

Galera Replication

—

Release 1.0

Seppo Jaakola, Codership
Alexey Yurchenko, Codership

Contents

1. Galera Cluster
2. Advanced Features
3. Release 1.0
4. Benchmarking
5. Installation & Management
6. Summary

Galera Cluster



clients

MySQL

wsrep

MySQL

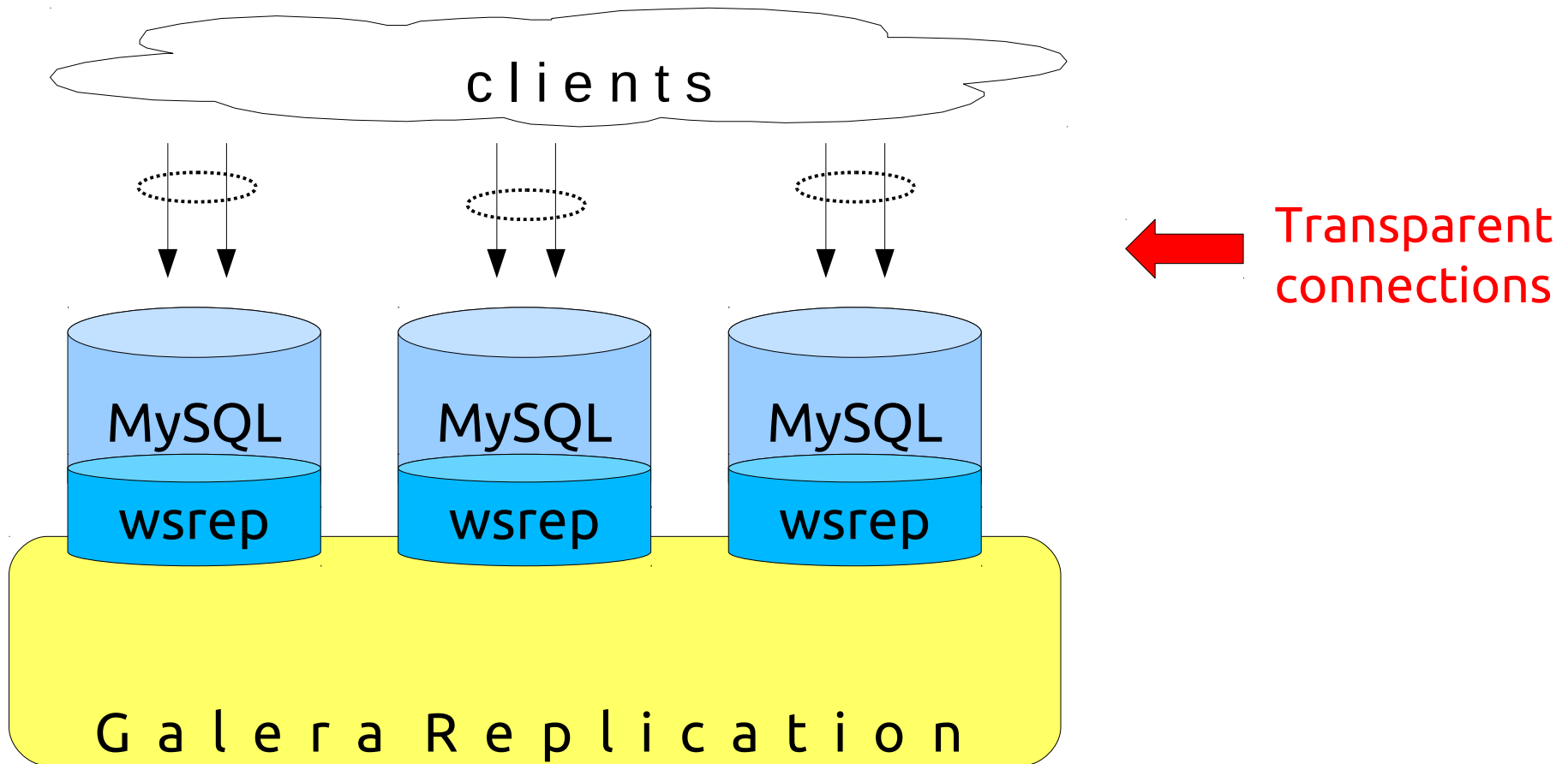
wsrep

MySQL

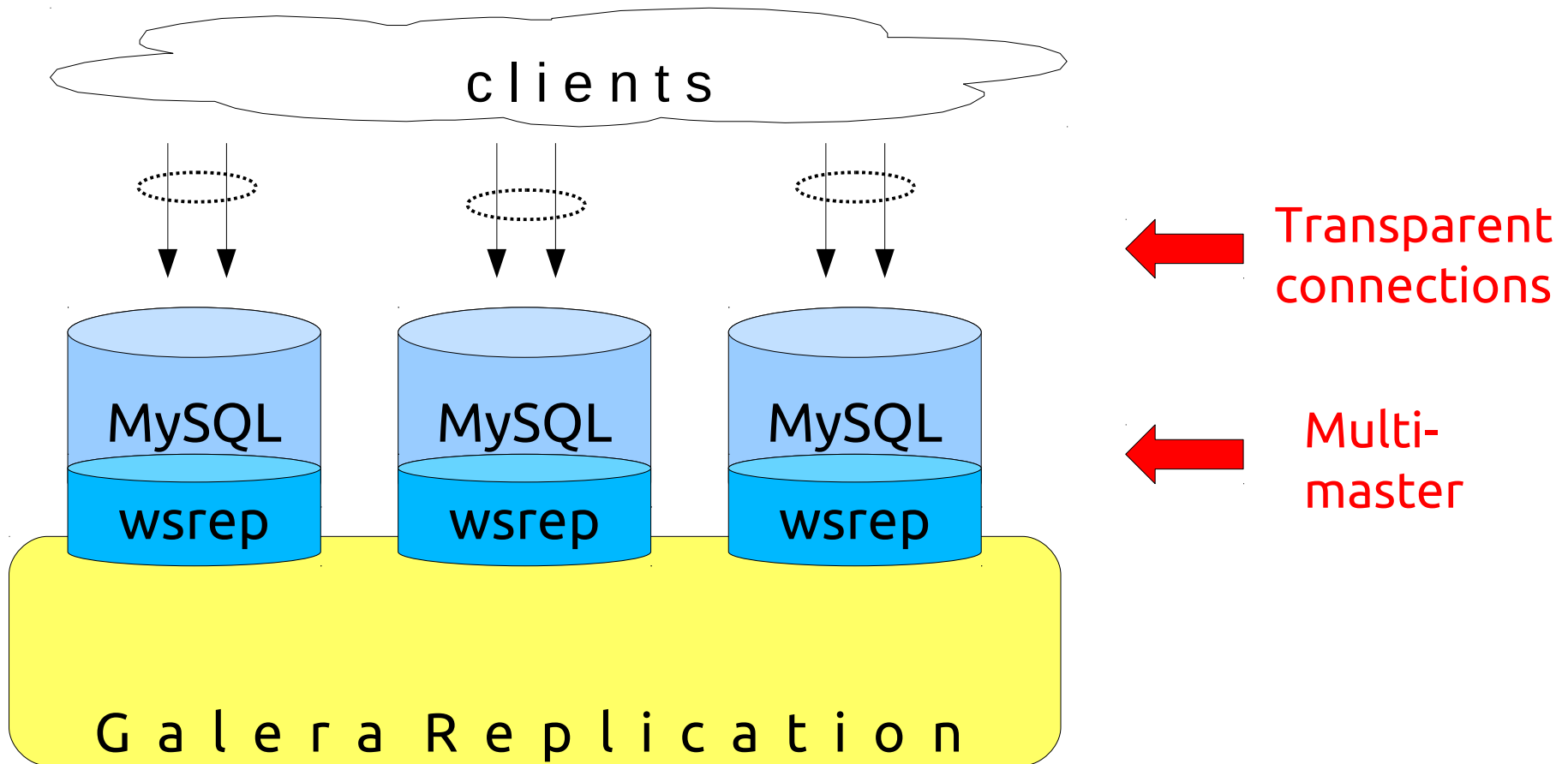
wsrep

Galera Replication

