

## MySQL 代码学习和调试



宋利兵

2012-7



- 一、获取 MySQL 源代码
- 二、编译 MySQL 源代码
- 三、运行 MySQL
- 四、调试 MySQL 源代码
- 五、阅读 MySQL 源代码







## 哪里有代码

- Launchpad

https://launchpad.net/mysq1

https://launchpad.net/maria

- MySQL Website

http://dev.mysq1.com/downloads

推荐从Launchpad 获取代码,因为Launchpad上的代码是存储在代码库中的。包含了一些重要的信息,对理解代码很有帮助。

# 从 launchpad 取代码

```
* bzr 命令克隆
```

```
bzr branch 1p:mysq1-server/5.1
bzr branch 1p:mysq1-server/5.5
bzr branch 1p:mysq1-server/trunk
```

#### \*直接拷贝bzr库目录

.bzr 是mysq1代码库中的隐藏目录 可以直接拷贝此目录来克隆 MySQL 代码库

#### \* 更新代码

```
bzr pull 1p:mysql-server/5.5
```

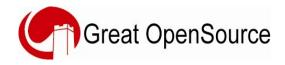


# 从 launchpad 取代码

#### \* 创建公共代码库

```
为各个分支创建一个公共的代码库可以节约空间和克隆时间
mkdir bzrroot
bzr init-repo
cd bzrroot
bzr branch 1p:mysq1-server/trunk
```

bzr branch 1p:mysq1-server/5.5



```
revno: 3820 [merge]
tags: clone-5.5.25-build, mysql-5.5.25
revision-id: annamalai.gurusami@oracle.com-20120510050316-6ct901cvswro4a5y
parent: qeorqi.kodinov@oracle.com-20120509140844-iit7r7vo2rtx6lm4
parent: annamalai.gurusami@oracle.com-20120510044831-d4xkpk2ky5sioeeq
committer: Annamalai Gurusami <annamalai.qurusami@oracle.com>
branch nick: mysql-5.5
timestamp: Thu 2012-05-10 10:33:16 +0530
message:
 Merging from mysql-5.1 to mysql-5.5.
    revno: 2661.810.8
    revision-id: annamalai.gurusami@oracle.com-20120510044831-d4xkpk2ky5sioeeq
    parent: sunanda.menon@oracle.com-20120508051914-6hjecro5b7tef2tx
    committer: Annamalai Gurusami <annamalai.gurusami@oracle.com>
    branch nick: mysgl-5.1
   timestamp: Thu 2012-05-10 10:18:31 +0530
   message:
     Bug #14007649 65111: INNODB SOMETIMES FAILS TO UPDATE ROWS INSERTED
      BY A CONCURRENT TRANSACTIO
     The member function QUICK RANGE SELECT::init ror merged scan() performs
      a table handler clone. Innodb does not provide a clone operation.
      The ha innobase::clone() is not there. The handler::clone() does not
      take care of the ha innobase->prebuilt->select lock type. Because of
```

北京万里开源软件有限公司



#### \* Revision Log

每次提交都会创建一个新的 revision. Revision log 中包含了丰富的信息。

- 版本信息 revno, revid, tags, parent
- 作者 committer
- 时间 timestamp
- 提交日志 message



#### \*提交日志

MySQL 的提交日志的格式

- Patch 的标题

```
Bug#ID <title>
```

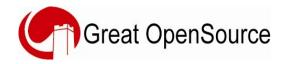
Bug#<oracle bug id> <mysq1 bug id> <title>

http://bugs.mysq1.com/

WL#ID <title>

http://forge.mysq1.com/worklog/

- Bug 或 Feature 的描述



- 显示 revision 日志 bzr log [—include-merges]

bzr log [—include-merges] -r <revision tag>

> bzr 1og -r 1i-bing.song@sun.com-20100328115733-41psnybf8hftnpug

\_\_\_\_\_

revno: 2661.584.23

committer: <Li-Bing.Song@sun.com>

branch nick: mysq1-5.1-bugteam

timestamp: Sun 2010-03-28 19:57:33 +0800

message:

Bug #50407 mysq1bin1og --database=X produces bad output for SAVEPOINTs

. . .



