Quick Note About Myself

- Database enthusiast.
- Working on MySQL for more than decade.
- Wide interest in data handling and management.
- During my tenure at Yahoo!, Teradata worked on some of the real big-data problems.
- Developed multiple features in InnoDB during 5.7 tenure while working for MySQL/Oracle.
Agenda

- Quick note on “What is Percona XtraDB Cluster (PXC)”?
- Current supported versions of PXC.
- MySQL/PS-8.0 features and how they affect the PXC-8.0.
- What’s new with the PXC-8.0?
- What more to expect with the PXC-8.0?
- Q&A
Percona XtraDB Cluster (PXC)

- Enhanced Security
- Multi-master
- Network protection (Geo-distributed)
- Flexible topology
- Scaling up and down
- Performance tuned
Current Supported Versions of PXC

- **PXC-5.5**
- **PXC-5.6**
- **PXC-5.7**
- **PXC-8.0**

- **EOL**
- **GA**
- **GA**
- **Under development**

Latest GA version:
- PXC-5.6.43
- PXC-5.7.25

Under-development:
- PXC-8.0.15
MySQL/PS-8.0 New Features
MySQL/PS-8.0 New Features

- Atomic DDL
- Data dictionary
- Introduction of ROLES, RESOURCE GROUP
- CATS scheduling algorithm
- Encryption support
- Cloud compatible features (SET PERSIST)

....
Resource Group and PXC-8.0
Resource Group and PXC-8.0

- Resource Group helps grouping of the available resources like vCPU, etc.
- MySQL threads can be assigned to a resource group for execution.
- Administrator can even associate single statement to a given resource group. (Say an important high priority query needs to get executed can be assigned to a specialized resource group).
Resource Group and PXC-8.0

- Given resources are local to a node
- RESOURCE GROUP statements are not replicated (create/alter/drop).

- Admin can now configure a RESOURCE GROUP for applier threads as actions of applier threads needs to run with HIGH PRIORITY. (Can help reduce Flow-Control).
Roles and PXC-8.0
Roles and PXC-8.0

- MySQL-8.0 introduced ROLES to facilitate user and privilege management. (CREATE/DROP/REVOKE/GRANT/….)
- ROLES management statements are replicated across the PXC cluster.
- MySQL needs ROLES to be activated before use. PXC replicates ROLE ACTIVATION too (done through SET command)
Roles and PXC-8.0

- PXC is already working on pre-defining some of the role for general action.
- Like we tried to define `mysql.pxc.sst.role` that can be assigned to backup-user (needed for SST)*.
Atomic DDL and PXC-8.0
Atomic DDL and PXC-8.0

- MySQL-8.0 introduced support for Atomic DDL.
- PXC-5.7 replicates non-atomic statement (like DDL) through TOI and atomic statement (like DML) through write-set based replication [allowing DML transaction to rollback and killed].
- PXC-8.0 continue to replicate DDL through TOI too.
Atomic DDL and PXC-8.0

Important difference to note:

- 2 tables t1 (local to n1) and t2 (replicated on n1 and n2)

<table>
<thead>
<tr>
<th></th>
<th>on-n1</th>
<th>on-n2</th>
</tr>
</thead>
<tbody>
<tr>
<td>drop table t1, t2;</td>
<td>t1, t2 dropped</td>
<td>t2 dropped</td>
</tr>
<tr>
<td>PXC-5.7</td>
<td>t1, t2 dropped</td>
<td>t2 dropped</td>
</tr>
<tr>
<td>PXC-8.0</td>
<td>t1, t2 dropped</td>
<td>no-table-dropped</td>
</tr>
</tbody>
</table>
Atomic DDL and PXC-8.0

- Eventually, PXC plans to make use of atomic nature of DDL and execute the DDL through write-set replication too.
- This will also take-care of blocking DDL issue that exist due to TOI based execution.
Non-Atomic DDL and PXC-8.0
Non-Atomic DDL and PXC-8.0

- MySQL-8.0 continue to support some non-atomic DDL operations
  - OPTIMIZE
  - REPAIR
  - ANALYZE

