MySQL Cluster mit Galera

DOAG Konferenz 2013 Nürnberg

Oli Sennhauser

Senior MySQL Consultant, FromDual GmbH

oli.sennhauser@fromdual.com

Über FromDual GmbH

- FromDual bietet neutral und unabhängig:
  - Beratung für MySQL und Galera
  - Support für MySQL und Galera
  - Remote-DBA Dienstleistungen
  - MySQL Schulungen
- Partner der Open Database Alliance (ODBA.org)
- Oracle Silver Partner (OPN)

www.fromdual.com

Inhalt

Galera Cluster

- Bestehende Probleme
- Was ist Galera Cluster für MySQL
- Eigenschaften
- Konfiguration
- Betrieb
- Demo
- Weiteres

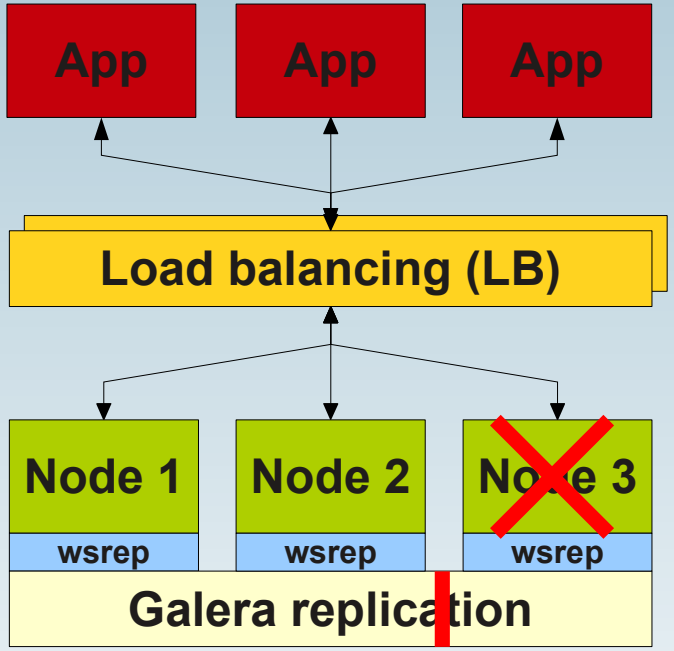
Bestehende Probleme

Probleme mit bestehenden Lösungen:

- Datenkonsistenz – M/S Replikation
- Asynchron, Slave-Lag (Hinterherhinken) – M/S Replikation
- Komplexität – a/p Failover-Cluster, NDB Cluster
- Downtime – M/S Replikation, a/p Failover-Cluster
- Nicht geeignet für komplexe Abfragen (= Joins) – NDB Cluster
- Single Point of Failure (I/O System, File-System) – a/p Failover-Cluster

- Was wäre wenn es eine Lösung gäbe, die all diese Probleme NICHT hat?

Galera Cluster



Eigenschaften von Galera

- Basiert auf der transaktionalen InnoDB SE
- Synchrone Replikation

  → Keine verlorenen Transaktionen mehr
- Echtes paralleles Replizieren auf Zeilenebene

  → Kein Slave Lag (Hinterherhinken) mehr
- Aktiv/aktiv multi-Master-Topologie

  → Lesen von und Schreiben auf beliebige Knoten möglich
- Lese-Skalierbarkeit und Erhöhung des Schreibdurchsatzes (SSD)
- Automatisches Knoten-Management
- Rolling Cluster Restart: Upgrade von Hardware, O/S, DB und Galera im Laufenden Betrieb...

