MySQL Cluster 实战初探

周彦伟

2013.04

先说三个闹心事儿

- 计数
- 用户个人信息
- 好友最新动态

MySQL的解决思路

- 切分
- 多slave
- 缓存

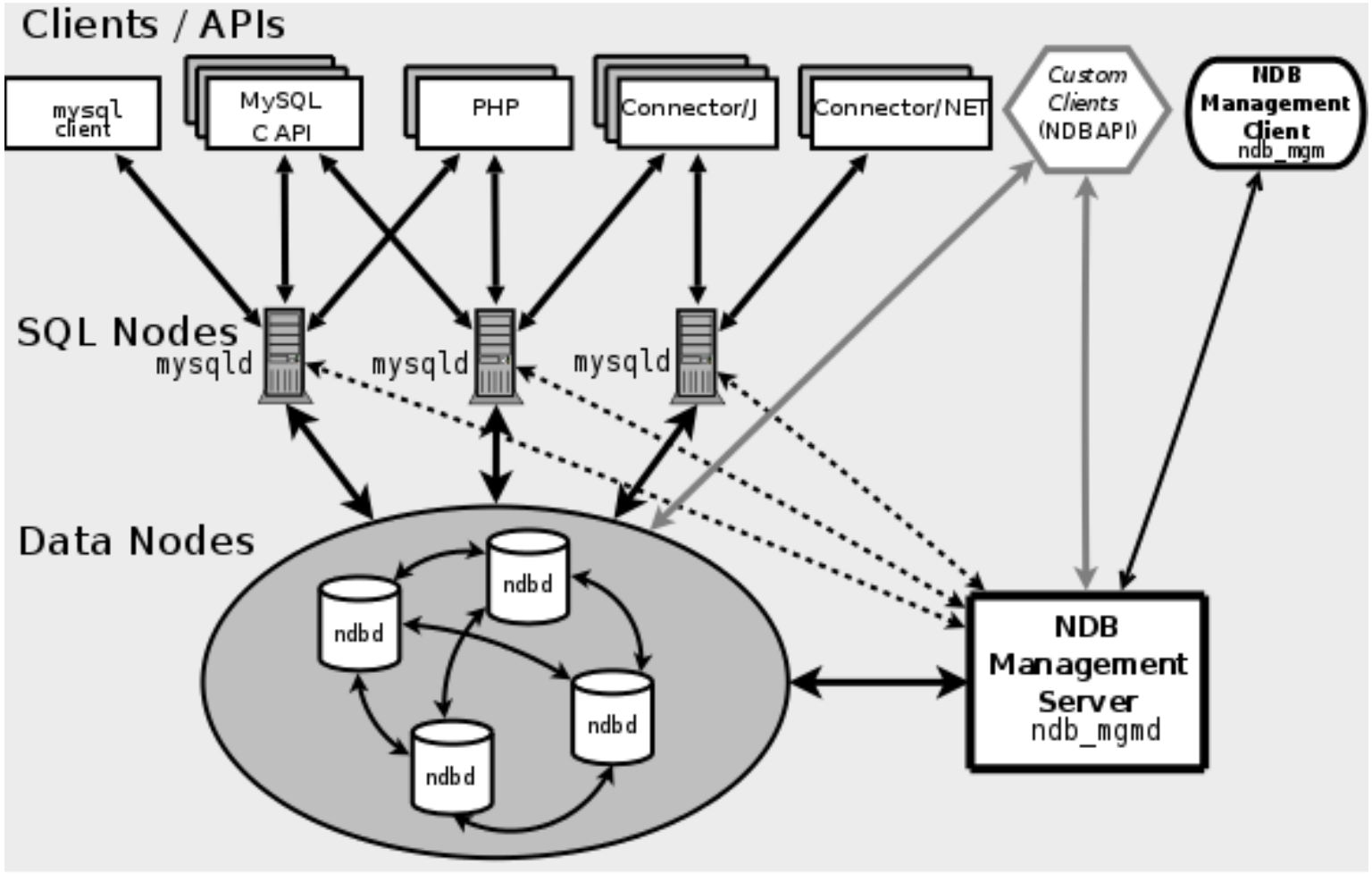
还有没有更合适的方案？

- MySQL Cluster --- 一个美丽的传说

MySQL Cluster

- NDB,NDBCLUSTER,Network Database
- “memory database”
- Share-nothing
- High-availability
- High-redundancy
- Distributed