- PXC continue to support them and there is no change in semantics of these non-atomic DDLs.
SRS (Spatial Reference System) and PXC-8.0
SRS (Spatial Reference System) and PXC-8.0

- MySQL-8.0 introduced SRS support. SRS can be created and dropped using CREATE/DROP command.
- PXC cluster replicates these SRS commands.
XID Inconsistency and PXC-8.0
XID Inconsistency and PXC-8.0

- Now that DDL are transactional, MySQL assigns xid to each DDL statement.
- MySQL has its own logic to assign XID and PXC (wsrep) has its own logic to assign XID.
- With 2 sub-systems in place it can cause duplicate XID generation that can eventually cause problem with recovery (especially with new improved recovery logic of MySQL-8.0 that checks for unique XID).
### XID Inconsistency and PXC-8.0

<table>
<thead>
<tr>
<th>mysql-bin.000001</th>
<th>3027</th>
<th>Gtid</th>
<th>2</th>
<th>3092</th>
<th>SET @@SESSION.GTID_NEXT= '1bc0f40c-8930-ee16-770b-15117ed32fd0:12'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>....</td>
</tr>
<tr>
<td>mysql-bin.000001</td>
<td>3248</td>
<td>Xid</td>
<td>2</td>
<td>3279</td>
<td>COMMIT /* xid=12 */</td>
</tr>
<tr>
<td>mysql-bin.000001</td>
<td>759</td>
<td>Gtid</td>
<td>2</td>
<td>824</td>
<td>SET @@SESSION.GTID_NEXT= '972566d2-51f7-11e9-a3b4-9cbe0ba1a1df:1'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>....</td>
</tr>
<tr>
<td>mysql-bin.000001</td>
<td>980</td>
<td>Xid</td>
<td>2</td>
<td>1011</td>
<td>COMMIT /* xid=12 */</td>
</tr>
</tbody>
</table>
XID Inconsistency and PXC-8.0

- While running in cluster-mode:
  - Only PXC sub-system will be used for assigning XID.
  - Statement that are not replicated (ALTER TABLE DISCARD/IMPORT TABLESPACE) are not logged to binlog.
  - If session turns off cluster-mode (wsrep_on=off) then it will automatically set sql_log_bin=off too.

- This ensures consistent XID generation from PXC sub-system avoiding XID inconsistency.
DD-tables Explicit Locks and PXC-8.0
DD-tables Explicit Locks and PXC-8.0

- PXC force-abort the local running transaction in favor of high-priority replicated transaction Brute-Force-Abort (BFA).
DD-tables Explicit Locks and PXC-8.0

- If the local session is holding an explicit lock (not-recommended in PXC setup) through `FLUSH TABLE WITH READ LOCK`, etc... then such local transactions are non-preemptable. Background high-priority thread has to wait.
With MySQL-8.0, DD-tables modification too establishes EXPLICIT locks as part of statement execution.

PXC tag these MDL locks as preemptable and transaction holding these locks can be force-aborted.
Replication Filter and PXC-8.0
Replication Filter and PXC-8.0

- MySQL-5.7 has concept of Global Replication Filter.
- MySQL-8.0 introduced per channel replication filter.
Replication Filter and PXC-8.0

- Complying with the updated design, PXC-8.0 inherits this concept and now has replication filter associated with wsrep channel.
- Given the multi-master nature of the product, it is not advisable to filter statement for wsrep-channel.
- PXC-8.0 doesn’t allow configuration of wsrep replication filter.
Upgrade Dependency and PXC-8.0
Upgrade Dependency and PXC-8.0

- Given REDO and DD changes, MySQL-8.0 data-directory is not compatible with MySQL-5.7.
- MySQL recommends running upgrade for even minor releases.
- PXC can support upgrade from **5.7-DONOR to 8.0-JOINER** but not vice-versa. SST script has been updated to enforce the needed checks.
Upgrade Dependency and PXC-8.0

- PXC will not support upgrade from 5.6 to 8.0 (5.6-DONOR to 8.0-JOINER). User should upgrade to 5.7 and then continue to upgrade further.
- PXC strongly recommend upgrading to PXC-5.7.25+ before upgrading to 8.0 to ensure proper enforcement of upgrade checks.
Upgrade post SST with PXC-8.0
Upgrade post SST with PXC-8.0