Galera Konfiguration

- my.cnf (conf.d/galera.cnf, conf.d/wsrep.cnf)

```
# wsrep_provider                  = none
wsrep_provider                    = …/lib/plugin/libgalera_smm.so

# wsrep_cluster_address           = "gcomm://"
wsrep_cluster_address             = "gcomm://node2,node3"

wsrep_cluster_name                = 'Galera Cluster'
wsrep_node_name                   = 'Node A'

wsrep_sst_method                  = mysqldump
wsrep_sst_auth                    = sst:secret
```

Starten des Clusters

Demo:

- Starten des ersten Knotens
- Staren der weiteren Knoten
- Cluster Status
- Starten des Load-Balancers (GLB)
- Load-Balancer Status
- Test-Applikation
  - Langsam
  - Schnell

Demo Eigenschaften

- Storage Engine
- Paralleles Replizieren
- Lesen und Schreiben von allen Knoten
- Knoten Management
  - Full Sync (SST)
  - Incremental Sync (IST)
  - Rolling Cluster Restart: InnoDB Buffer Pool Size
- DB-Upgrade
- Galera-Upgrade

Online Schema Upgrade (OSU)

- Schema Upgrade = DDL ausgeführt gegen die DB
  - Ändern der DB Struktur
  - Nicht transaktional!
- 2 Methoden:
  - Total Order Isolation (TOI) (default)
  - Rolling Schema Upgrade (RSU)
- `wsrep_osu_method = {TOI|RSU}`

Online Schema Upgrade

- Total Order Isolation (TOI) (default)
  - DDL wird auf allen Knoten in der selben Reihenfolge ausgeführt
  - Ein Teil der DB wird während des DDLs gesperrt

    + Einfach, vorhersagbar und garantierte Datenkonsistenz

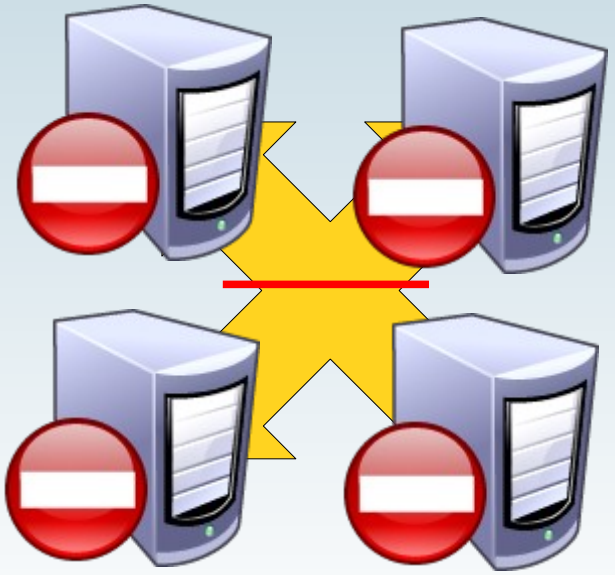
    - Sperrende Operation
  - Gut für schnelle (= kleine) DDL Operationen
- Rolling Schema Upgrade (RSU)
  - DDL wird nur auf einem Knoten aufs Mal ausgeführt
  - Knoten sind für die Dauer des DDL desynchronisiert
  - Nach DDL, werden die fehlenden Write Sets (= Transaktionen) nachgeführt, ähnlich wie im IST.
  - DDL müssen von Hand auf jedem Knoten ausgeführt werden.

    + nur ein Knoten aufs Mal wird blockiert.

    - Potentiell unsicher, kann fehlschlagen, wenn altes und neues Schema nicht kompatibel sind
  - Gut für langsame (= grosse) DDL Operationen