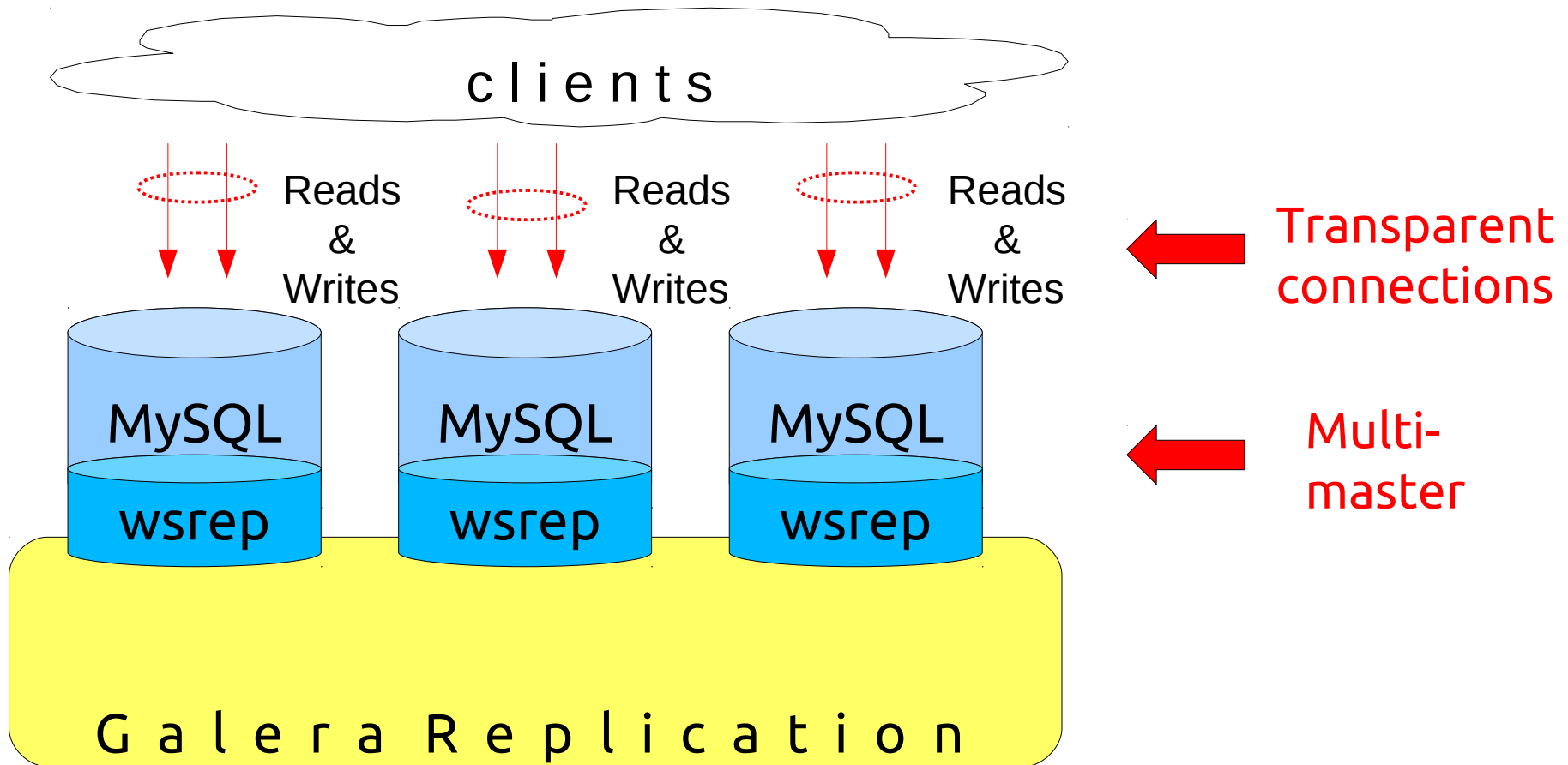
Galera Cluster



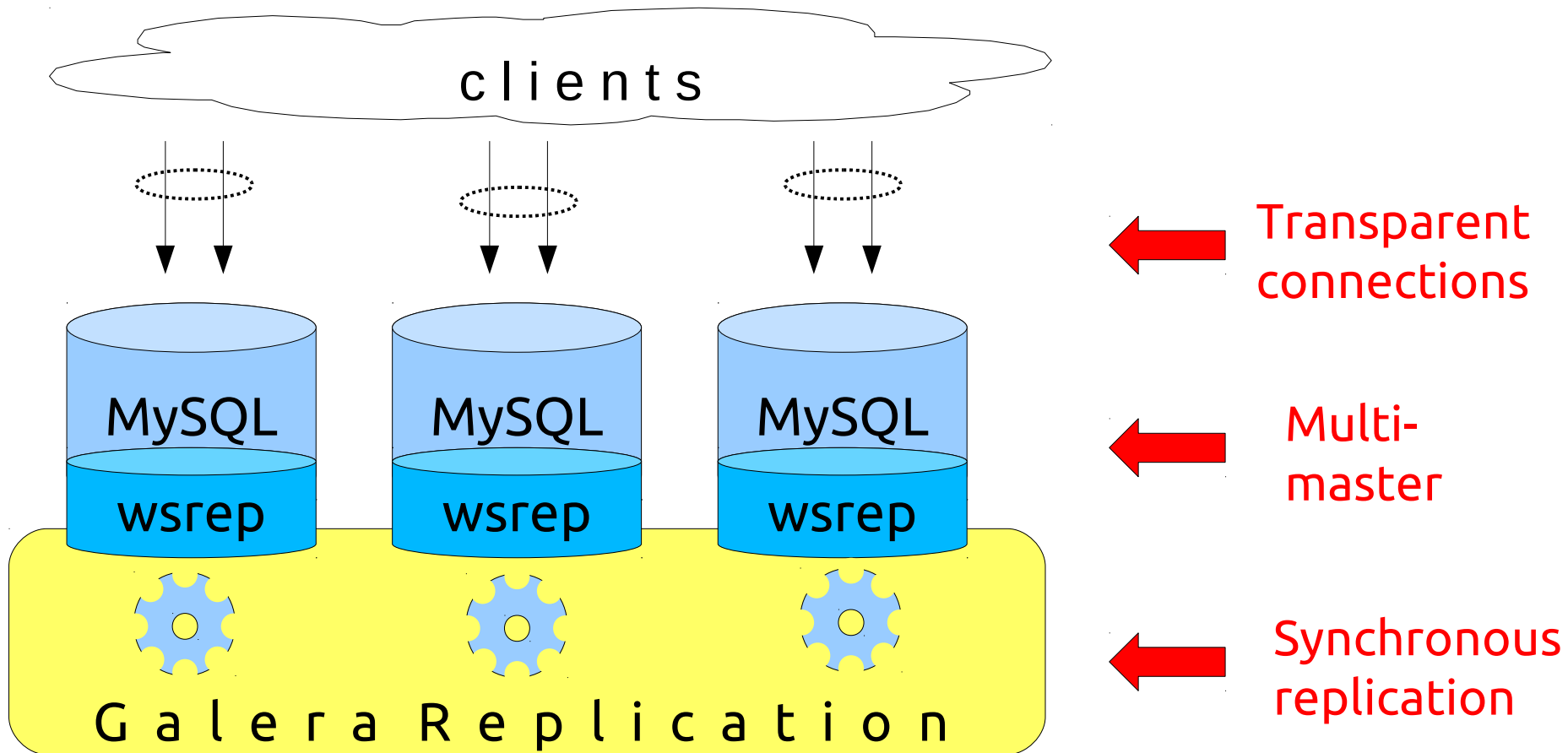
Galera Cluster



Galera Cluster



Galera Cluster



Galera Replication

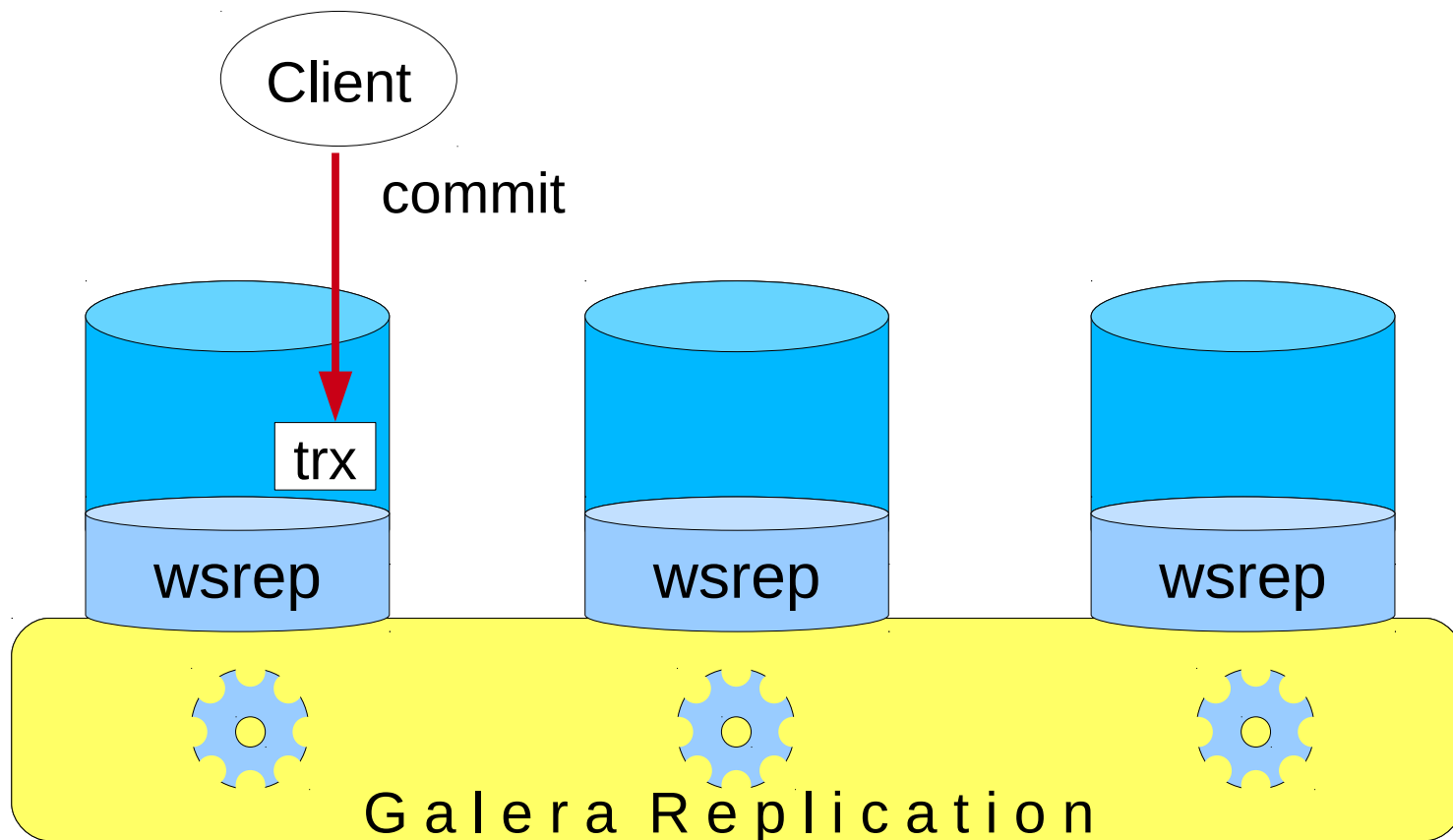
- Synchronous multi-master replication
 - High Availability
- No middle-ware, direct DBMS connections
 - Transparency
- Row events, row level locking
- Parallel Applying
 - Write scalability