DONOR node

backup

XB-backup

XB-prepare (REDO log recovery)

JOINER node

XB-move

SST using XB
Upgrade post SST with PXC-8.0

- MySQL-8.0 has changed REDO log format.
- This means copying over live 5.7-data-directory and trying to recover it through XB-8.0 binaries (that are MySQL-8.0 compatible) will not work.
- This requirement demands:
  - If DONOR is < 8.0 then use XB-2.4 to prepare.
  - If DONOR is > 8.0 then use XB-8.0 to prepare.
Upgrade post SST with PXC-8.0

- Node can’t have 2 packages of XB installed on the same node.
- Also, in past we have seen user facing restriction in using different XB binaries due to its dependency on PXC.

Both of these pain-point are now addressed with PXC-8.0.

PXC-8.0 package will now ship its own version of XB binaries to allow auto-node provisioning from 5.7-DONOR and 8.0-JOINER.
Upgrade post SST with PXC-8.0

- PXC-8.0 will only support XB based SST.
- mysqldump was deprecated in 5.7 and has been removed from PXC-8.0.
- rsync is not supported due to REDO log format change. Will be removed from PXC-8.0.
Error/Info/Warn Logging with PXC-8.0
MySQL-8.0 improved the error/info/warning logging framework.

User can route all the messages through logging component(s) configurable through log_error_services.

This features delayed logging of the message during server start till the logging module is loaded. Messages are cached and once the the logging module is loaded messages are routed accordingly.
Error/Info/Warn Logging with PXC-8.0

- MySQL-8.0 also emit the module name representing where the message is originating from

2019-05-14T07:22:50.264963Z 3 [Note] [MY-011089] [Server] Data dictionary restarting version '80014'.
2019-05-14T07:22:50.791202Z 3 [Note] [MY-012357] [InnoDB] Reading DD tablespace files
2019-05-14T07:22:50.794591Z 3 [Note] [MY-012356] [InnoDB] Validated 6/6 tablespaces
Error/Info/Warn Logging with PXC-8.0

- PXC-8.0 inherit this new logging framework and PXC messages are now logged with module name too.
- PXC defines 3 modules:
  - WSREP
  - GALERA
  - WSREP-SST
- Also, worth noting that SST originating messages from an external process (wsrep-xtrabackup-v2) are now redirected to MySQL process for consistent logging and processing.
2019-05-14T09:07:36.728088Z 0 [Note] [MY-000000] [WSREP-SST] (debug) Cleaning up temporary directories
2019-05-14T09:07:36.736437Z 0 [Note] [MY-000000] [Galera] 0.0 (n1): State transfer to 1.0 (n2) complete.
2019-05-14T09:07:36.736459Z 0 [Note] [MY-000000] [Galera] Shifting DONOR/DESYNCHED -> JOINED (TO: 0)
2019-05-14T09:07:36.736758Z 0 [Note] [MY-000000] [Galera] Member 0.0 (n1) synced with group.
2019-05-14T09:07:36.736771Z 0 [Note] [MY-000000] [Galera] Shifting JOINED -> SYNCHED (TO: 0)
2019-05-14T09:07:36.736859Z 2 [Note] [MY-000000] [WSREP] Synchronized with group, ready for connections
2019-05-14T09:07:36.736887Z 2 [Note] [MY-000000] [WSREP] Setting wsrep_ready to true
2019-05-14T09:07:36.736913Z 2 [Note] [MY-000000] [WSREP] wsrep_notify_cmd is not defined, skipping notification.
2019-05-14T09:07:48.011987Z 0 [Note] [MY-000000] [Galera] 1.0 (n2): State transfer from 0.0 (n1) complete.
2019-05-14T09:07:48.012500Z 0 [Note] [MY-000000] [Galera] Member 1.0 (n2) synced with group.
As mentioned above, if there is no error during the startup, MySQL will delay logging/processing of the messages till log module it loaded.

In PXC, log module is loaded post SST and so if there is no error message during SST then appearance of first log message may take time.