认识MySQL Cluster



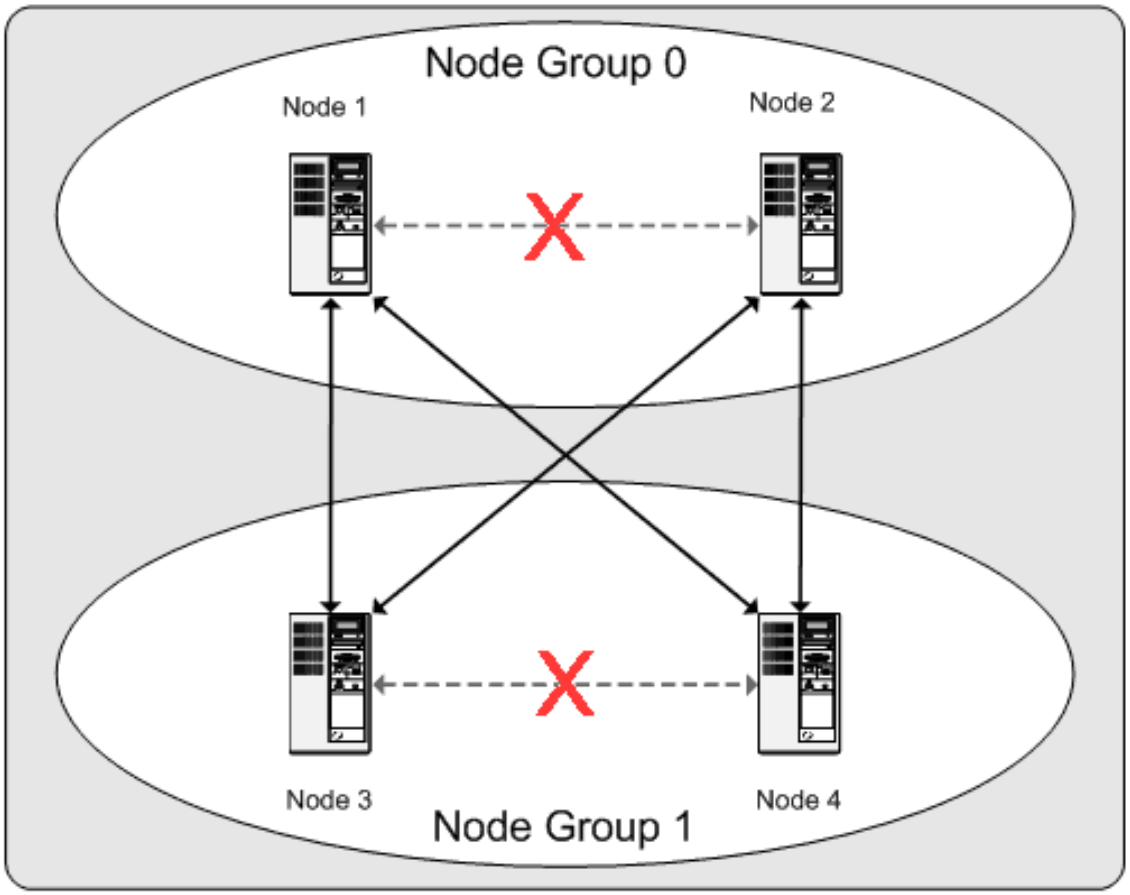
MySQL Cluster

- Managerment node
- Data node
- SQL node
- Management client
- NDB client
- MySQL client