Galera Replication

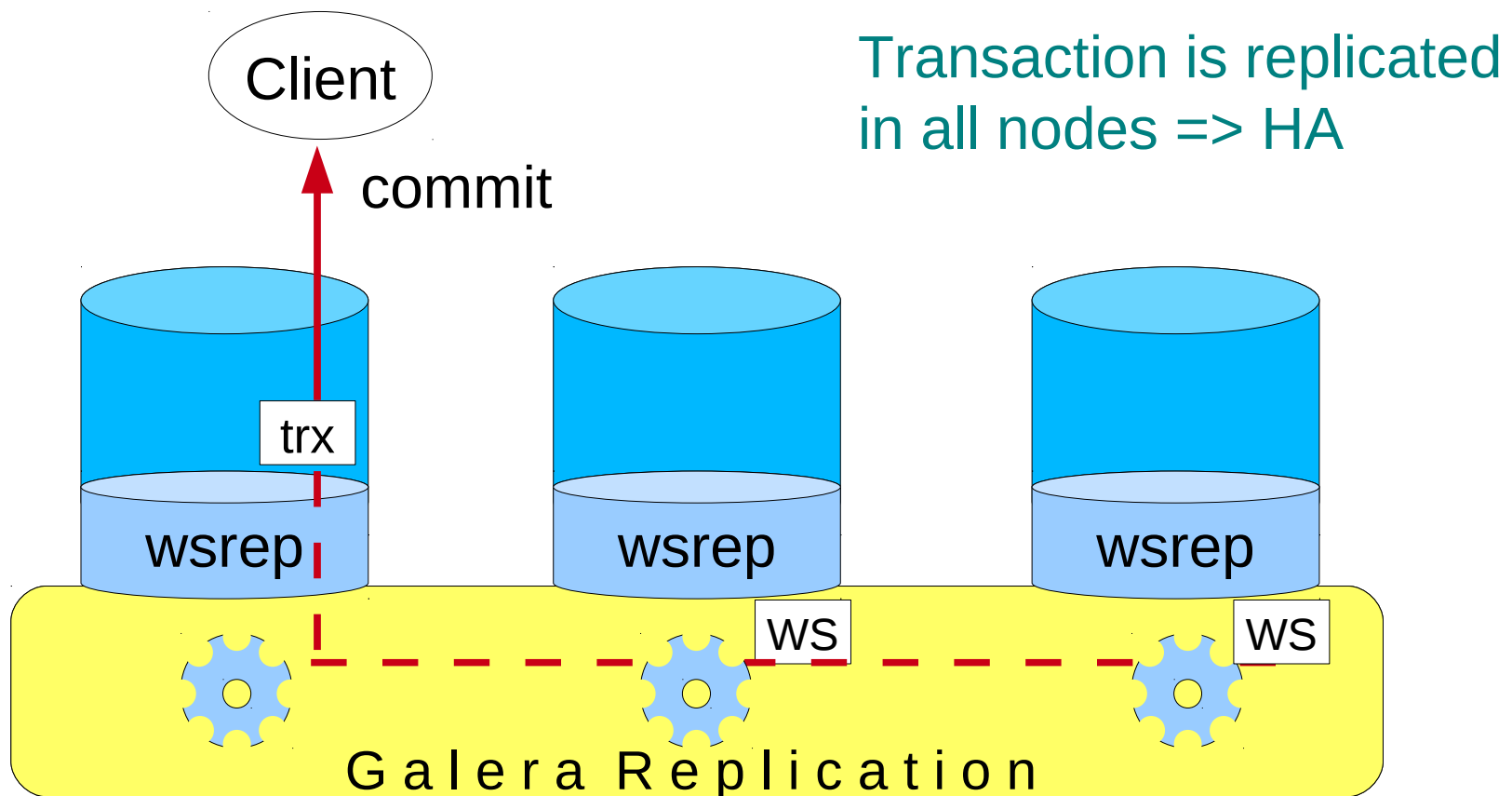
- Global Transaction ID
- Failure Detection
- Automatic node join/provisioning
 - Simple cluster management
- Certification based replication method
- Group Communication System