In case below, after starting node-2, messages started appearing after 44 seconds

```
...

2019-05-15T08:00:08.957156Z 0 [Note] [MY-011245] [Server] Plugin mysqlx reported: 'Scheduler 'network', create threads'
```
CATS and PXC-8.0
CATS and PXC-8.0

- MySQL-8.0 introduced Contention Aware Transaction Scheduling algorithm. A transaction that can unblock majority of the waiting/blocked transactions, is first to get hold of the lock.
- PXC has a concept of applier that needs to take priority, irrespective of the number of waiting locks/transactions.
- PXC-8.0 doesn’t yet support CATS instead continue to use FIFO scheduling.
XPlugin and PXC-8.0
XPlugin and PXC-8.0

- In MySQL-8.0 xplugin is enabled by default (to be used with new mysql shell).
- X-Plugin loading sequence (during boot) fire queries against server.
- PXC can’t start accepting queries till PXC is in SYNCED mode.
- This means with PXC, X-plugin is loaded once PXC enter synced state.
- If SST or IST is involved, sync may take longer and so would start of mysqlx plugin.
Deprecated/Removed and PXC-8.0
XPlugin and PXC-8.0

- `mysqldump`: deprecated in PXC-5.7
  - Logical backup. No takers
- `rsync`: limited support (deprecated/will be removed in PXC-8.0)
  - REDO log format change.
- `wsrep_force_binlog_format`: deprecated in PXC-5.7
  - PXC-8.0 operates only with `binlog_format=ROW`.
- `wsrep_convert_lock_to_trx`: deprecated in PXC-5.7.
- `wsrep_preordered`: deprecated in PXC-5.7
  - With performance fix this is no more needed.
XPlugin and PXC-8.0

- **innodb_disallow_writes**: deprecated in PXC-5.7
  - Deprecated in favor of innodb_read_only.
  - Was used by rsync. Rsync is no more supported.
- `session level binlog_format=STATEMENT`: deprecated in PXC-5.7.
- **wsrep_drupal_282555_workaround**: deprecated in PXC-5.7.
  - Auto-increment bug that caused duplicate value generation is now fixed.
Create Table As Select and PXC-8.0
Create Table As Select and PXC-8.0

- Known to introduce GTID inconsistency with MySQL-8.0 so blocked with gtid_mode=ON.
- PXC doesn’t recommend use of CTAS (already blocked with pxc_strict_mode=ENFORCING).
- CTAS further leads to XID inconsistency with MySQL-8.0. Reported here (bug#93948).
- PXC-8.0 will now replicate CTAS through TOI (Total Order Isolation) just like other DDLs.
Create Table As Select and PXC-8.0

- This has the following repercussions:
  - `CREATE TABLE dest SELECT * from src;`

<table>
<thead>
<tr>
<th></th>
<th>n1-node</th>
<th>n2-node</th>
</tr>
</thead>
<tbody>
<tr>
<td>wsrep-replicate-myisam=off</td>
<td>dest table created and loaded</td>
<td>dest table created but not loaded</td>
</tr>
<tr>
<td>PXC-5.7</td>
<td>dest table created and loaded</td>
<td>dest table created and loaded</td>
</tr>
<tr>
<td>PXC-8.0</td>
<td>dest table created and loaded</td>
<td>dest table created and loaded</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>n1-node</th>
<th>n2-node</th>
</tr>
</thead>
<tbody>
<tr>
<td>source table local</td>
<td>dest table created and loaded</td>
<td>dest table created and loaded</td>
</tr>
<tr>
<td>PXC-5.7</td>
<td>dest table created and loaded</td>
<td>dest table created and loaded</td>
</tr>
<tr>
<td>PXC-8.0</td>
<td>dest table created and loaded</td>
<td>dest table not-created</td>
</tr>
</tbody>
</table>
SET PERSIST (mysqld-auto.cnf) and PXC-8.0
SET PERSIST (mysqld-auto.cnf) and PXC-8.0

- MySQL-8.0 allows user to persist the changed configuration across restart. To facilitate this, it creates a file named mysqld-auto.cnf.
- This file is not copied over as part of SST from DONOR to JOINER.
What’s new with PXC-8.0?
What’s new with PXC-8.0?