MySQL Cluster

- Node Group
- Partition
- Replicas

HA



部署容易？

需要做的事

- 准备机器（>3)
- 下载两个安装包(client,server)
- 修改两个配置文件
  - my.cnf
  - config.ini

my.cnf

```
[ndb_mgmd]
config-file=/data/mysql-cluster/config.ini

[ndb_mgm]
ndb-connectstring=10.4.26.31:1186;10.4.26.32:1186


[mysql_cluster]
ndb-connectstring=10.4.26.31:1186;10.4.26.32:1186


[mysqld]

ndbcluster

binlog_format=row
log-bin=fb-bin

server-id = 2631
```

Config.ini

```
[ndb_mgmd default]
DataDir=/data/mysql-cluster/var

[ndb_mgmd]
HostName=10.4.26.31
NodeId=231

[ndb_mgmd]
HostName=10.4.26.32
# NodeId=232


[ndbd default]
NoOfReplicas=2
DataMemory=40G
IndexMemory=20G


[ndbd]
HostName=10.4.26.33
NodeId=33
# more [ndbd] ...

[mysqld]
HostName=10.4.26.31
NodeId=101
# more [mysqld] ...
```

启动

- ndb_mgmd –initial
- ndbd –initial
- mysqld_safe –user=mysql &

```
ndb_mgm> show
Connected to Management Server at: 10.4.26.31:1186
Cluster Configuration
---------------------
[ndbd(NDB)]     4 node(s)
id=33   @10.4.26.33  (mysql-5.5.29 ndb-7.2.10, Nodegroup: 0, Master)
id=34   @10.4.26.34  (mysql-5.5.29 ndb-7.2.10, Nodegroup: 0)
id=35   @10.4.26.35  (mysql-5.5.29 ndb-7.2.10, Nodegroup: 1)
id=36   @10.4.26.36  (mysql-5.5.29 ndb-7.2.10, Nodegroup: 1)

[ndb_mgmd(MGM)] 2 node(s)
id=231  @10.4.26.31  (mysql-5.5.29 ndb-7.2.10)
id=232  @10.4.26.32  (mysql-5.5.29 ndb-7.2.10)

[mysqld(API)]   4 node(s)
id=101  @10.4.26.31  (mysql-5.5.29 ndb-7.2.10)
id=102  @10.4.26.32  (mysql-5.5.29 ndb-7.2.10)
id=103 (not connected, accepting connect from any host)
id=104 (not connected, accepting connect from any host)
```

运维-备份

```
ndb_mgm> start backup
Waiting for completed, this may take several minutes
Node 1: Backup 2 started from node 254
Node 1: Backup 2 started from node 254 completed
 StartGCP: 1965887 StopGCP: 1965890
 #Records: 2089 #LogRecords: 0
 Data: 53900 bytes Log: 0 bytes
```

```
[r[redacted] BACKUP-2]# pwd
/data/mysql-cluster/var/BACKUP/BACKUP-2
[r[redacted] BACKUP-2]# ls
BACKUP-2-0.4.Data  BACKUP-2.4.ctl  BACKUP-2.4.log
```

