Practical Distributed Processing

Using MySQL Built-In Functionality

Bob Burgess – radian6 Technologies
MySQL User Conference 2010
Who Am I

- Database Analyst, radian

- Several Years Experience:
  - Sybase, Oracle, MySQL
  - AIX, Linux
  - Programming
Scope

- MySQL concepts used in this system
- How I solved our particular problem
- Practical set-up of our distributed system
Scope – NOT

- Complete course on distributed processing
- Complete coverage of related MySQL functionality
The Problem

- “Influencer” calculation: complex and getting more complex
- More and more data
- More and more customers and profiles
- Not scalable to run the stored procedure on just the main database
Design Goals

- Move calculation off of main database
- Keep existing stored procedure
- No single controller
- Horizontally scalable!
“Influencer” calculation

Who’s most influential? (For this topic.)

- subset of database
- sphinx index
- do stuff
- influencer data (two MyISAM tables)
MySQL Functionality Used
MySQL Functionality Used

- Replication
- Blackhole Storage Engine
- Federated Storage Engine
- Public-Key Authentication
Replication

Master Database

binlogs
INSERT...
UPDATE...
INSERT...
DELETE...

Replica Database

relay logs
INSERT...
UPDATE...
INSERT...
DELETE...
Replication

server_id = 1
log_bin = /store/log/mysql-bin

grant replication slave on *.*
to 'repl'@'%' identified by 'secret';

/store/log/mysql-bin.000001
Replication

**Server Configuration**

```sql
server_id = 2
relaylog = /store/relaylog/relaylog

change master to
  master_host="masterserver",
  master_port=3306,
  master_user="repl",
  master_password="secret",
  master_log_file="mysql-bin.000001",
  master_log_pos=0;
start slave;
```

**Database Replica**

- Replica Database
- Relay Logs
  - Insert...
  - Update...
  - Insert...
  - Delete...
- `/store/relaylog/relaylog.000001`
Black Hole Tables

create table t (a int) engine=blackhole;
Query OK, 0 rows affected (0.02 sec)

insert into t values (1);
Query OK, 1 row affected (0.00 sec)

select * from t;
Empty set (0.00 sec)
create trigger t_bi
before insert on t
for each row
insert into real_table values (new.a);
Federated Tables

- Points to a table in another database

create table tm (a int) engine=federated
connection="mysql://user:pass@server1/schemaname/t"

create table t (a int)

Not enabled by default in 5.1!
Federated Tables

- Points to a table in another database

```sql
create table t (a int)
create table tm (a int) engine=federated
connection="mysql://user:pass@server1/schemaname/t"
```
Public-Key Authentication

- Secure shell (ssh) or secure copy (scp) without a password
Public-Key Authentication

# scp file mysql@main:/...
# cd /root/.ssh
# ssh-keygen
# cat id_rsa.pub
# vi /home/mysql/.ssh/authorized_keys
Distributed System – Why It Works For Us

- All involved *database* data fits on the worker database
- Calculation produces a self-contained data set
- MyISAM can be used for the resulting data
Distributed System – General Architecture

main db server

main db

replication

federated

worker db

worker 1

MYD & MYI

scp file copy
Distributed System –
General Architecture

main db server

worker 1

worker 2

worker 3

main db

worker db

MYD & MYI

worker db

MYD & MYI

worker db

MYD & MYI
# The Non-Queue... Queue

--- Topics ---
<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>Apples</td>
</tr>
<tr>
<td>102</td>
<td>Oranges</td>
</tr>
<tr>
<td>103</td>
<td>Grapes</td>
</tr>
<tr>
<td>104</td>
<td>Bananas</td>
</tr>
<tr>
<td>105</td>
<td>Pears</td>
</tr>
</tbody>
</table>

--- To-do List ---
- 101
- 104
- 105
The Non-Queue... Queue

--- Topics ---

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Recalc -Reqd</th>
<th>Recalc -Date</th>
<th>Priority</th>
<th>Running</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>Apples</td>
<td>Full</td>
<td>10:00am</td>
<td>high</td>
<td>101</td>
</tr>
<tr>
<td>102</td>
<td>Oranges</td>
<td>No</td>
<td>8:00am</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>103</td>
<td>Grapes</td>
<td>No</td>
<td>7:30am</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>104</td>
<td>Bananas</td>
<td>Partial</td>
<td>10:10am</td>
<td>high</td>
<td>302</td>
</tr>
<tr>
<td>105</td>
<td>Pears</td>
<td>Full</td>
<td>10:05am</td>
<td>low</td>
<td>0</td>
</tr>
</tbody>
</table>

- not running
- highest priority
- earliest recalc date
Processing Cycle

Get work...

