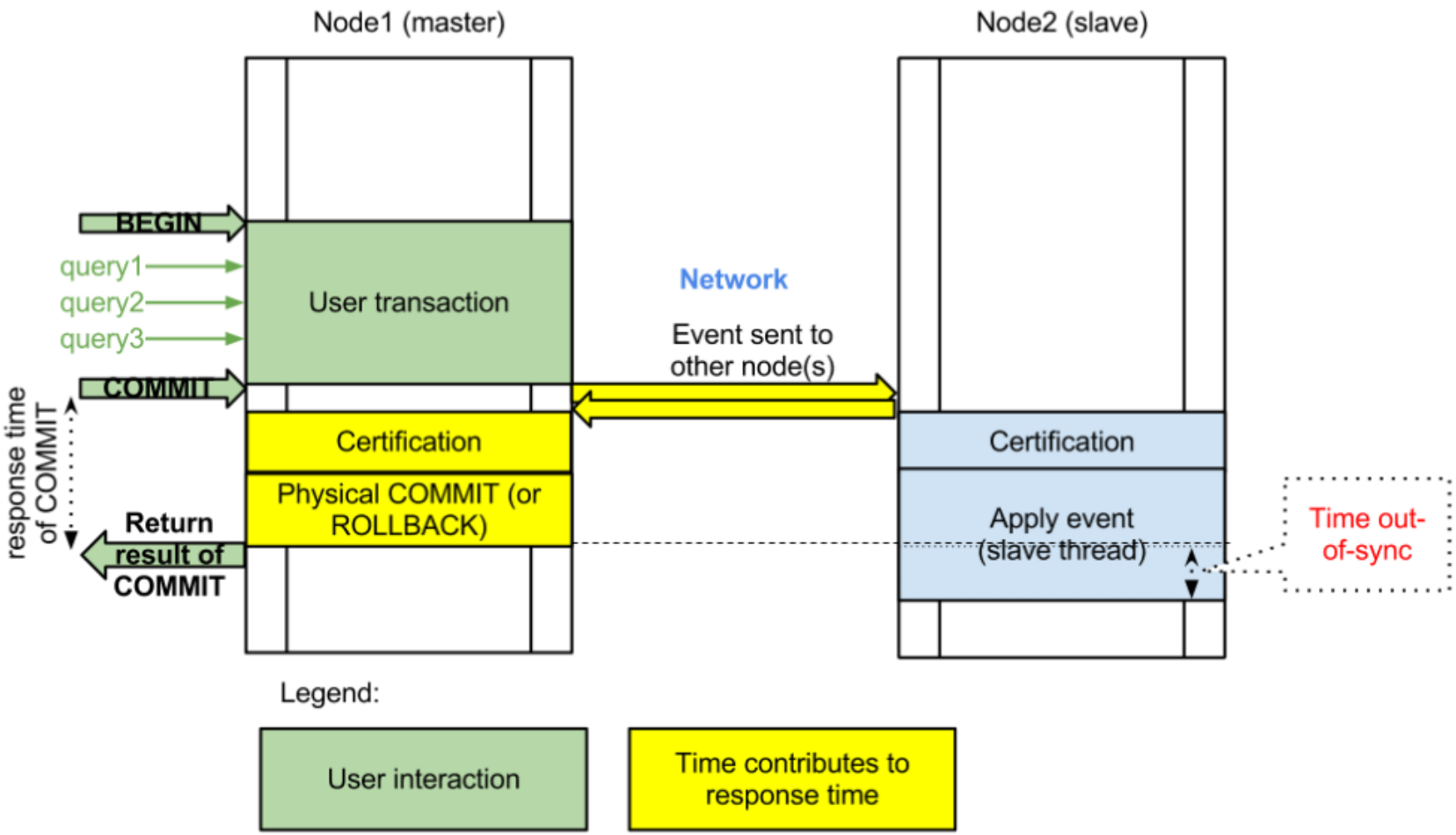
Reaction to one node loss



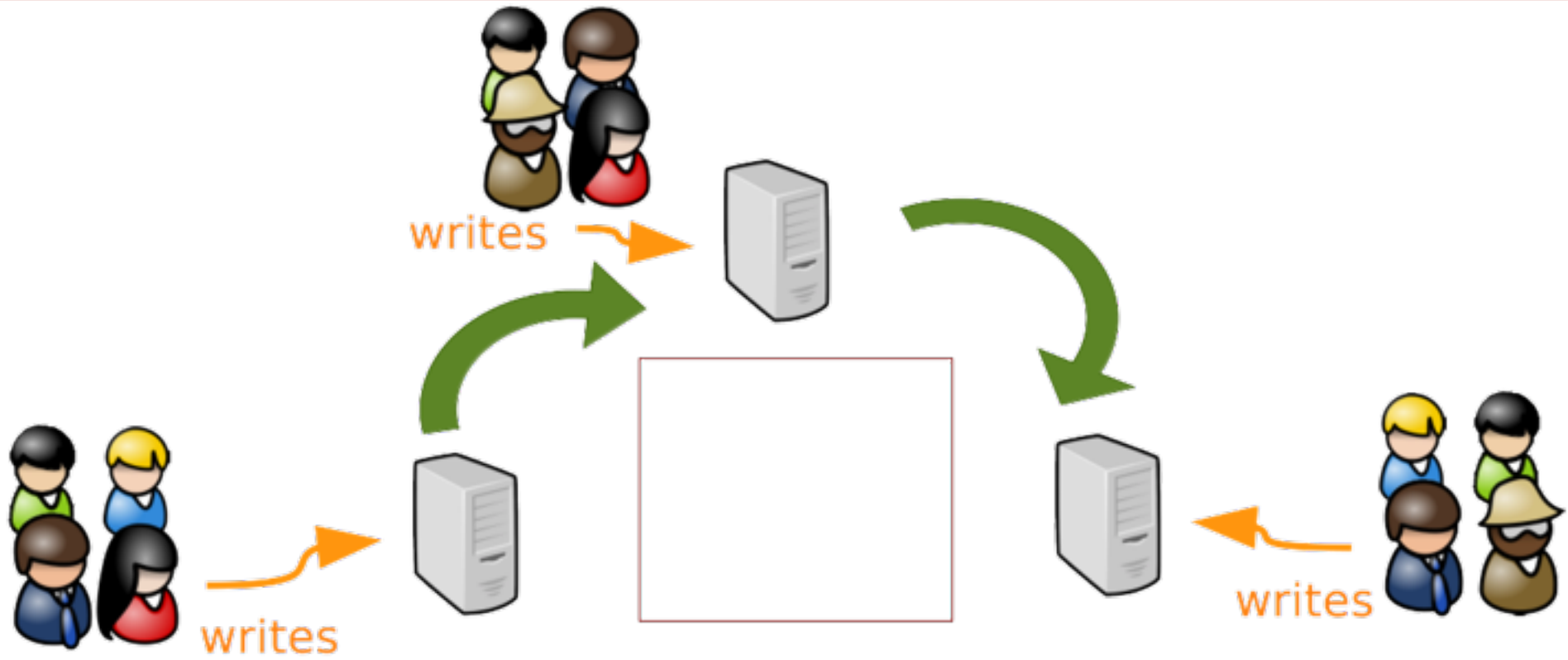
Majority loss



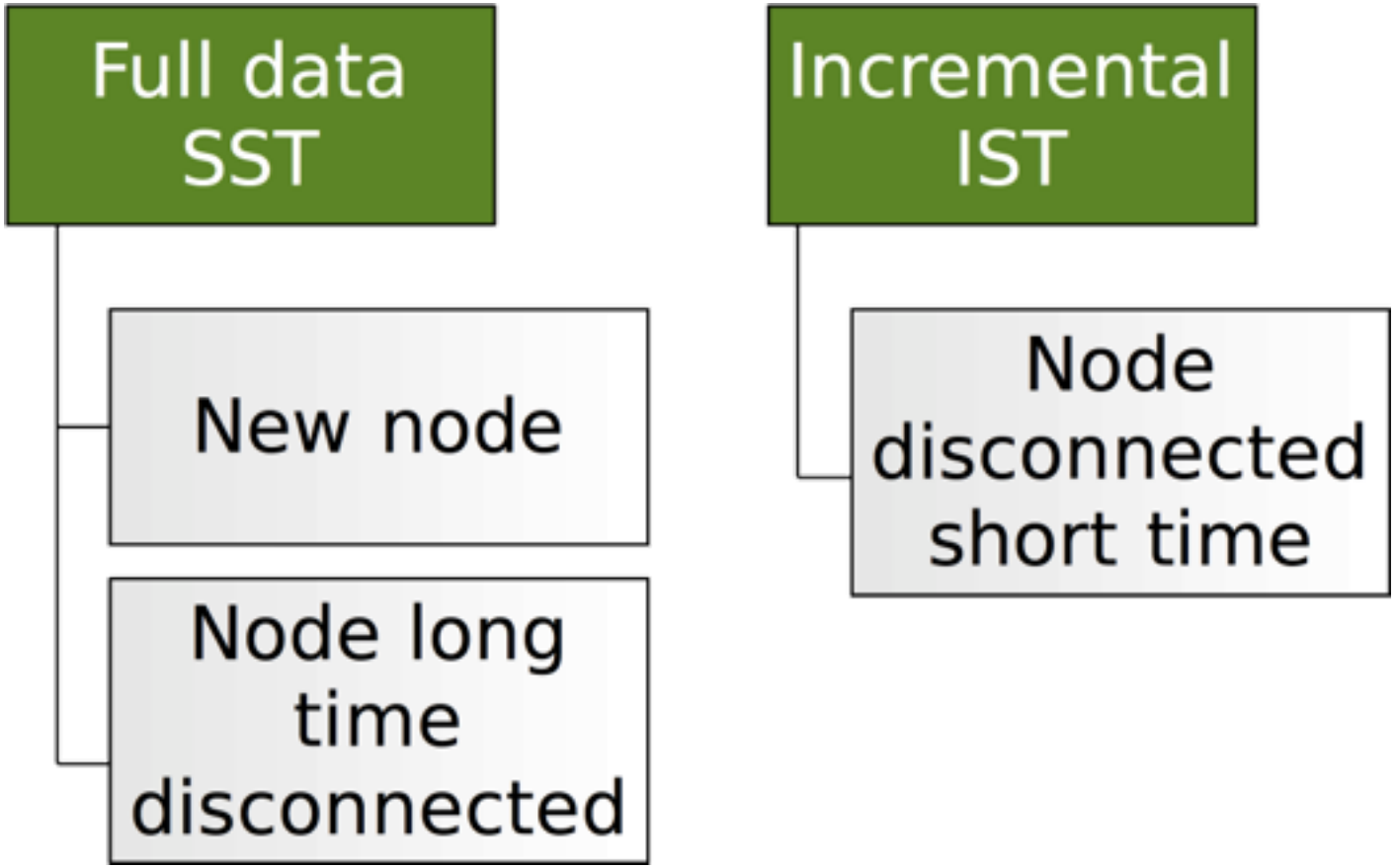
Data consistency



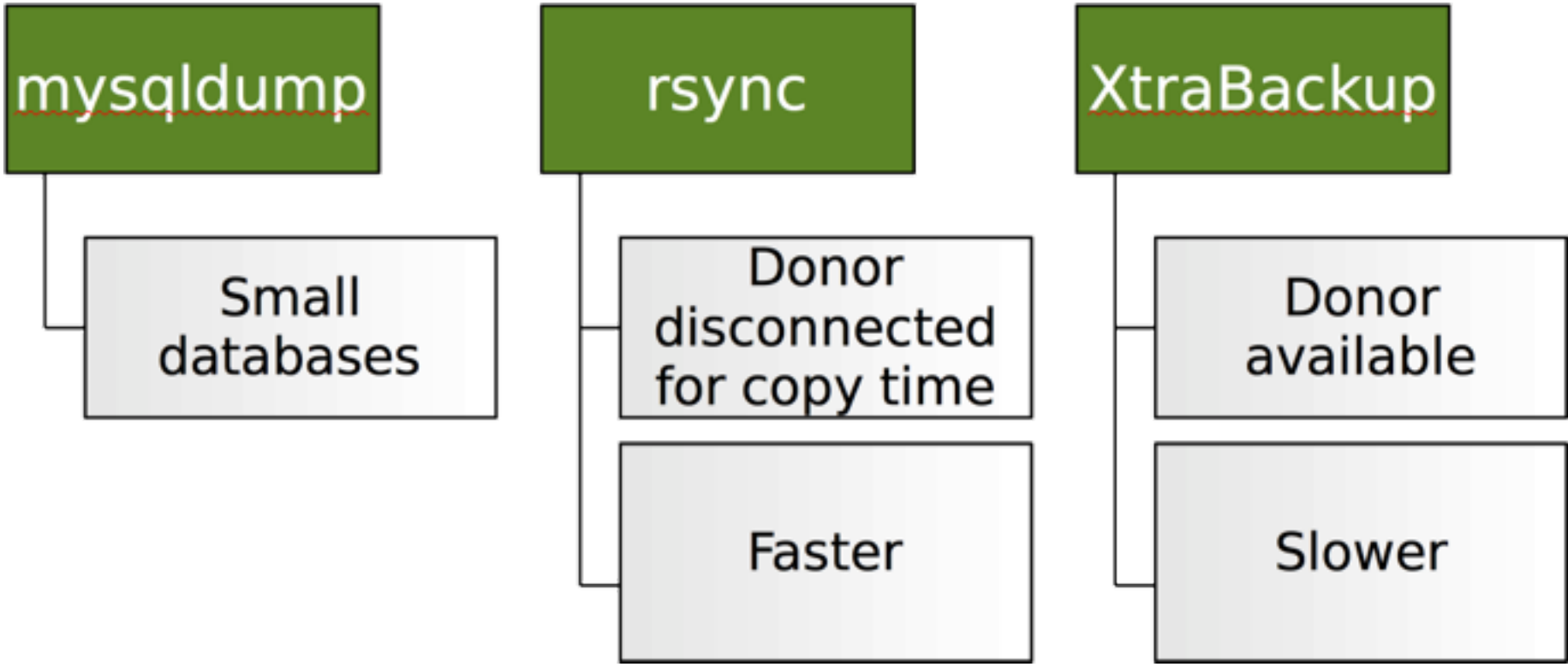
Multi Master Support



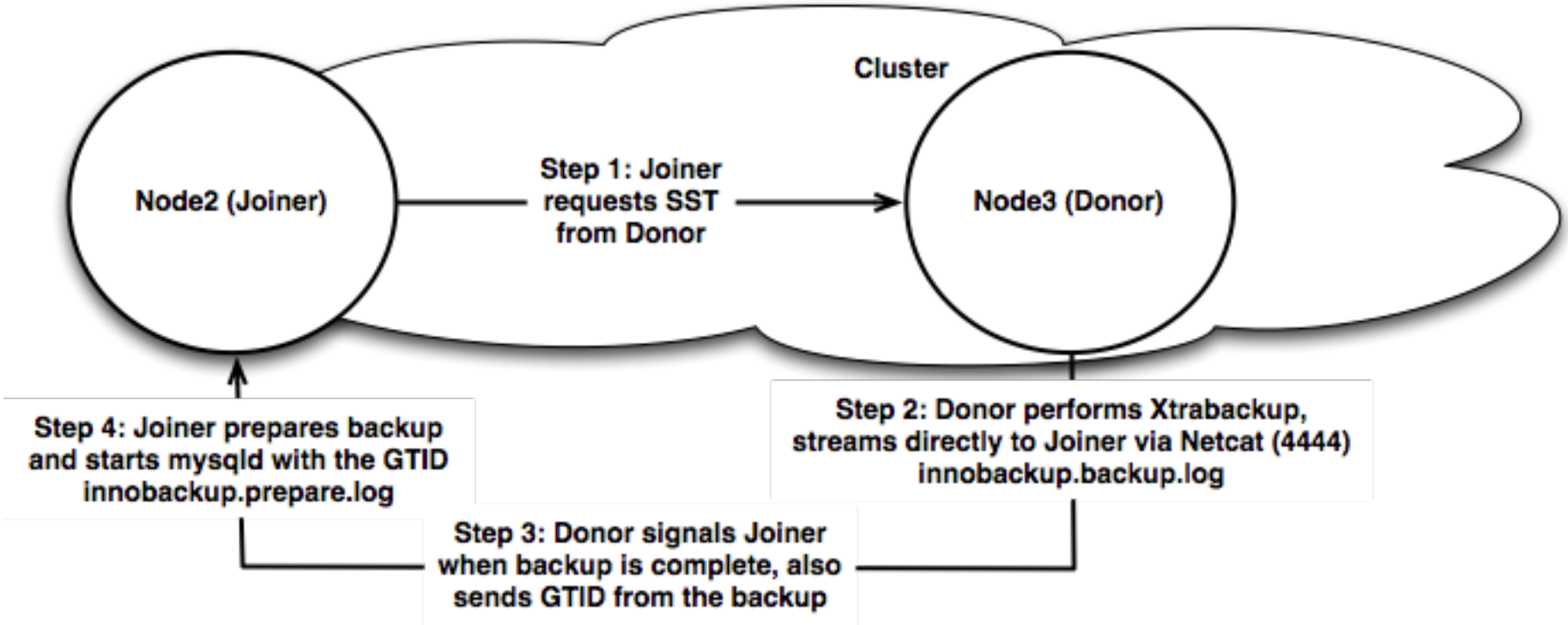
How will it auto recover from failure?



Statesnapshot



Autojoin of new node



The goal of the system