运维-恢复

- ndb_restore

```
[sh]# ndb_restore -n 4 -b 2 -r  --backup-path=/data/mysql-cluster/var/BACKUP/BACKUP-2
Nodeid = 4
Backup Id = 2
backup path = /data/mysql-cluster/var/BACKUP/BACKUP-2
Opening file '/data/mysql-cluster/var/BACKUP/BACKUP-2/BACKUP-2.4.ctl'
File size 29832 bytes
Backup version in files: ndb-6.3.11 ndb version: mysql-5.5.29 ndb-7.2.10
Stop GCP of Backup: 1965889
Connected to ndb!!
Opening file '/data/mysql-cluster/var/BACKUP/BACKUP-2/BACKUP-2-0.4.Data'
File size 12116 bytes
_____________________________________________________
Processing data in table: mysql/def/NDB$BLOB_7_3(8) fragment 3
_____________________________________________________
Processing data in table: mysql/def/ndb_index_stat_head(4) fragment 3
_____________________________________________________
Processing data in table: test/def/t(11) fragment 3
_____________________________________________________
Processing data in table: mysql/def/ndb_schema(7) fragment 3
_____________________________________________________
Processing data in table: mysql/def/ndb_apply_status(9) fragment 3
_____________________________________________________
Processing data in table: sys/def/NDB$EVENTS_0(3) fragment 3
_____________________________________________________
Processing data in table: sys/def/SYSTAB_0(2) fragment 3
_____________________________________________________
Processing data in table: mysql/def/ndb_index_stat_sample(5) fragment 3
Opening file '/data/mysql-cluster/var/BACKUP/BACKUP-2/BACKUP-2.4.log'
File size 52 bytes
Restored 3 tuples and 0 log entries

NDBT_ProgramExit: 0 - OK
```

Rolling restart

- 修改配置
- 增加节点
- 维护机器

Add Data node online

- 修改config.ini
- Rolling restart
- Start new data nodes
- Create nodegroup
- Alter online table … reorganize partition
- Optimize table …