Advanced Features

Synchronous Replication



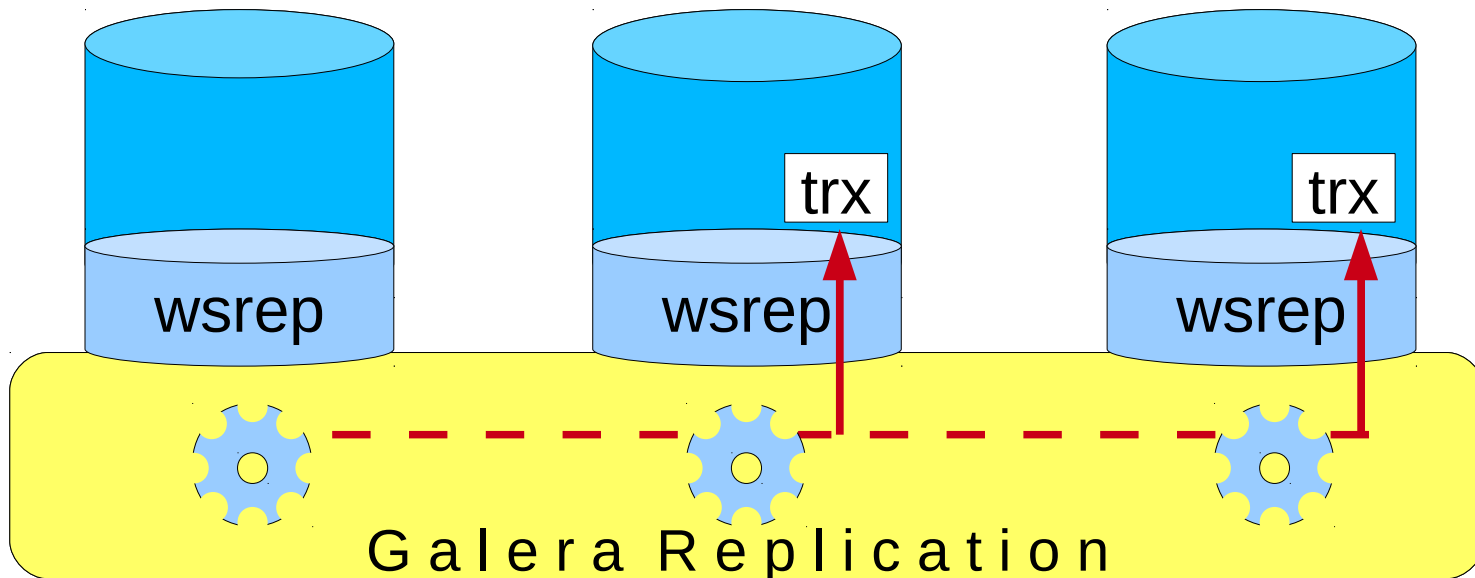
Synchronous Replication



Synchronous Replication

Client

Transaction is applied at later time
=> virtually synchronous



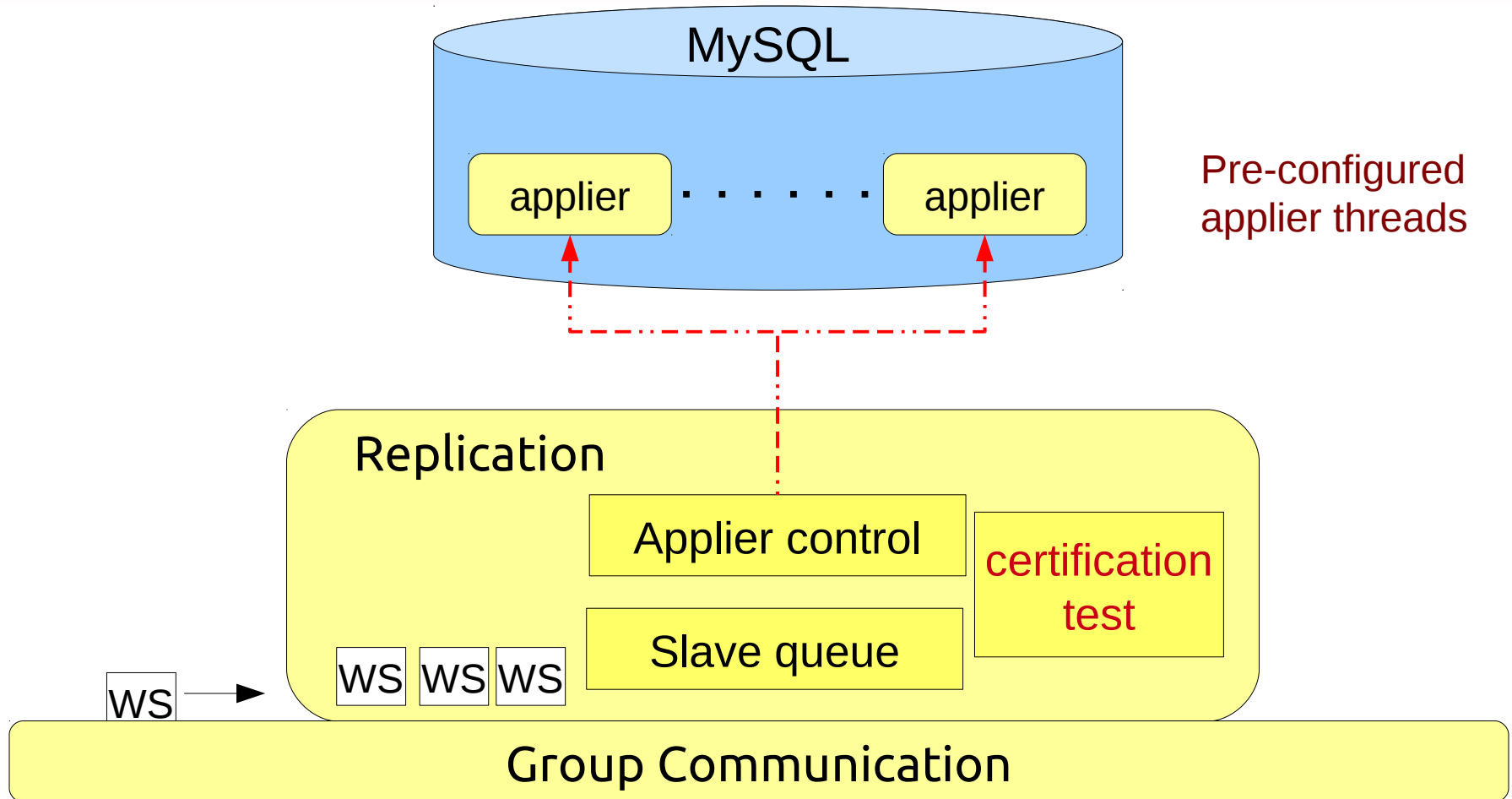
Certification Based Replication

- Transactions process independently in each cluster node
- Transaction write sets will be replicated at commit time
- Cluster wide conflicts resolved by certification test

Parallel Applying

- Galera assigns non-conflicting WS tasks to parallel appliers
- Applier threads launched at MySQL startup
- Any number of appliers can be launched
- Optimal applier count depends on work load

Parallel Applying



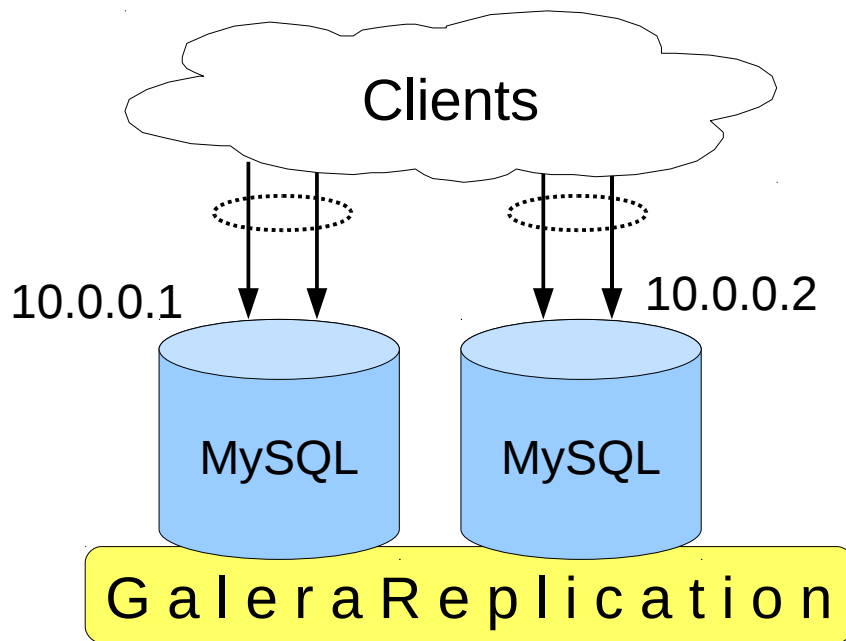
Global Transaction ID

- Galera assigns Global Transaction ID for all replicated transactions
- Transactions can be uniquely referenced in any node
- Helps in provisioning new nodes

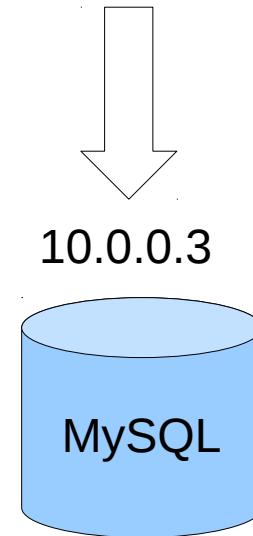
Node Provisioning

- Automatic node joining
- Cluster chooses 'donor' for the 'joiner'
- State Snapshot Transfer
- Scriptable interface, currently implemented:
 - `mysqldump`
 - `rsync`
 - `Xtrabackup` by Percona

