

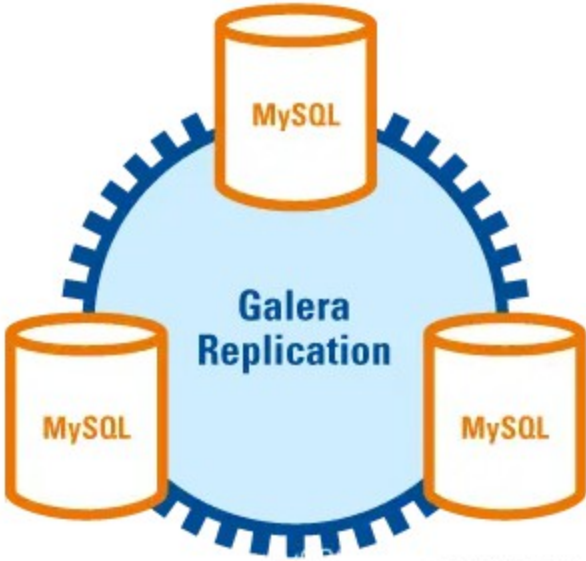


Galera Cluster一致性问题

Willem97 2021-05-12 1,848 阅读3分钟

TRAE 2.0 SOLO 出道，一键贯通从灵感火花到上线部署的全程协作

MariaDB Galera Cluster + ProxySQL 方案

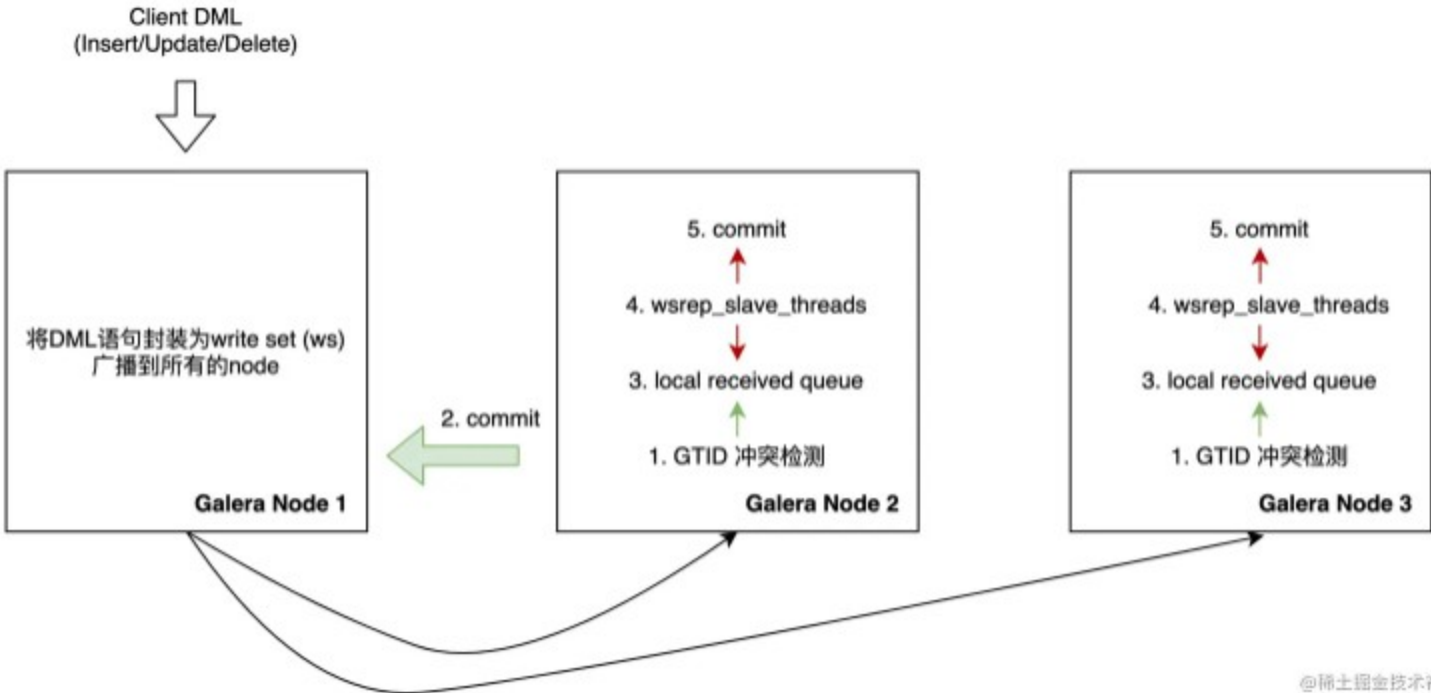


@稀土掘金技术社区

三个实例组成一个集群，为多主模式，即 **Multi-Master**，可以支持同时多点写入，多主同步复制。

Galera Cluster 提供了基于验证的复制，其是一种乐观的同步复制。每一个节点在复制事物时，都会拿这些写集 (**write-set**)与正在 **APPLY** 队列的写集作比对，如果没有冲突，这个事物就可以继续提交，或 **APPLY**，及时任务这个事务被提交，之后在数据库层面，进行事务上的提交操作。

Galera Cluster 的复制，也被称为**虚拟同步复制**，实际一种**逻辑同步**，因为每个节点的操作是相互独立的，更为准确的说是异步的。逻辑同步指的是一个事务在一个node上执行成功后，保证它在其他的node上一定会被成功执行，但不保证及时同步。



@稀土掘金技术社区



Willem97 LEVEL 3
软件开发工程师

20
文章

28k
阅读

5
粉丝

关注

私信

目录

收起 ^

常见问题汇总

参考

相关推荐

MariaDB Galera Cluster

662阅读 · 1点赞

MySQL（十三）MYSQL集群---PXC

2.4k阅读 · 4点赞

极数云舟签约Codership，成中国唯一G...

318阅读 · 0点赞

数据库小白变大咖：社招面试MySQL高...

32阅读 · 0点赞

mysql-wsrep+galera集群实现真正意义...

670阅读 · 2点赞

精选内容

糟糕，生产环境频繁Full GC，怎么办？

苏三说技术 · 153阅读 · 3点赞

ZeroTier 源码解析 (5) 交换机 (Switch)

重启的码农 · 15阅读 · 0点赞

ZeroTier 源码解析 (4) 对等节点 (Peer)