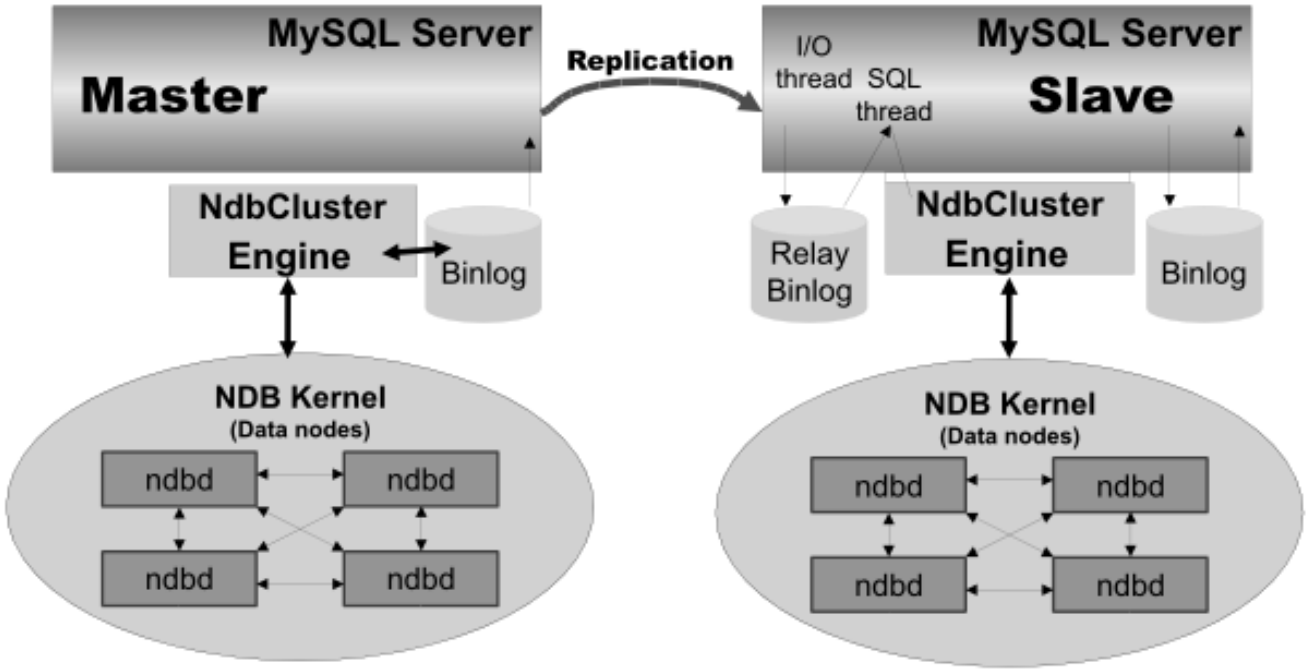
Disk data tables

- Tablespaces
  - Data files
- Undo log files
  - log file grop

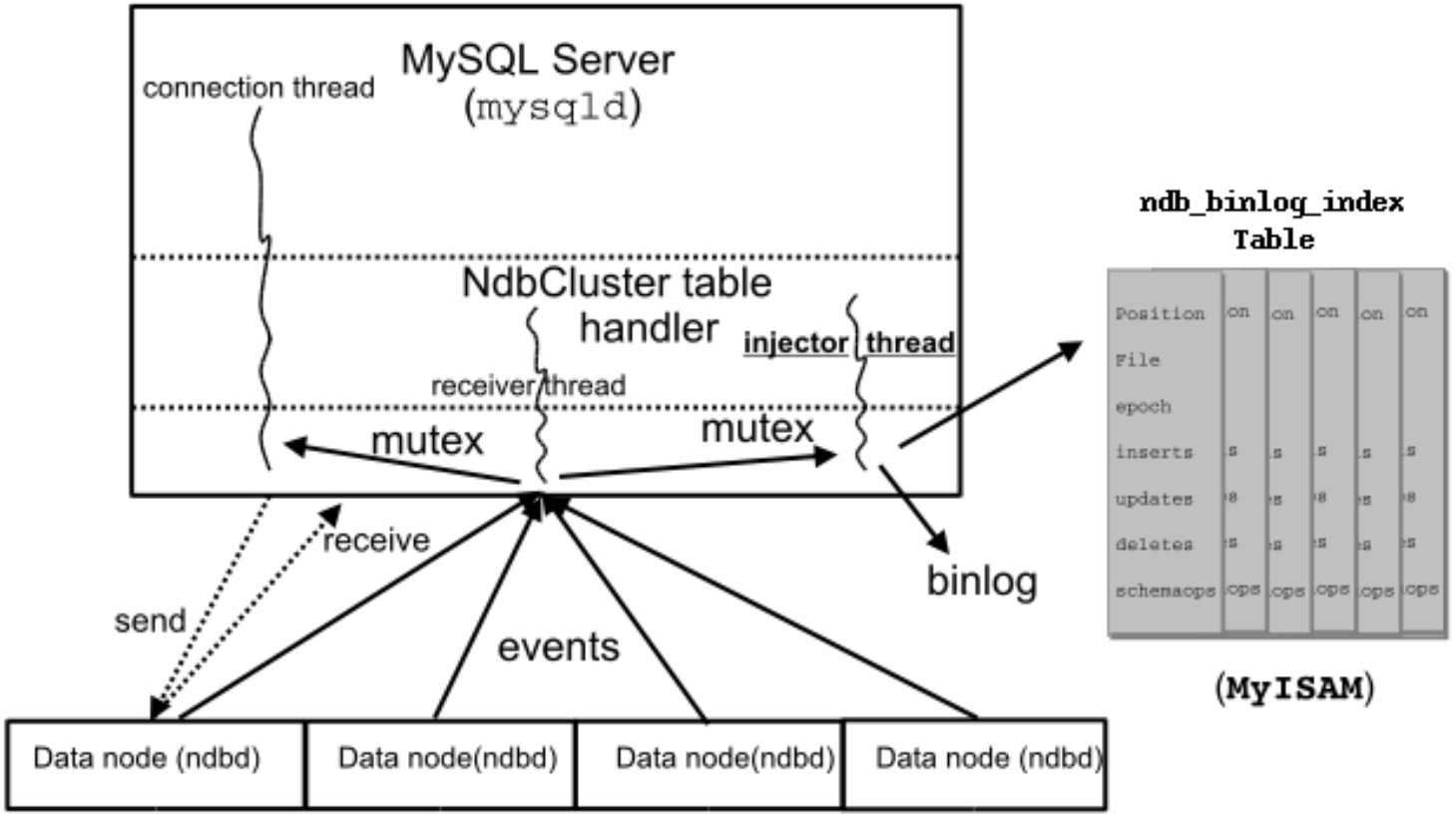
Disk data tables

- create logfile group lg ... engine ndbcluster
- create tablespace ts ... engine ndbcluster
- create table t (...) tablespace ts storage disk engine ndbcluster