- 显示 revision 的 patch
bzr log [—include-merges] -r <revision> -p

```
> bzr log -p -r
li-bing.song@sun.com-20100328115733-41psnybf8hftnpug
```

```
diff:
=== modified file 'client/mysqlbinlog.cc'
--- client/mysqlbinlog.cc 2010-02-17 18:07:28 +0000
+++ client/mysqlbinlog.cc 2010-03-28 11:57:33 +0000
@@ -730,9 +730,7 @@
```



参考:http://www.mysq1ops.com/2011/11/16/bazaar-annotate-快速找到引入 代码的版本.htm1



- Revision
- \* Local Revision Number(revno)

It is a number, e.g. 1, 21, 1001, etc.

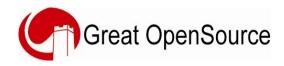
\* Global Revision ID(revid)

包含了作者,时间等信息.

'anders.song@greatopensource.com-20110115055451-dps4tyqx1bkn7vgv'

当在分支之间进行合并时, revid 也会被合并进去。所以是全局唯一的。通常使用 revid 来查找一个 patch 的日志.

bzr help revisionspec



## Bazaar 的常用命令

bzr commit 提交代码到本地库中

bzr uncommit 从本地库中移除当前的 revision

bzr revert 将工作目录 (worktree) 代码返回到之前的某

个版本

bzr status 查看工作目录的状态信息

bzr push 将本地的代码变更推送到其他的分支库中

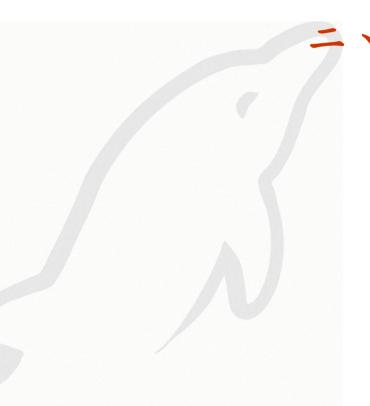
bzr merge 合并分支直接的代码变更

bzr help 查看帮助信息

bzr shelve 将正在修改的代码放到一边(架子上)

bzr unshelve 从架子上取回某个正在修改的代码





# 、编译 MySQL 源代码



## MySQL 编译脚本

#### \* MySQL 的编译脚本

MySQL 的代码中提供的一组编译脚本。开发调试时,使用这

些脚本编译更简单快捷。这些脚本在 BUILD 目录中。anders@anders-laptop:~/bzrwork/mysql-5.5\$ cd BUILD/

```
anders@anders-laptop:~/bzrwork/mysql-5.5/BUILD$ ls
autorun.sh
                                compile-amd64-valgrind-max
build mccge.sh
                                compile-darwin-mwcc
                                compile-dist
check-cpu
cleanup
                                compile-hpux11-parisc2-aCC
cmake configure.sh
                                compile-ia64-debug-max
compile-alpha
                                compile-irix-mips64-mipspro
compile-alpha-debug
                                compile-ndb-autotest
compile-amd64-debug-max
                                compile-pentiument generate
compile-amd64-debug-max-no-ndb
                                compile-pentium64
compile-amd64-gcov
                                compile-pentium64-debug
compile-amd64-gprof
                                compile-pentium64-debug-max
                                compile-pentium64-gcov
compile-amd64-max
compile-amd64-max-sci
                                compile-pentium64-aprof
```



## 编译脚本分类

#### \*不同的平台

compile-ppc-\*
compile-pentium-\*
compile-pentium64-\*

#### \*不同的功能

compile-\*-debug 包含 MySQL 的调试代码 compile-\*-max 包含所有的用户功能 compile-\*-valgrind 包含 valgrind 调试功能 有的脚本同时包含机制功能如: compile-\*-debug-max



## 手动配置编译参数

- \* 最常用的是 ./BUILD/compile-pentium64-debug-max
- \* configure

如果没有合适的方法,可以用 configure 配置后进行编译.