重启的码农 · 19阅读 · 0点赞

🌟 PHP 重载（Overloading）——不是...

用户307459698... · 32阅读 · 0点赞

【附录】为什么说 Spring 中 BeanFacto...

dylan_QAQ · 29阅读 · 0点赞

从上图中可以看到，在客户端收到 **DML请求 (Insert/Update/Delete)** 时给到node1后，node1会将其封装为**写集(Write Set)**，并广播到 **node2、node3**， **node2、node3** 收到**写集(Write Set)**后，进行冲突检查，如果没有冲突，先通知 **node1** 进行 **commit**， **node2** 和 **node3** 将数据写入队列，之后再由 **wsrep_slave_threads** 读取进行 **commit**。故其同步必定会存在一定的延迟。

「常见问题汇总」

1. 如果主节点（负责写入的节点）写入过大，apply_cd时间过长，导致数据更新操作时间过长，怎么处理？
- wsrep_slave_threads** 参数配置成cpu的个数或者1.5倍。
2. 脑裂
- 任何命令执行出现unknown command，表示出现脑裂，集群中任意两个节点间通信的4567端口不通，并且无法对外提供服务。

SET GLOBAL wsrep_provider_options="pc.ignore_sb=true"； 通过这个命令来强制恢复出现脑裂的节点。
3. 并发写
- 如果在集群多个节点进行写/更新操作，有可能同时不同节点update同一行操作时就会出现锁死问题，出现>Error: 1213 SQLSTATE:4001.解决：指定更新和写入都在都一个节点操作。
4. DDL全局锁
- 采用 **pt-online-schema-change**
5. 只支持Innodb引擎，表结构必须要有主键，不然会造成集中每个节点的data page里的数据不一致。
- 不支持表级锁，即不能lock/unlock tables，使用行级锁
6. 新节点加入加入&故障节点恢复加入集群，此时不能有写操作，不然会导致被写入的那台库DDL死锁。所以需要暂停集群业务写操作，等数据一致后在开启写操作。