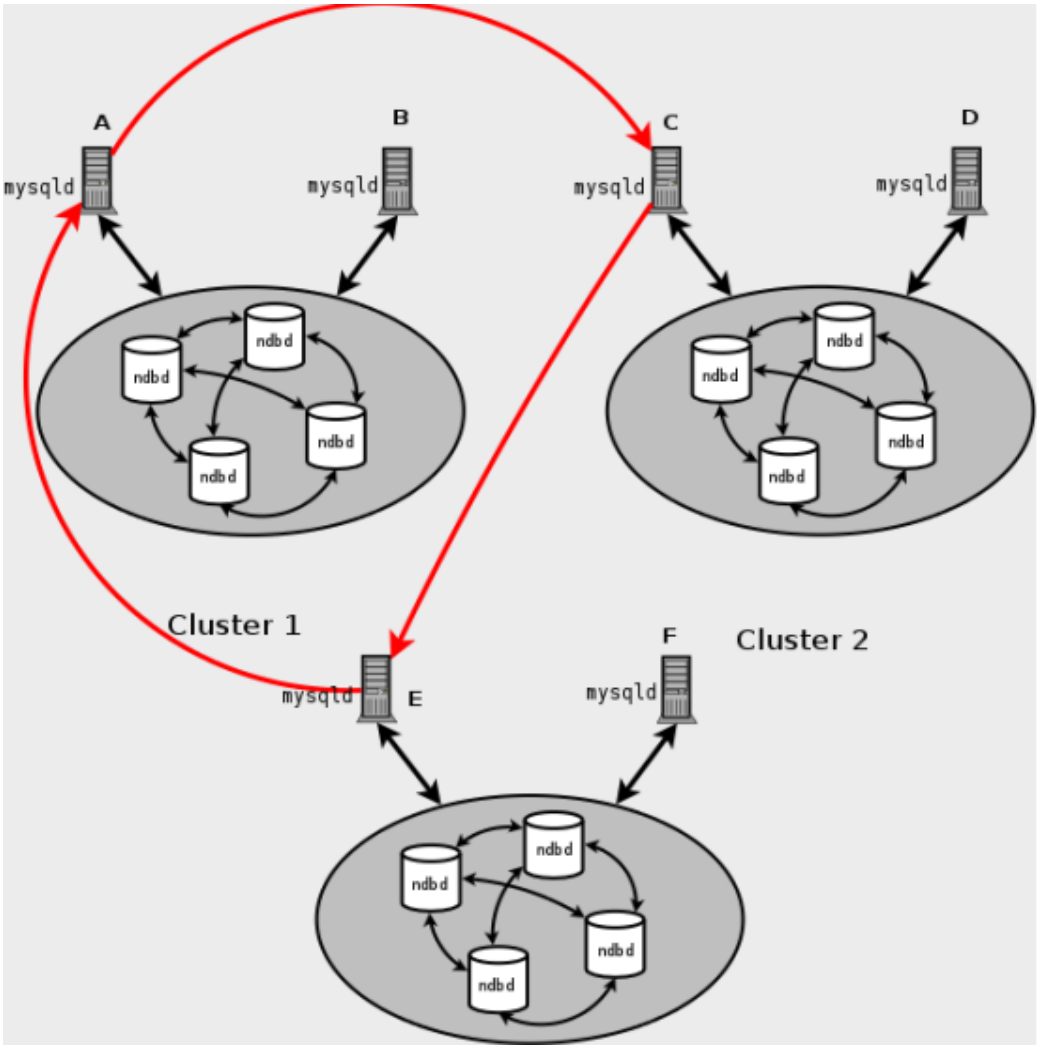
MySQL Cluster replication



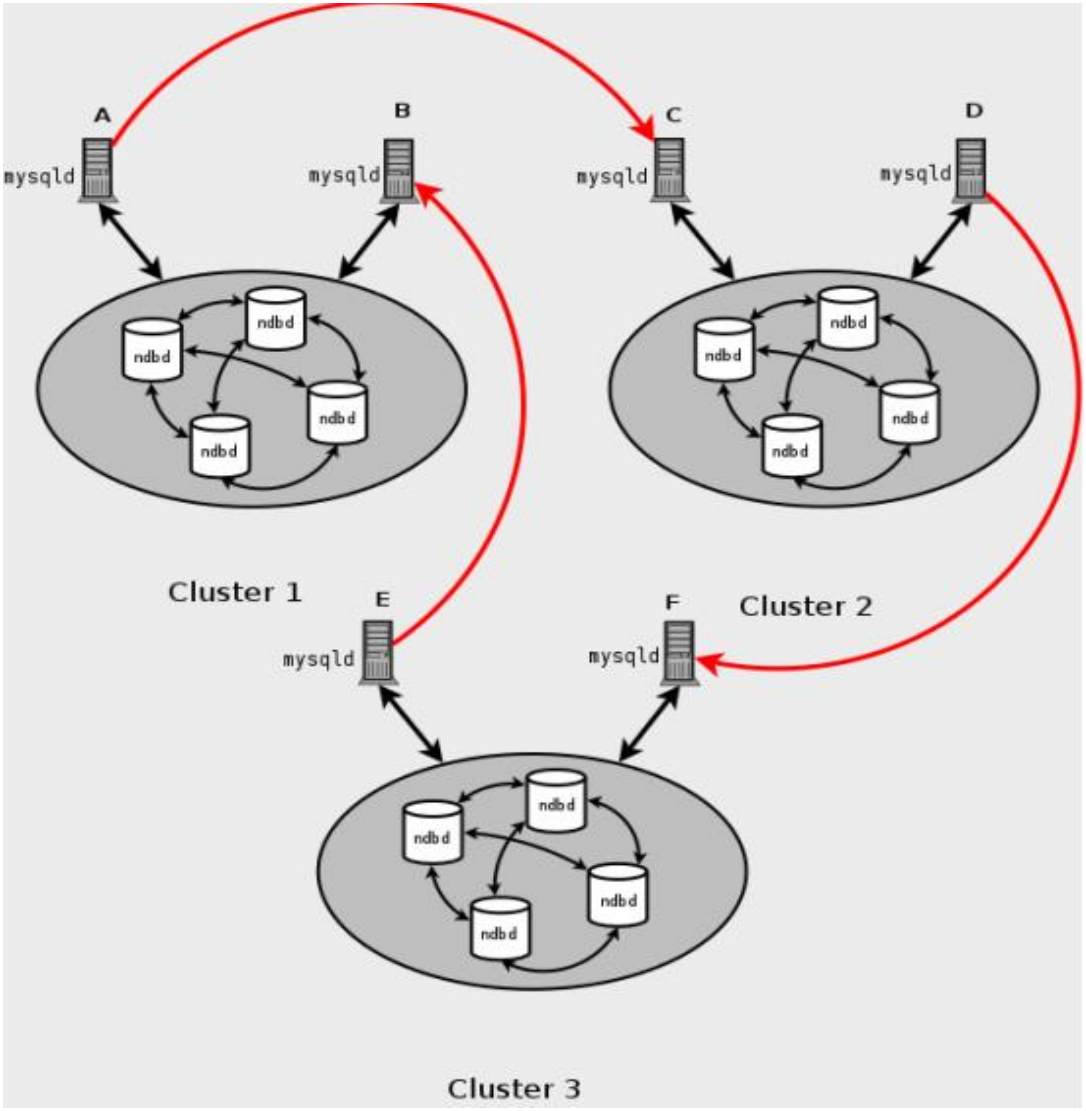
Injector thread