- Internal SST user (getting rid of wsrep_sst_auth).
- Auto-upgrade.
- Avoid setting up automatic slave if DONOR is slave.
- Additional stats through SHOW STATUS.
Internal SST User
Internal SST User

- For doing SST, PXC demands presence of user on DONOR node that can then be used to take backup from DONOR.

- This creates 3 problems for administrator:
  - Administrator needs to create the said user, grant it correct privileges.
  - User is retained even post SST (as next SST may need it).
  - Credentials for the said user are specified in human readable format in configuration file. Raising Security Alarm.
Internal SST user

- PXC-8.0 got rid of all this. wsrep_sst_auth is now removed from PXC-8.0.
- PXC-8.0 introduces concept of internal-sst-user (mysql.pxc.sst.user).
- **On DONOR:**
  - This user is created on donor when sst action start and is removed on completion of sst action.
  - User is assigned a random on the fly generated password.
  - Needed privileges are worked out internally and assigned to user to take backup.

*(JOINER doesn’t need user for completing SST).*
Auto-Upgrade

- MySQL recommends running upgrade even for minor version upgrade.
- With PXC DONOR and JOINER it could be challenging.
- Let’s take a use-case:
  - DONOR is 5.7 and JOINER is 8.0.
  - JOINER will get SST from DONOR. SST is prepared and restore using XB-2.4 (now PXC ships XB-2.4 and XB-8.0 binaries as part of PXC packages).
  - Post SST, server startup resumes and fails as data-directory is not compatible.
  - User now needs to restart server in standalone mode.
  - Run upgrade.
  - Restart server in cluster mode hoping it will join using IST (else repeat).
- This defeats the purpose of auto-node provisioning
PXC introduces auto-upgrade.

Simply boot the JOINER. JOINER and DONOR will handshake to see if the version are compatible ensuring DONOR version <= JOINER version.

On successful SST, JOINER will automatically restart the server for mysql-upgrade and upgraded data-directory is then handed over to main flow for further processing.

Takes care of minor version upgrade too.

*Very helpful with Cloud operation.*
Skip Setting Multiple Slave Post SST
One other problem that is often reported post SST is automatic spawn-up of multiple slaves when user has configured only one of the PXC node as slave.

If DONOR node is acting as active async slave then post SST the slave_master_info is copied over there-by creating an unwanted additional async slave.

No more unconfigured slaves with PXC-8.0.
Skip Setting Multiple Slave Post SST

- PXC-8.0 as part of post-processing executes `RESET SLAVE ALL` to clear the slave information that is copied over as part of SST.

- This all happens behind the scene as part of SST process. So SST process is not merely copying over of the data-directory.

**PXC-8.0 is equipped with SMART SST.**
Additional Stats Through SHOW STATUS
PXC has concept of monitor. Monitor is configured with condition that decide which transactions are allowed to pass through it and execute in critical section.

For example:
- ApplyMonitor allows parallel application of the transaction.
- CommitMonitor allows only one transaction (ordered by sequence number) to proceed with action.
- LocalMonitor is meant to order local action including sync-wait/pause.

These monitors play important role in understanding which transactions/threads are operating in which part of critical section.

Additional Stats through SHOW STATUS
Additional Stats through SHOW STATUS

- Stats for this monitor is now exposed through SHOW STATUS

```
| wsrep_monitor_status (L/A/C) | [ (17, 10), (3, 3), (3, 3) ] |
```

- L/A/C represent Local/Apply/Commit and pair represent (last_entered_, last_left_)
- Say for L = 17, 10 that suggest last_left_ = 10 and currently local monitor is being held by write-set with seqno=11 and last_entered_ = 17 (that is waiting) to enter critical section.
What More to Expect From the PXC-8.x?
What More to Expect From the PXC-8.x?

.... And list is not done yet

More features are being worked out.
- improved resilient cluster.
- cloud friendly cluster.
- inheriting complete encryption support form upstream.
- improved packaging.
- .....
Thank You to Our Sponsors
Rate My Session

Introducing gh-ost: triggerless, painless, trusted online schema migrations

Rate & Review
Tap a star to rate
Feedback (optional)

SUBMIT