Joining New Nodes

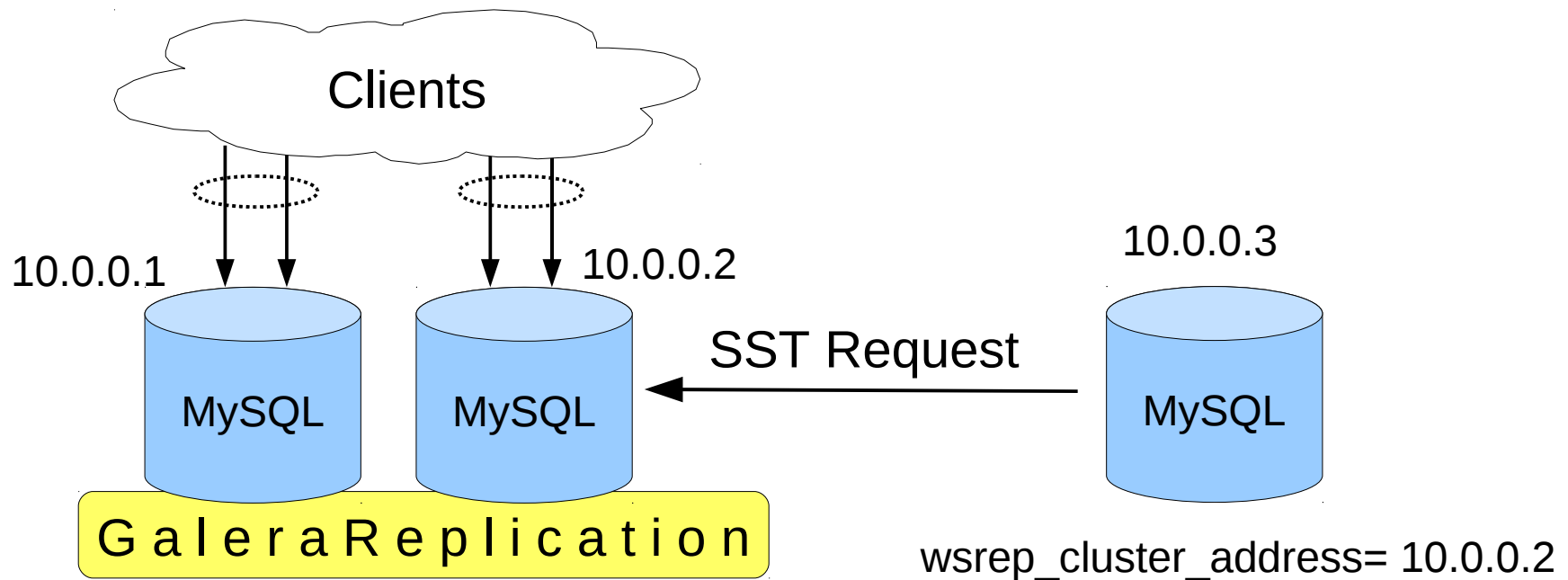


Install Galera



A c t i v e c l u s t e r

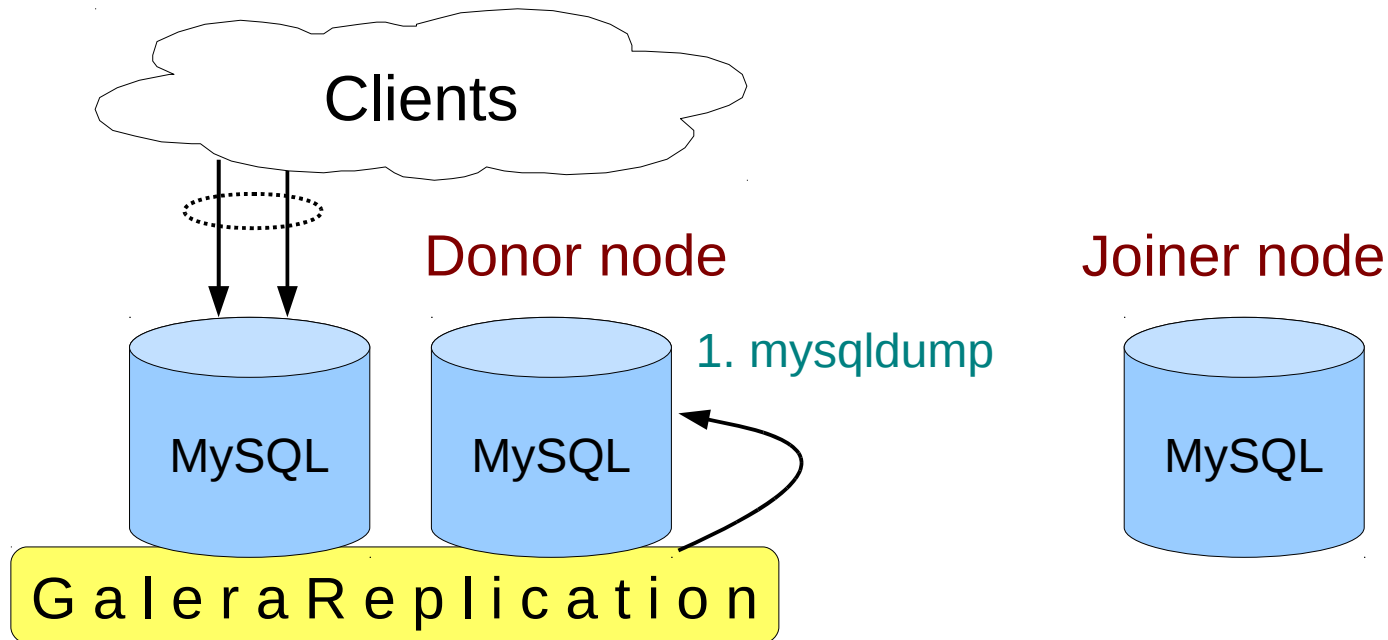
Joining New Nodes



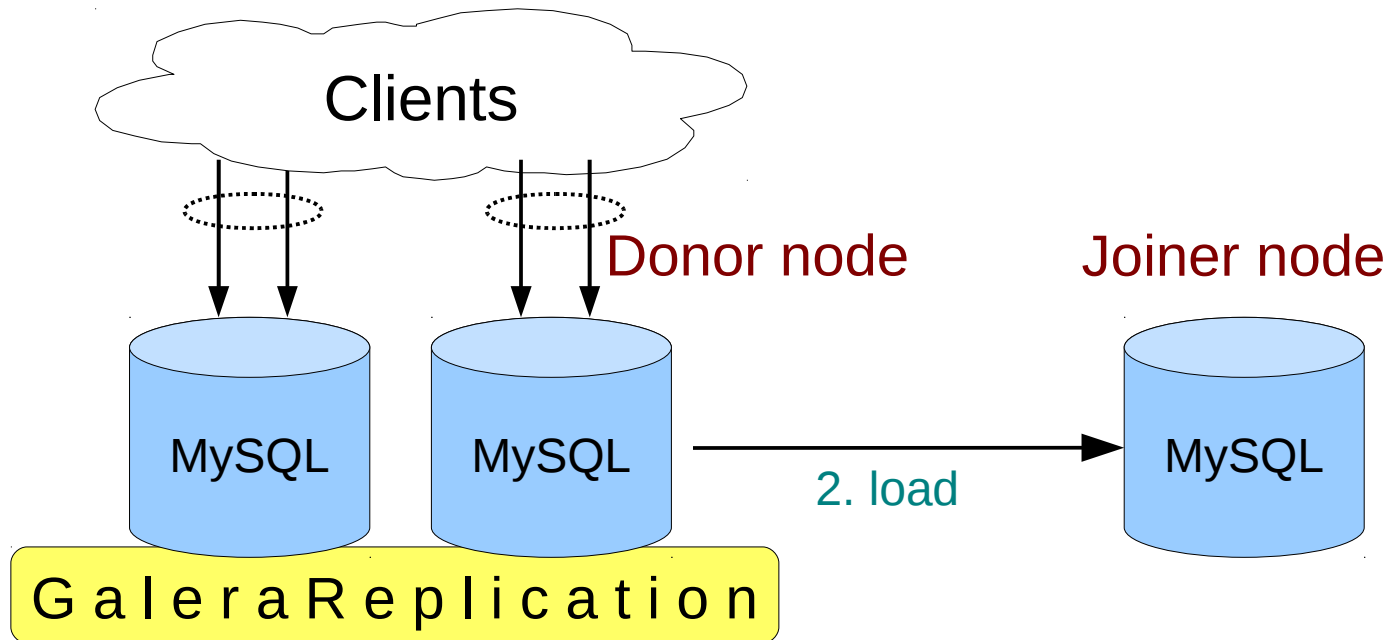
Active cluster

Joining node

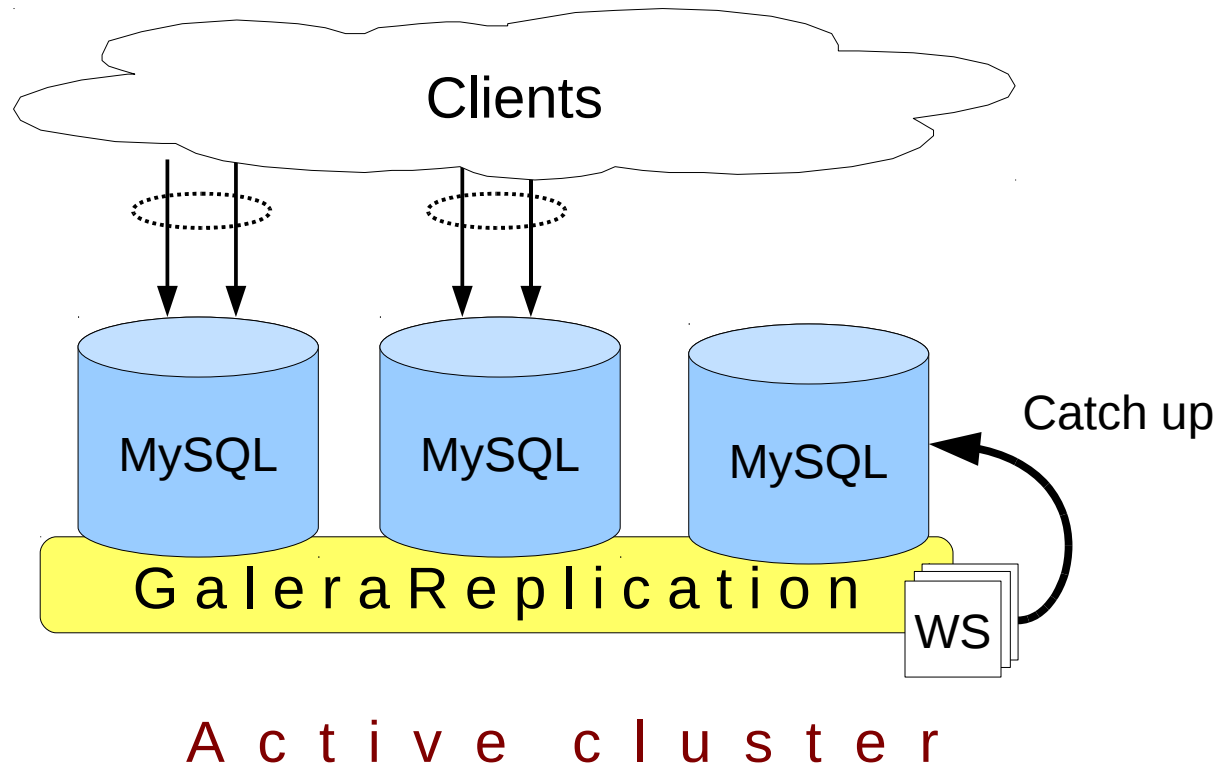
Joining New Nodes



Joining New Nodes

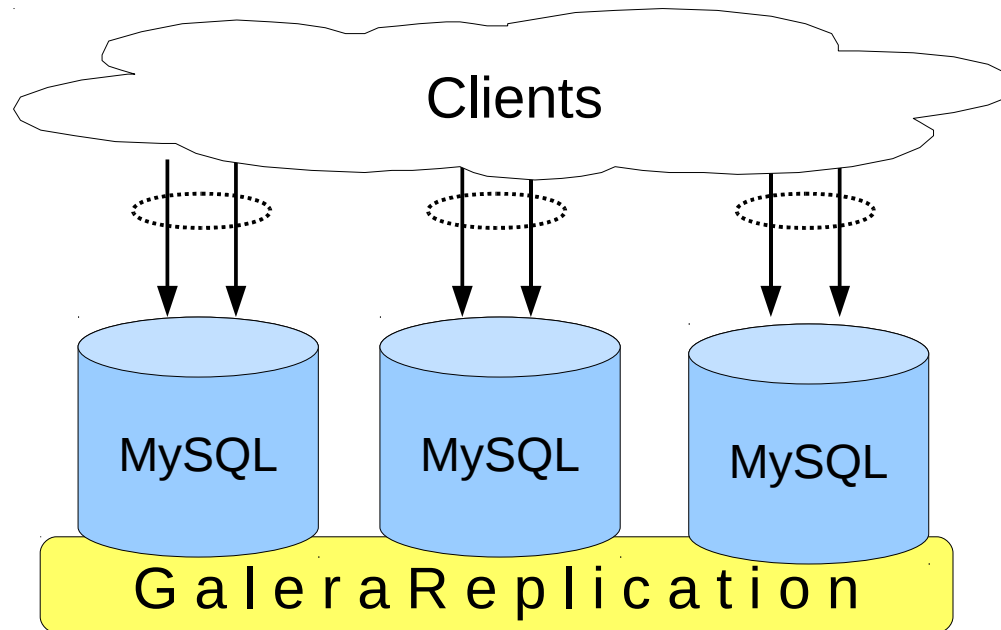


Joining New Nodes



A c t i v e c l u s t e r

Joining New Nodes



A c t i v e c l u s t e r

Release 1.0

New Features

- MySQL 5.5 Support
- On-disk write set caching
- Fully synchronous reads
- SSL encrypted replication
- Garbd – lightweight arbitrator daemon