BUILD/autorun.sh

./configure

make

sudo make install

'configure' 命令有很多的参数可以使用,可以使用下面的命令查询。

configure --help



### 编译 MySQL 代码

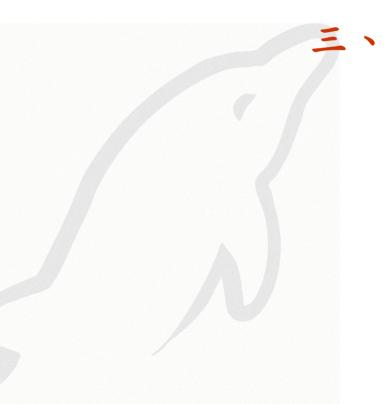
#### \*提示\*

- 如果是第一次编译 MySQL, 可能会碰到一些错误。多数是因为依赖的库和编译工具没有安装。根据错误提示进行安装就可以了。
- 这些脚本使用了-Werror等来进行严格的错误检查。因此用这些脚本可以尽早的发现 patch 中引入的代码问题。

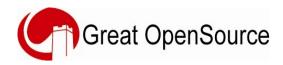
#### 对于写 patch 的同学,建议要使用编译脚本。

- 这些脚本使用了-g 参数,因此编译后代码可以很方便的进行调试。





# 、运行 MySQL



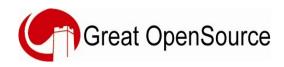
## MySQL 的测试平台 mtr

#### \* mtr

```
mysq1_code_root/mysq1-test/mtr
```

组成: perl 测试框架 + mysqltest + 测试用例

```
anders@anders-laptop:~/bzrwork/mysql-5.5/mysql-test$ ls
                    include
                                         mysql-test-run.pl t
aaa.txt
                                         purify.supp
                    lib
                                                           test.sh
a.ann
CMakeFiles
               Makefile
                                                           test.sh~
cmake install.cmake Makefile.in
                                                           valgrind.supp
                                         README
CMakeLists.txt
                                         README.qcov
                mtr
                                                           var
collections
                mtr.out-of-source
                                         README.stress
CTestTestfile.cmake mysql-stress-test.pl std data
                    mysql-test-run
                                         suite
extra
```



## mtr 的特点

#### \*简单

无需安装、无需配置、自动运行多实例 不改变系统环境、多版本的实例运行互不干扰

#### \*功能丰富

支持各种各样的 MySQL Server 配置 支持运行 MySQL 客户端程序和 shell 命令 支持 gdb, ddd, valgrind 压力测试 mysql-stress-test.pl 简单改造后,可以测试 mysql的中间件

#### 用mtr 运行 MySQL Server

#### \*启动 MySQL 单实例

```
cd mysq1-test
echo "s1eep 1000000;" > t/my_run.test
./mtr my_run —testcase-timeout=1000000
```

```
TEST RESULT TIME (ms) or COMMENT

worker[1] Using MTR_BUILD_THREAD 300, with reserved ports 13000.513009the output
```

使用的端口从13000开始,一个实例占一个端口。

#### 用mtr 运行 MySQL Server

#### \*启动 MySQL 双实例复制

```
cd mysq1-testecho "source include/master-slave.inc;" > suite/rp1/t/my_run.testecho "sleep 10000000;" >> suite/rp1/t/my_run.test./mtr my_run —testcase-timeout=1000000Master -> mysq1d.1 第一个端口Slave -> mysq1d.2 第二个端口
```