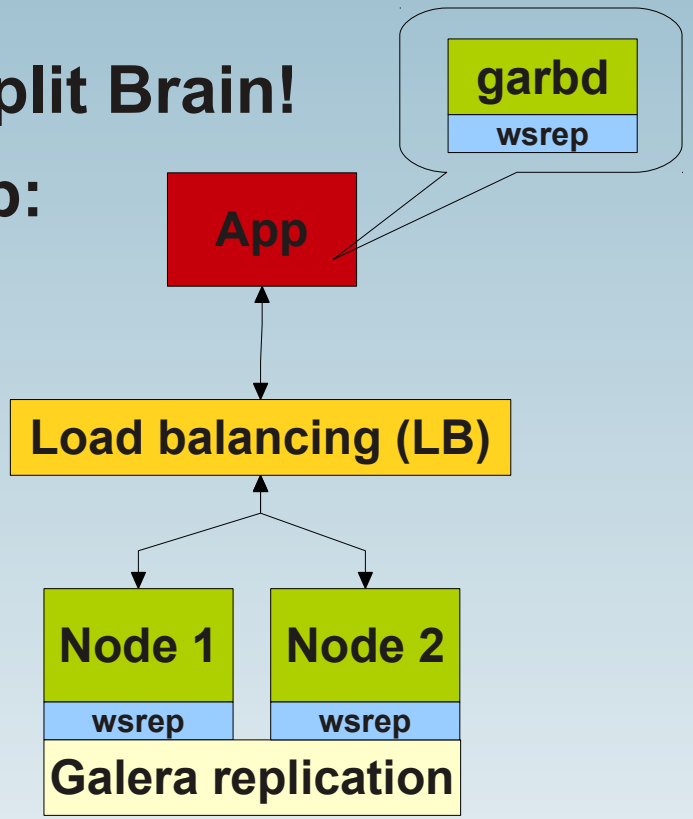
Best of all worlds!

Quorum



2 + 1 Knoten Cluster

- 2 Knoten ist böse → Split Brain!
- Minimalistisches Setup:
  2 + 1

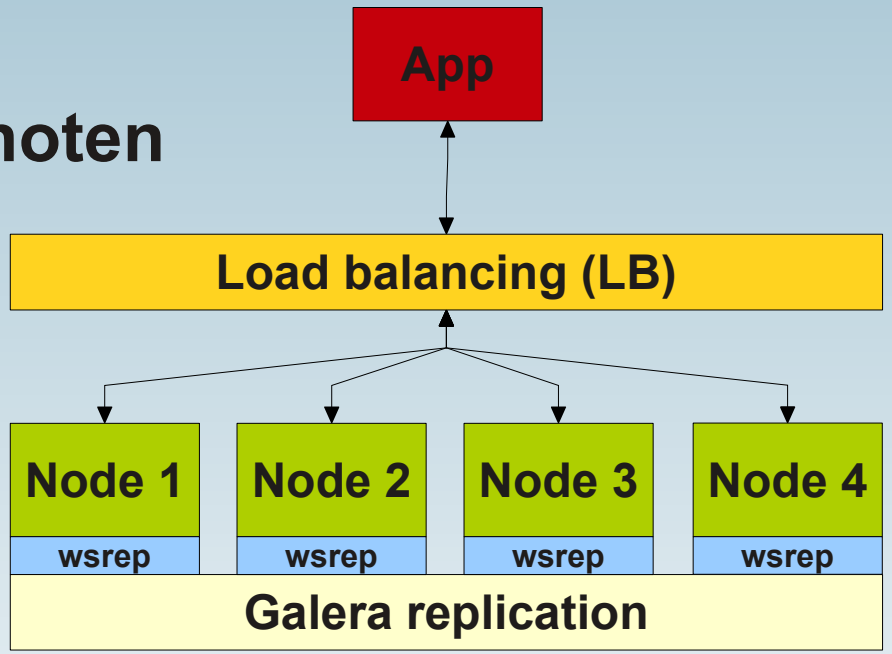


- "Unser M/S-Replikation hat jetzt nur 2 Knoten!" oder
- "Ich will nicht zu viel für Hardware ausgeben!"