MySQL Support

- MySQL 5.5
 - Development head, all new features go here
- MySQL 5.1
 - Still actively maintained
 - Bug fixes
- Percona Server
 - Developed by Percona
 - <https://code.launchpad.net/~percona-dev/percona-server/percona-server-galera-5.5.15>

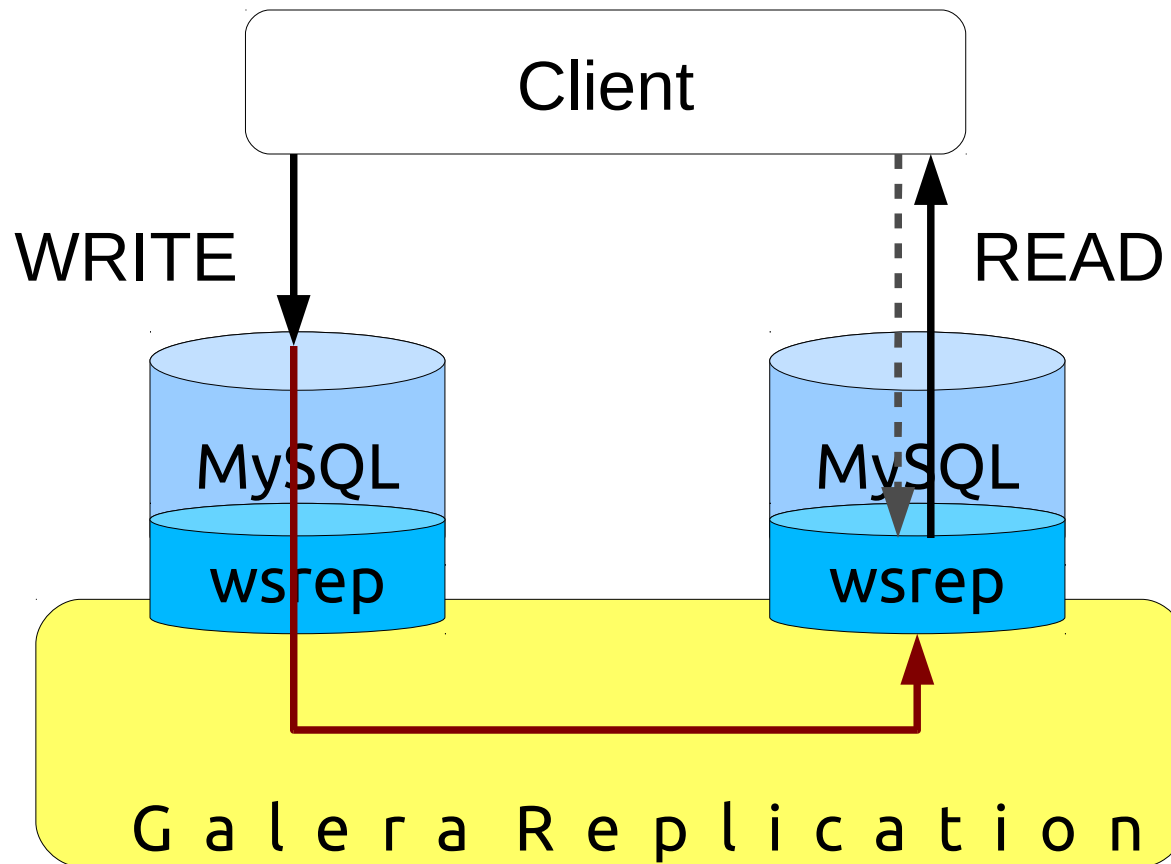
On-disk Write Set Cache

- **All write sets are allocated in mmapped files.**
- **Full use of OS buffering, negligible overhead.**
- **Used in:**
 - Node Provisioning
 - State Transfer
 - Rolling Schema update

On-disk Write Set Cache

- **Preallocated fixed-size ring buffer:**
 - KEEPS SOME HISTORY FOR RECOVERY
AFTER SHORT CONNECTIVITY BREAKS OR
GRACEFUL RESTARTS
- **Additional dynamic pages if ring
buffer is too small:**
 - DISK SPACE IS THE LIMIT

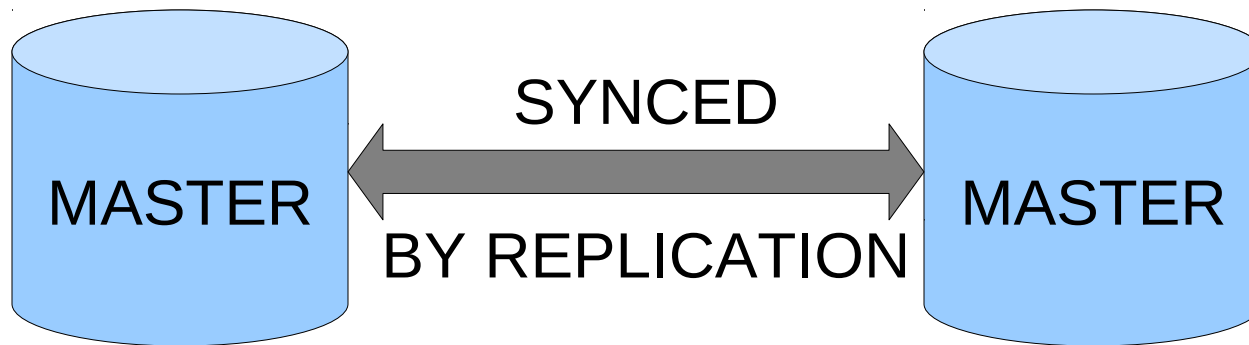
Synchronous Reads



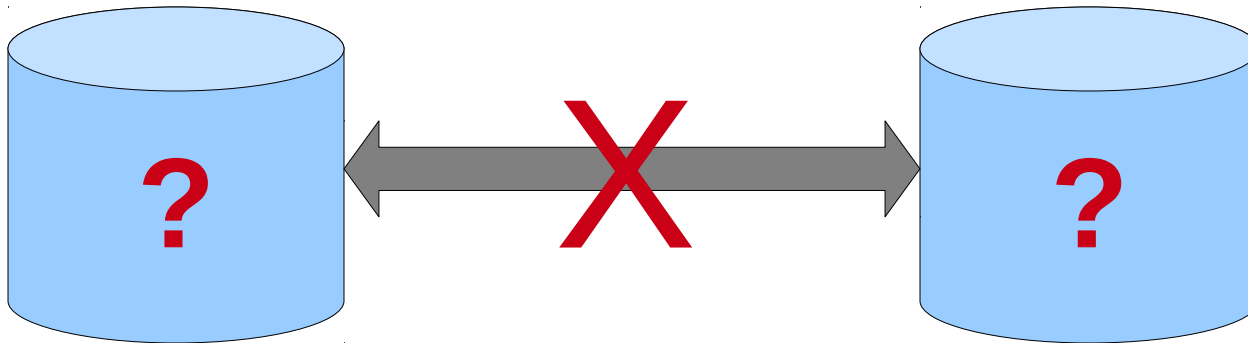
Encrypted Replication

- TLS/SSL v3 support added in 0.8.2 release
- Whole cluster replicated either in plain text or through encryption
- VPN tunneling is optimal for data center linking