如何掌控Galera Cluster的同步延迟，使其同步延迟减至最小

- 1、调整 **wsrep_slave_threads** 参数：有多少threads可以同时由queue读取资料及commit，建议设置为cpu的1~1.5倍
- 2、调整 **flow control** 参数: 避免任一Node交易时落后其它Node太多，用于协调每个Node，保证commit的速度可优于queue的增长速度

sql 体验AI代码助手 代码解读 复制代码

```
1 # mysql 查询
2 MariaDB [(none)]> show variables like "wsrep_sst_method";
3 +-----+-----+
4 | Variable_name | Value |
5 +-----+-----+
6 | wsrep_sst_method | rsync |
7 +-----+-----+
8
9 1 row in set (0.001 sec)
10
11 MariaDB [(none)]> show variables like "wsrep_slave_threads";
12 +-----+-----+
13 | Variable_name | Value |
14 +-----+-----+
15 | wsrep_slave_threads | 1 |
16 +-----+-----+
17
18 1 row in set (0.001 sec)
```

There might be a small period of time when a replica is out of sync. This happens because the source may apply events faster than the replica. And if you do read from the replica, you may read the data that has not changed yet. You can see that from the diagram.

However, this behavior can be changed by setting the **wsrep_causal_reads=ON** variable. In this case, the read on the replica will wait until the event is applied (this will obviously increase the response time of the read).

The gap between the replica and the source is the reason why this replication is called *virtually synchronous replication*, and not *real synchronous replication*.

参考

- www.sre-devops.info/galera-clus...
- www.percona.com/doc/percona...

找对属于你的技术圈子
回复「进群」加入官方微信群



标签：

MySQL

评论 0



登录 / 注册

即可发布评论!

暂无评论数据

为你推荐

JoinMemory 解决数据装配问题一

coderrxdh | 1年前 | 👁 1.5k | 👍 10 | 💬 2

后端

Java 并发之Future模式（异步）

KittyGuy | 4年前 | 👁 2.3k | 👍 2 | 💬 评论

Java

后端

redis集群解析和水平扩展

Darkness_Icy | 3年前 | 👁 2.2k | 👍 4 | 💬 3

Redis

Redis网络模型

方块海绵 | 1年前 | 👁 374 | 👍 1 | 💬 评论

Redis

JUC- 集合不安全

遇见_line | 4年前 | 👁 782 | 👍 3 | 💬 评论

Java

api设计最佳实践

creny | 4年前 | 👁 1.8k | 👍 12 | 💬 1

API

Protobuf使用性能优化

MasterJiang | 9月前 | 👁 344 | 👍 点赞 | 💬 评论

后端

Java

【前端设计模式】之建造者模式

俊刚 | 1年前 | 👁 412 | 👍 点赞 | 💬 评论

前端

设计模式

Hmily: 高性能异步分布式事务TCC框架

Dromara开源社区 | 4年前 | 👁 1.7k | 👍 5 | 💬 评论

后端

Vue 适时清理 keepalive 缓存方案

ZTStory | 3年前 | 👁 4.9k | 👍 45 | 💬 8

前端

面试中js常问的12个设计模式

器张农民 | 1年前 | 👁 1.5k | 👍 17 | 💬 4

前端

面试

程序员

【前端设计模式】之调停者模式（中介者模式）

俊刚 | 1年前 | 👁 320 | 👍 点赞 | 💬 评论

前端

设计模式

Redisson使用笔记

乘以零 | 3年前 | 👁 2.8k | 👍 6 | 💬 评论

Redis

后端

十六，Redis网络模型

normaling | 3月前 | 👁 61 | 👍 点赞 | 💬 评论

Redis

异步 Promise 以及 Async / Await

去洗澡了晚安 | 5年前 | 👁 1.5k | 👍 3 | 💬 1

JavaSc...