  → 2 + 1 = 2 Galera Knoten + 1 Galera Arbitrator

Lese Scale-out

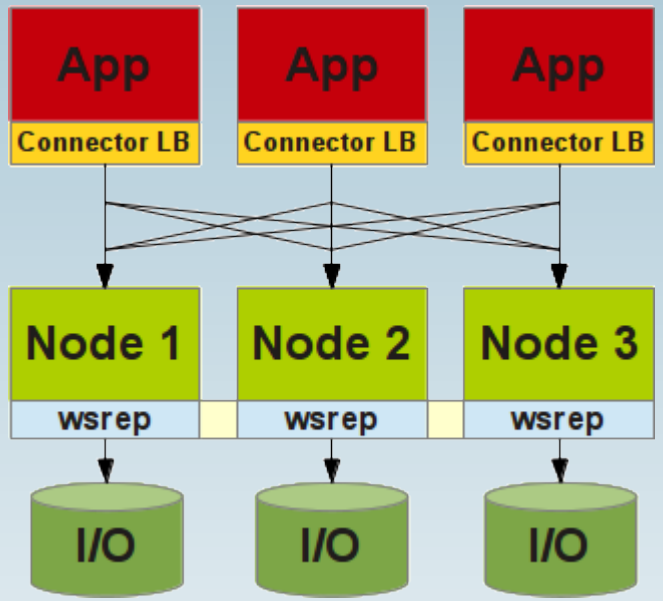
- 4 und mehr Knoten Cluster
  - Backup-Knoten
  - Dedizierter SST-Donor Knoten
  - Reporting-Knoten, etc.
- Ungerade Anzahl ist empfohlen!
  - Ansonsten → gewichtetes Quorum?
  - Gerade Anzahl: Split Brain!
- Grösster Cluster, nur so zum Spass: 17 Knoten!

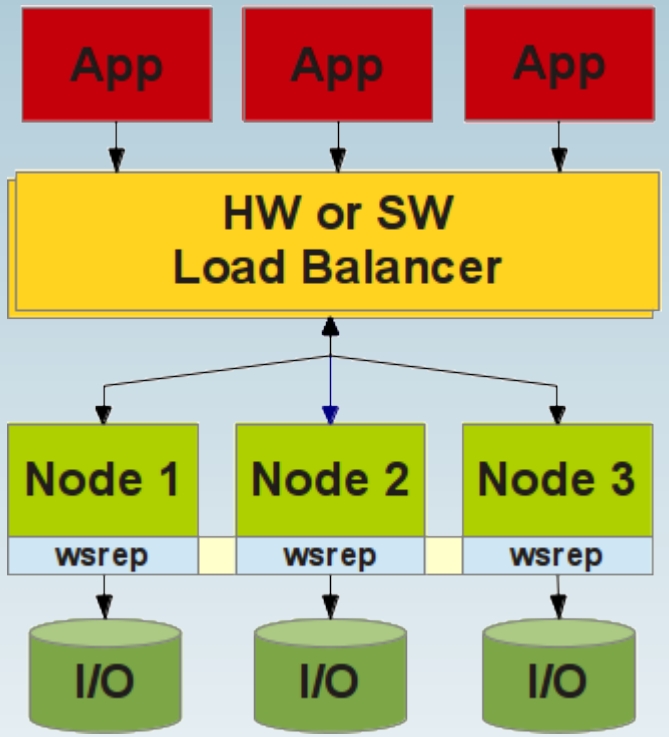


Lastverteilung

- Connectors
  - Connector/J
  - PHP: MySQLnd Replikations- und Load Balancing Plug-in
  - `libglb`
- SW Load Balancer
  - GLB, Pen, LVS/IPVS/Ldirector, Ultra Monkey, HAProxy, MySQL Proxy
- HW Load Balancer

Lage der Lastverteilung





Wir suchen noch:

- **Erfahrene/r MySQL DBA / Open-Source Enthusiast/in für MySQL Support / remote-DBA**

  und

- **Guter C++ Entwickler/in (mit Affinität zu DBs, MySQL, Replikation und Cluster)**

Q & A

Fragen ?

Diskussion?

Wir haben Zeit für ein persönliches Gespräch...

- **FromDual bietet neutral und unabhängig:**
  - **Beratung**
  - **Remote-DBA**
  - **Support für MySQL, Galera, Percona Server und MariaDB**
  - **Schulung**

www.fromdual.com/presentations