Split-Brain



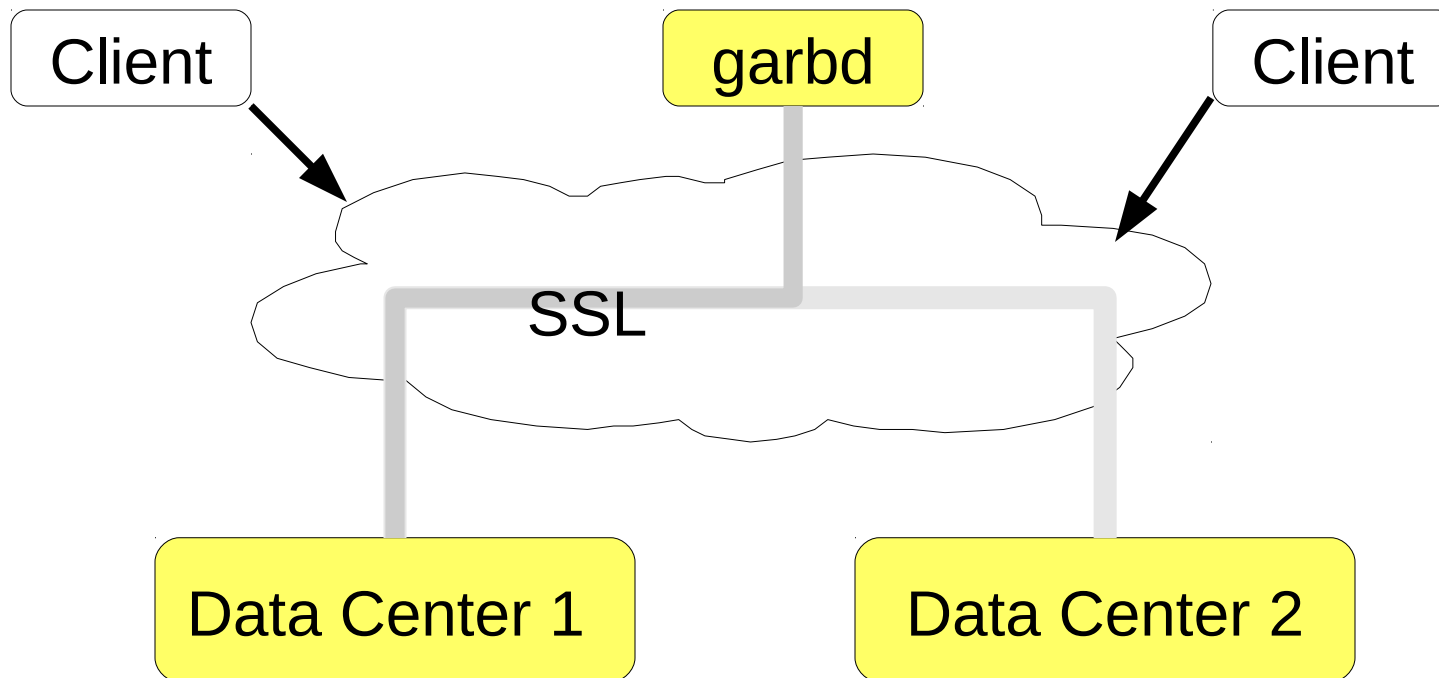
Split-Brain



Galera Arbitrator (garbd)

- Resolves split-brain
- Stateless
- Can be strategically placed to improve quorum resolution odds.

WAN Replication

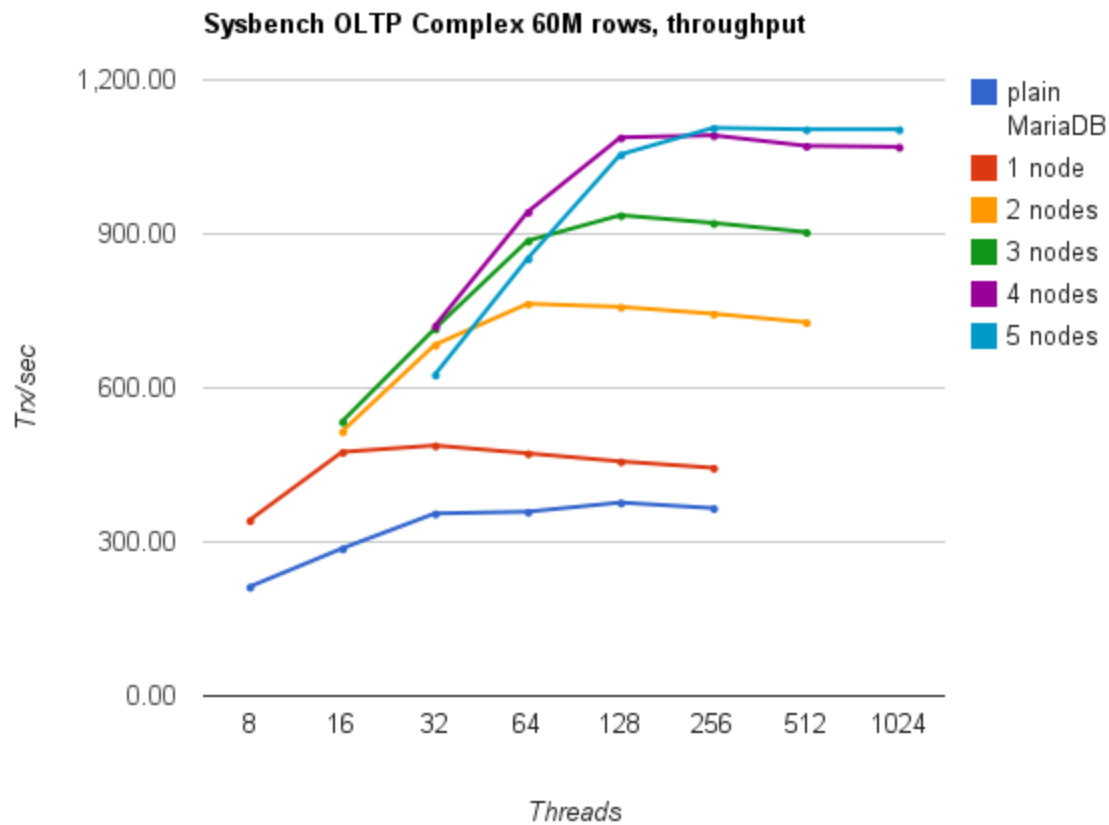


Benchmarking

Benchmarking

- Tested with several benchmarks
 - Sysbench, dbt2, DOTS, osdb, jmeter, sqlgen...
- Tested with 'physical hardware' and with Amazon EC2 instances
- In general, shows good scalability even with write intensive work loads

Scale Out: Throughput



Installation & Management

Installing MySQL/Galera

- Download from www.codership.com
- Distributions choices:
 - 1.Pre-built RPM or Debian package
 - 2.demo tar distribution
 - 3.Source build

Demo Distribution

- Pre-built 32/64 bit linux binaries
- Installs in one directory path
- Contains a sample database
- Good for testing/evaluation

wsrep Variables

```
mysql> show variables like 'wsrep%';
```

Variable_name	Value
wsrep_auto_increment_control	OFF
wsrep_certify_nonPK	OFF
wsrep_cluster_address	gcomm://?gmmcast.listen_addr=tcp://127.0.0.1:4568
wsrep_cluster_name	my_wsrep_cluster
wsrep_convert_LOCK_to_trx	OFF
wsrep_data_home_dir	/codership/data/galera-nod1/
wsrep_debug_option	
wsrep_debug	OFF
wsrep_drupal_282555_workaround	OFF
wsrep_local_cache_size	20971520
wsrep_max_ws_rows	65636
wsrep_max_ws_size	0
wsrep_node_incoming_address	10.1.198.1:3307
wsrep_node_name	nod1
wsrep_notify_cmd	
wsrep_on	ON
wsrep_provider	/codership/nod1/mysql-5.1.52/galera/lib/libmmgalera++.so
wsrep_provider_options	
wsrep_retry_autocommit	OFF
wsrep_slave_threads	1
wsrep_sst_auth	test:testpass
wsrep_sst_donor	
wsrep_sst_method	mysqldump
wsrep_sst_receive_address	AUTO
wsrep_start_position	00000000-0000-0000-0000-000000000000:-1
wsrep_ws_persistency	OFF

wsrep Status

wsrep_local_state_uuid	a398eaf8-2aba-11e0-0800-432d0098b829
wsrep_last_committed	2989366
wsrep_replicated	122
wsrep_replicated_bytes	161514094
wsrep_received	0
wsrep_received_bytes	0
wsrep_local_commits	110
wsrep_local_cert_failures	0
wsrep_local_bf_aborts	0
wsrep_local_replays	0
wsrep_local_send_queue	0
wsrep_local_send_queue_avg	0.007752
wsrep_local_recv_queue	0
wsrep_local_recv_queue_avg	0.000000
wsrep_flow_control_paused	0.000000
wsrep_flow_control_sent	0
wsrep_flow_control_recv	0
wsrep_cert_deps_distance	1.750000
wsrep_apply_oooe	0.000000
wsrep_apply_ool	0.000000
wsrep_apply_window	1.000000
wsrep_local_state	4
wsrep_local_state_comment	Synced (6)
wsrep_cluster_conf_id	4
wsrep_cluster_size	2
wsrep_cluster_state_uuid	a398eaf8-2aba-11e0-0800-432d0098b829
wsrep_cluster_status	Primary
wsrep_local_index	1
wsrep_ready	ON

Severalnines ClusterControl

- Severalnines has developed Galera support for the ClusterControl tool for:
 - Configurator
 - Monitoring
 - Management
- Install in a separate management server

ClusterControl - Configurator

1. Fill your architecture specs in:

<http://www.severalnines.com/galera-configurator>

→ You'll get a tarball in mail,

2. Untar and run deploy script

3. Sit back and relax

Galera Cluster
'myGaleraCluster'(17/4)

MySQL Servers

- 10.30.30.31
- 10.30.30.32
- 10.30.30.33

Galera Cluster - myGaleraCluster

Current Cluster Load

queries/sec	inserts/sec	selects/sec	updates/sec	deletes/sec	DB Connections active/connected/max
11437	178	11259	0	0	18 / 34 / 46