## 设置 MySQL 的参数

\* mtr 参数 'mysqld'

```
./mtr my_run.test -mysqld='--innodb-file-per-table'
```

\*为测试用例添加 option 文件

```
echo "--innodb-file-per-table" >t/my_run-master.opt
echo "--innodb-file-per-table" >t/my_run-slave.opt
```

\*为测试用例添加 cnf 文件

```
t/my_run.cnf
```

启动多实例,就是在配置文件中配置多个mysq1d

### 连接到 MySQL Server

#### \* TCP 连接

- ../client/mysql -u root -protocol=tcp -P13000
- ../client/mysql -u root —protocol=tcp -P13001

#### \* Unix Domain Socket

mtr 的 socket 文件在 var/tmp 目录中

- ../client/mysql -u root -S var/tmp/mysqld.l.sock
- ../client/mysql -u root -S var/tmp/mysqld.2.sock

## MySQL 的数据和日志

\* var 目录

mysq1\_root/mysq1\_test/var 数据和临时文件等所在目录

\*数据文件

数据 var/mysqld.\*/data/

\*日志文件

错误日志、调试日志 var/log/ General,Slow log var/mysqld.\*/

#### 写测试用例

```
* SQL 语句
 CREATE TABLE t1(c1 INT);
  --CREATE TABLE t1(c1 INT)
 delimiter /
* 引用其他的测试文件
  --source include/have binlog format row.inc
  --source extra/rp1 tests/rp1 innodb.test
*切换 mysqltest 和 mysql server 的连接
  --connection default
  --connection master
  --connection slave
```

### 写测试用例

#### \*复制同步

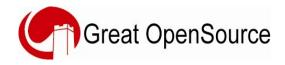
```
--sync_slave_with_master #on master connection
--save_master_pos # on master
--sync_with_master # on slave
```

#### \*创建更多的连接

```
--connect(conn_name, localhost, root,,)
--connect(conn_name, 127.0.0.1, root,, test, $SLAVE_MYPORT)
```

#### \*参考

```
mysqltest 手册
https://dev.mysql.com/doc/mysqltest/2.0/en/index.html
其他的测试用例 t/ suite/rpl suite/binlog suite/innodb
```



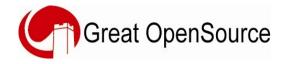
## 其他常用的功能

```
启用 binlog --source include/have_log_bin.inc
启用 innodb --source include/have_innodb.inc
使用指定的 --source include/have_binlog_format_*.inc
复制模式 *→ statement, row, mixed
重启 Server
          --source include/restart mysqld.inc
           --source include/rpl start server.inc
           --source include/rpl_stop_server.inc
重启 Slave
           --source include/rpl start slaves.inc
           --source include/rpl_stop_slaves.inc
检查主备数
           --source include/rpl_diff.inc
据的一致性
执行外部命
           --exec $MYSQL test -e "select * from t1\g"
令
           --exec $MYSQL DUMP --compact test t1
           --exec $MYSQL_BINLOG —short-form ···
           --exec <shell command>
期望返回特
           --error <error no>
定的错误
```

北尔刀里开源拟件有限公司



# 四、调试 MySQL 源代码



# gdb 调试

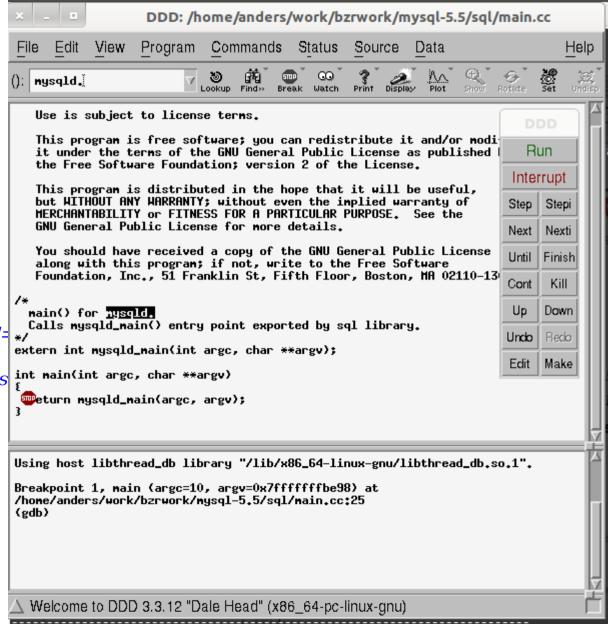
- \* gdb
  - ./mtr my\_test --gdb
- \* ddd
  - ./mtr my test —ddd
- \*单线程运行 MySQL