Get next Topic to do

is there any?

sleep

no

yes
Processing Cycle

Is data current? Let replication catch up!

replication stopped?

yes → sleep 10 min.

replication behind?

yes → sleep 1 min.
Processing Cycle

Replace stale Topic Profile – someone else likely has it...

waited for replication?

get a new Topic to do
Processing Cycle

Playing Nice with Others

mark Topic with my ID

sleep short random time

Topic still has my ID?

no

sleep 1 min.

begin again
Processing Cycle

- run the stored proc

  success ?

  - yes
  - no

    - log the error
    - sleep 1 min.
    - begin again

- yes

- no
flush the data table for this Profile (locally)

copy to main db server

flush the data table for this Profile (on main db)

update Profile table on main: *not running*
Processing Cycle

back to the top
trap 'STOP=1' TERM
STOP=0

while [ $STOP -eq 0 ]; do

done
Code – who am I?

```plaintext
NODEID_QUERY="select @@server_id-1000;"
INSTANCE=$2

NODEID=`$MYSQL -e"$NODEID_QUERY"`
UNIQUE_INSTANCEID=$((NODEID*100+INSTANCE))
```
NEEDY=`$MYSQL_MASTER -e"$NEEDY_TP_QUERY" 2>/tmp/mysql.err`
set -- $NEEDY
TPID=$1; RECALC_SCOPE=$2; RECALC_DATE=$3
Code – who’s next, the SQL

NEEDY_TP_QUERY="
    select topicFilterId,influencerRecalcReqd,
    unix_timestamp(influencerRecalcDate) RecalcDate
    from TopicFilter
    where influencerRecalcReqd>0
        and influencerRunning=0
        and influencerRecalcDate<=now()
        and active in (1,2)
        and influencerPriority>=$MINPRIORITY
    order by influencerPriority desc,
        influencerRecalcDate asc,
        topicFilterId
    limit 1;"
while [ ! -z "${TPIID}" -a $STOP -eq 0 ]; do
done
Code – how far behind?

REPLICATION_DELAY=
`$MYSQL -e'show slave status\G'
  | grep "Seconds_Behind_Master"
  | cut -f2 -d':'`
Code – call the proc

$MYSQL -e "call proc(${TPID},${RECALC_SCOPE})"
  >>${INFLUENCER_LOG}${TPID}.${INSTANCE}.log
  2>&1
RUN_SUCCESS=`tail -1
  ${INFLUENCER_LOG}${TPID}.${INSTANCE}.log
  | grep -c SUMMARY`
if [ $RUN_SUCCESS -eq 1 ]; then
  copy tables
else
  report error
fi
Code – send the result back

$MYSQL -e"use influencer;
            flush tables Dynamics_${TPID};"
if scp -p -B ${TABLE_SOURCE}/Dynamics_${TPID}.*
             ${TABLE_MASTER_PUSH_LOC} 2>&1 && \
    scp -p -B ${TABLE_SOURCE}/Post_${TPID}.*
             ${TABLE_MASTER_PUSH_LOC} 2>&1
then
    copy files to backup
    remove tables from worker
else
    report error
fi
Code – was I killed?

if [ $STOP -eq 1 ]; then
    echo "---- Script stopped by Kill (TERM signal) ----"
fi
create table BlogPost (...)  
engine=blackhole;

create table PostAuthor (  
    blogPostId bigint not null primary key,  
    author varchar(189))  
default character set=utf8;

create trigger BlogPost_bir  
before insert on BlogPost for each row  
begin  
    if new.mediaProviderId>1 then  
        insert ignore into PostAuthor (blogPostId,author)  
            values (new.blogPostId, new.author);  
    end if;  
end if;
end;
Launch Script

- Shell script to launch workers
- On
  - Starts several workers
- Off
  - sends a signal to each worker
  - stops after this stored proc call
- Kill
  - abort and clean up database
Monitor Scripts

- Replication
- Disk space
- Errors from the worker script
  - Procedure fails
  - scp fails
  - gzip fails
  - replication is behind
Design Decisions

- MyISAM vs. InnoDB
  - Size
  - Tables easily copied
  - Table locking / Replication: Smaller Queries Required

- SSD Storage
Development Directions

- Migrate stored proc to Java
- More workers
- Problems in Code – not atomic:
  - flush table / copy table
  - copy result tables to main db
Take Away...

- Replication and Federated tables aren't hard.
- Thousands of MyISAM tables isn't crazy.
- Distributed processing isn't rocket surgery.
Thank you!

Bob Burgess
bob.burgess@radian6.com

Slides on the conference site.
Email me for sample scripts.