Galera Servers

Hostname	MySQL Status	Galera Stats	Server Stats	Host Stats	Version	Last HB	Details
10.30.30.31	●	State: Synced (6) wsrep_local_send_queue_avg: 0.991813 wsrep_local_recv_queue_avg: 0.000000 wsrep_flow_control_paused: 0.000000	Queries/s: 60 DB Connections: 14 Uptime: 8h 24m	ping: ok cpu util: 7.5% uptime: 12days 17h 27m		2011-10-25 10:31:08	view graphs
10.30.30.32	●	State: Synced (6) wsrep_local_send_queue_avg: 0.000000 wsrep_local_recv_queue_avg: 0.713612 wsrep_flow_control_paused: 0.000000	Queries/s: 5537 DB Connections: 6 Uptime: 8h 24m	ping: ok cpu util: 99.4% uptime: 12days 17h 27m	5.5.15-wsrep_21.1	2011-10-25 10:31:08	view graphs
10.30.30.33	●	State: Synced (6) wsrep_local_send_queue_avg: 0.901160 wsrep_local_recv_queue_avg: 0.618791 wsrep_flow_control_paused: 0.009267	Queries/s: 7794 DB Connections: 14 Uptime: 8h 23m	ping: ok cpu util: 99.5% uptime: 4days 19h 5m	5.5.15-wsrep_21.1	2011-10-25 10:31:08	view graphs

Cluster - myGaleraCluster

SQL Layer

- [Processlist \(global\)](#)
- [Query analyzer \(global\)](#)

Management

- [Alarms \(4\)](#)
- [Backups](#)
- [Config mgmt](#)
- [Host mgmt](#)
- [Job Messages \(17\)](#)
- [Node mgmt](#)
- [Performance mgmt](#)
- [Process mgmt](#)
- [Scale](#)
- [Schema mgmt](#)
- [Software Packages](#)
- [Upgrade](#)

Troubleshooting

- [Logs](#)

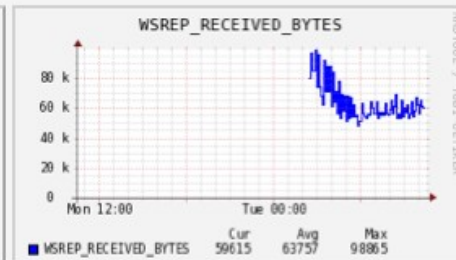
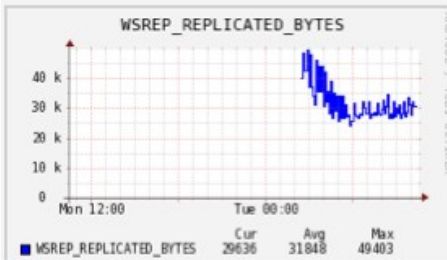
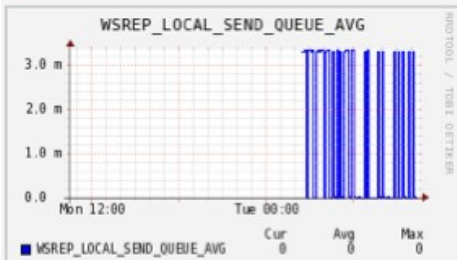
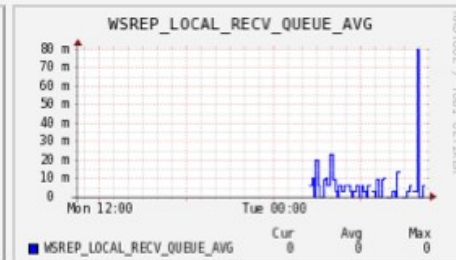
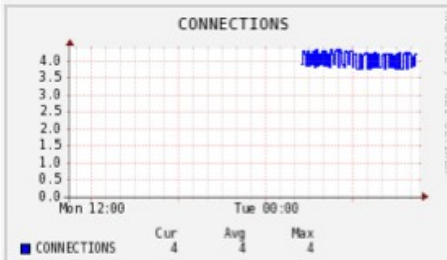
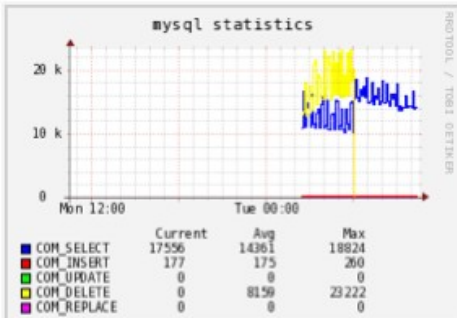
Administration

- [Email notifications](#)
- [Cluster settings](#)

Dashboard - Monitored Clusters

Galera Cluster - 'myGaleraCluster'

cluster id	cluster name	status	heartbeat	actions
1	myGaleraCluster	● STARTED ●	2011-10-25 10:30:05	View Cluster



Host Alarms

	hostname	component / resource	severity	alarm	alarm count	description	recommendation	updated	dismiss
●	10.30.30.32	RAM	CRITICAL	Excessive RAM Usage	5	RAM Utilization for 10.30.30.32 is 90 percent	Upgrade Node with more RAM	2011-10-25 10:30:12	<input type="button" value="dismiss"/>
●	10.30.30.30	RAM	CRITICAL	Excessive RAM Usage	5	RAM Utilization for 10.30.30.30 is 97 percent	Upgrade Node with more RAM	2011-10-25 10:30:14	<input type="button" value="dismiss"/>
●	10.30.30.31	RAM	CRITICAL	Excessive RAM Usage	5	RAM Utilization for 10.30.30.31 is 90 percent	Upgrade Node with more RAM	2011-10-25 10:30:14	<input type="button" value="dismiss"/>
●	10.30.30.30	RAM	WARNING	SWAP space used	2007	10.30.30.30 is swapping	Upgrade Node with more RAM, check MYSQL configuration	2011-10-25 10:30:14	<input type="button" value="dismiss"/>

- Galera Cluster 'myGaleraCluster'(17/4)
- MySQL Servers
 - 10.30.30.31
 - 10.30.30.32
 - 10.30.30.33

Performance Management - Probe 1 - myGaleraCluster

Period: 30 min
 Probe: 1

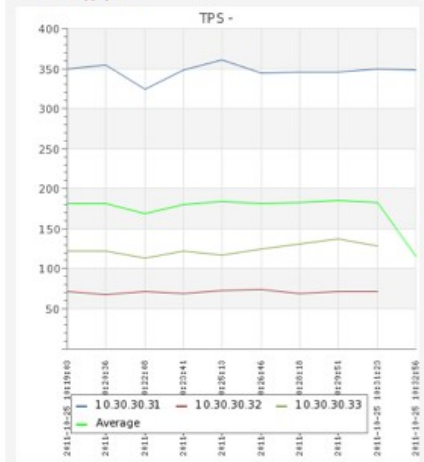
Query: insert into t1(b) values('hello galera')
 Active: Yes

Hostname	DB Connections	Connection Pool	Execution Count	Rows returned	Avg Response Time (us) (stdev)	95th Percentile (us)	Throughput (tps) (stdev)	Last updated
10.30.30.31	8	0	1349	0	22984 (17364)	56400	348 (21)	2011-10-25 10:32:58
10.30.30.32	8	0	278	0	111344 (94679)	265290	71 (20)	2011-10-25 10:31:57
10.30.30.33	8	0	499	0	62064 (71979)	229718	128 (24)	2011-10-25 10:32:07
TOTAL	8	0	2126	0	65464	190489	547	-

change

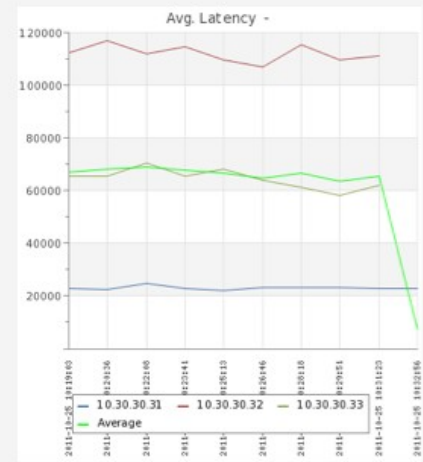
Throughput

Total (tps) : 542
 Average (tps) : 181
 Last value (tps) : 182



Latency

Average (response time) : 66526
 Last value (response time) : 65464



- Cluster - myGaleraCluster
- SQL Layer
 - Processlist (global)
 - Query analyzer (global)
- Management
 - Alarms (4)
 - Backups
 - Config mgmt
 - Host mgmt
 - Job Messages (17)
 - Node mgmt
 - Performance mgmt
 - Process mgmt
 - Scale
 - Schema mgmt
 - Software Packages
 - Upgrade
- Troubleshooting
 - Logs
- Administration
 - Email notifications
 - Cluster settings

ClusterControl

- Use in management server:
 - Monitor to visualize the cluster state
 - Manage your cluster
- Monitor/manage cluster as a whole

Summary

Galera Support

- Codership offers support services for MySQL/Galera users
- We are building support partner network with local support providers
 - Local timezone
 - Local Language & Culture
 - Codership on level 3 support to back up
- FromDual starting now

Summary

- Galera is Replication Redefined
 - No slave lag, no lost transactions
 - Native InnoDB look & feel
 - Ultimate performance
 - WAN/LAN/Cloud
- Severalnines ClusterControl
 - ease of use experience
- Support available
 - FromDual

codership

- R&D consulting services
- Galera Support

- Web-site: <http://www.codership.com>
- Downloads: <https://launchpad.net/codership-mysql>
- Mailing list: codership-team@googlegroups.com