./mtr my\_test —ddd --mysq1d= \*/

'—thread-handling=no-threads

\*禁用编译优化

优化后的执行代码和源 代码不匹配,妨碍调试





## MySQL 的 debug 信息

#### \* trace dump

DBUG\_ASSERT

```
./mtr my_test --debug<sup>T@3</sup>
var/log/mysqld.*.trac
dbug package 一个升源
的代码追踪和调试工具
DBUG_ENTER,
DBUG RETURN,
DBUG VOID RETURN
DBUG PRINT,
DBUG DUMP,
```

```
>build bitmap for nested joins
                       <build_bitmap_for_nested_joins_penSource</pre>
T@3
                       >optimize_cond
                       <optimize cond</pre>
                       >optimize cond
                       <optimize cond</pre>
                       info: No tables
                     <JOIN::optimize</pre>
                                                * trace dump
                     >JOIN::exec
                       proc info: /home/anders/work/bzrwork/mysql-5.5/sql/sql_se
T@3
T@3
                       >PROFILING::status_change
                       <PROFILING::status_change
T@3
                                                   生成 var/log/mysgld.*.
T@3
                       >send result set metadata
T@3
                         packet_header: Memory: 0x7fc024912060 Bytes: (4)
01 00 00 01
                         >alloc root
T@3
                                                  DBUG ENTER, DBUG
T@3
                           enter: root: 0x3400358
T@3
T@3
                           exit: ptr: 0x7fc01c0068d0
                                                   DBUG PRINT, DBUG 1
                         <alloc root
T@3
                         >Protocol::write
                         <Protocol::write
                         packet_header: Memory: 0x7fc024912010 Bytes: (4)
33 00 00 02
                         packet_header: Memory: 0x7fc024912010 Bytes: (4)
05 00 00 03
                       <send result set metadata</pre>
T@3
                       >select_send::send_data
T@3
                         >Protocol::send_result_set_row
T@3
                         <Protocol::send result set row
T@3
                       <select_send::send_data
                       >Protocol::write
```

### MySQL 的 debug point

```
* MySQL 调试宏

DBUG_EXECUTE

DBUG_EXECUTE_IF

添加调试代码,通过 MySQL 变量控制调试代码何时执行
```

- \* 代码中添加调试点

  DBUG\_EXECUTE\_IF( "dbug\_point\_name", sleep(5););
- \*启动执行调试点代码 SET SESSION debug="d,dbug point name";
- \*提示: dbug point 大多都有测试用例 代码中搜索 DBUG\_EXECUTE\_IF 测试用例中搜索代码中定义的 <dbug point name>

## MySQL 的 debug 信息

#### \*输出信息到 error 日志

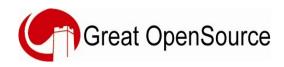
```
sql_print_error
sql_print_warning
sql_print_information
```

\* strace-server

```
./mtr my_test —strace-server
var/log/mysqld.*.strace
使用 strace 命令来跟踪 MySQL 的系统调用和 Signal
```