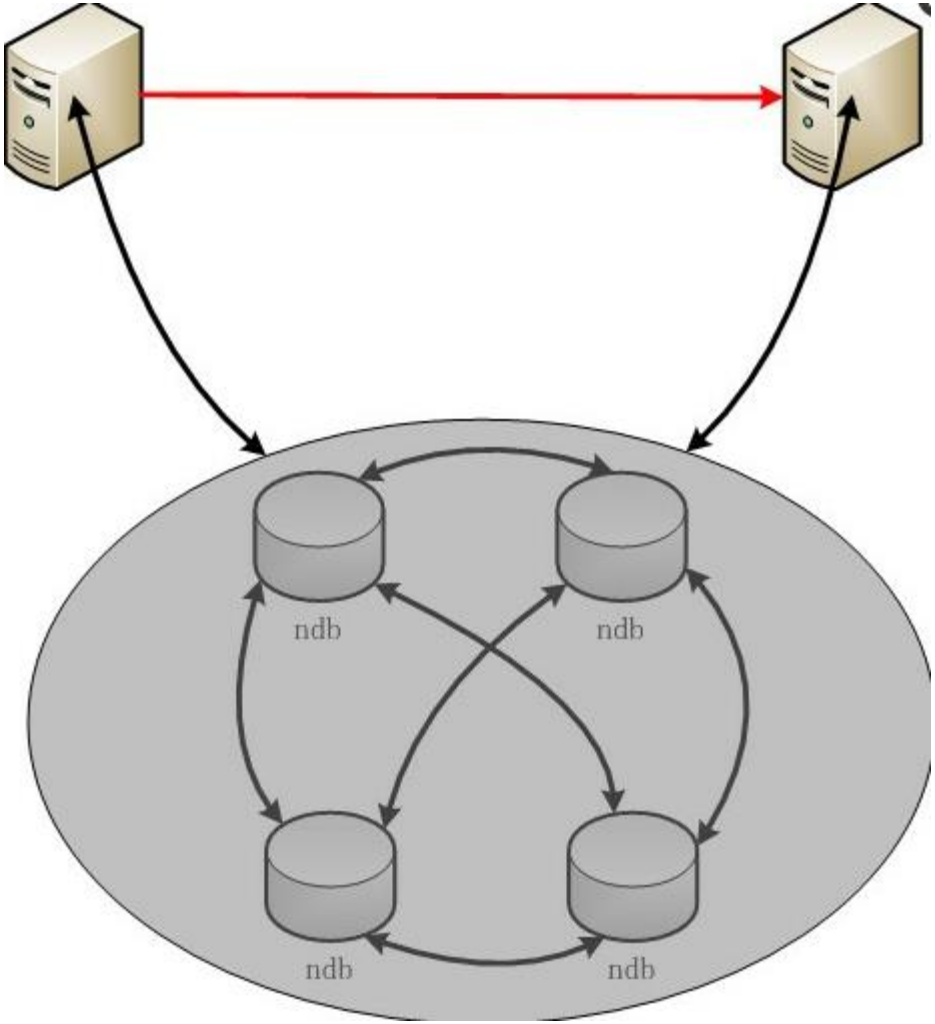
MySQL Cluster replication



MySQL Cluster replication



MySQL Cluster replication



讨论

- 适用场景
  - 密集写
  - 密集读
  - 数据量可控
  - 替代部分缓存
- 使用成本
  - 硬件
    - 用磁盘换内存
  - 运维
    - 标准化，自动化

EOF

@周彦伟
zhouyanwei@gmail.com