Proxy solutions

- Percona XtraDB cluster will provide you a high available dataset, but the connections to the database are not managed.
- You can use the following options:
  - VIP failover using pacemaker or PRM
  - HAproxy with clustercheck
  - MaxScale
  - Third party solutions like ScaleArc, ELB

So OpenStack how, what where.

Multi node writes are not always optimal

On heavy loaded openstack environments

Openstack components have dedicated tables which potentially will limit the multi node performance issues.

Limit large transactions (large cleanups can create performance issues.)

Alter table will lock your tables!

OpenStack items to keep in mind.

- innodb_flush_log_at_trx_commit=0
  - an option you can activate because you are using a clustered solution.
- Keep in mind that flow control can kick in.
  - Replication feedback - any node of the cluster has a list of transactions to do, it will lock the cluster in case the cluster is slowing down.
- Auto increment handling is on.
- Wan segmentation, if you have several cluster nodes in multiple DC's —> limiting the roundtrip time.
- Parallel apply but serialised components!!!

Don’t let it bite you in the face!

Operating MySQL/PXC

- Monitoring using:
  - Percona toolkit
  - Percona monitoring plugins
  - Severalnines Clustercontrol
  - Orchestrator
- Backup
  - Xtrabackup
    - Recovery time objective
    - Restore time objective
  - mydumper

Conclusion

- Use the tool optimal for your requirements.
- Percona XtraDB cluster might be a good fit for high availability, but test it thoroughly.