# 五、阅读 MySQL 源代码



# 主要的代码目录

| 目录         | 功能描述                    |
|------------|-------------------------|
| sq1        | Server 的实现,和 SQL 解析优化执行 |
| storage    | 各种引擎代码                  |
| client     | 各种客户端工具代码               |
| plugin     | 各种插件的代码                 |
| mysys      | 对系统调用的封装和一些共用算法         |
| vio        | 对各种网络调用的封装              |
| sq1-common | server和 client 共用的代码    |
| 1ibmysq1   | 客户端 API                 |
| 1ibmysq1d  | 服务器端 API                |

最频繁查看的代码在 sql, storage/innobase, client

## 从哪里开始阅读代码

\* sq1/mysq1d.cc

main()

启动、配置文件、系统变量、网络、引擎、复制等各种功能初始化

\* sq1/sq1\_parse.cc

dispatch\_command()

各种客户端命令的处理(包括 SQL), COM\_QUERY, COM\_XXX

mysq1\_execute\_command()

各种 SQL 语句的执行, SQL COM\_XXX

从这两个文件开始,几乎总能找到你想找的代码。

## 从哪里开始阅读代码

\* storage/handler/ha\_innodb.cc

InnoDB 对 MySQL 引擎 API 的实现.每个引擎都有自己的实现,主要代码的入口都在这里。

\* sq1/log\_event.cc

XXX\_log\_event::do\_apply\_event()

各种 binary log 在 slave 上的执行逻辑

\* sq1/slave.cc

handle\_slave\_io() IO 线程的处理逻辑

handle\_slave\_sq1() SQL 线程的处理逻辑



## 从哪里开始阅读代码

\* sq1/sq1\_yacc.yy SQL 语句解析的代码

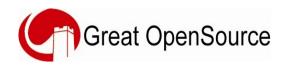




\* THD(sq1/sq1\_c1ass.h)

每个 session 有一个 THD, THD 中包含了这个线程特有的所有上下文信息。

| 变量名         | 描述   |
|-------------|--|
| 1ex         | 解析后的 SQL 语句结构  |
| query       | 当前连接的 SQL 语句   |
| mem_root    | 线程的内存池,处理语句时不需要从系统中分配内存,从线程的内存池中分配。语句结束时释放回池中。               |
| variables   | SESSION 变量。 THD 之外,另有 global_system_variables 存储 GLOBAL 变量 . |
| host, user  | 当前连接的用户信息  |
| transaction | 事务相关的信息  |
|             |  |



\* TABLE(sq1/sq1\_table.h)

一个数据表可以有多个 TABLE 结构。每个线程在打开一个表时, 会创建一个 TABLE, TABLE 不在线程间共享。

| 变量名                   | 描述   |
|-----------------------|--|
| S                     | TABLE_SHARE,每个数据表只有一个 TABLE_SHARE,在所有的 TABLE 之间共享.包含一些元数据信息。   |
| file                  | 引擎的 handler 实例,和引擎的交互通过 file 来完成。  |
| in_use                | TABLE 所在的 THD  |
| field                 | 表的字段信息   |
| record[2]             | 临时存储当前行的数据。在进行数据读写时,引擎需要传入或传出行数据. select, insert, delete 只需要其中一个. Update 时需要 2 个, record[1] 指向修改前的数据, record[0]指向修改后的数据。 |
| read_set<br>write_set | 在引擎传入或传出一条记录时,并不一定是完整的一行数据。当某些字段不需要时,可以通过 read_set, write_set 告诉引擎。它们维护了一个字段是否需要(出现)的标志位。                                |



\* SELECT\_LEX(sq1/sq1\_lex.h)

解析后一个 SELECT 语句的结构信息.如果一个 SQL 中有多个 SELECT,需要用多个 SELECT\_LEX 来表示.多层次的 SELECT\_LEX 通过 SELECT\_LEX\_UNIT 来组织到一起.

| 变量名                               | 描述                                  |
|-----------------------------------|-------------------------------------|
| table_list                        | SELECT 中的用到的表, CREATE TABLE 时,也放在其中 |
| where, having                     | WHERE 和 HAVING 中的内容                 |
| <pre>group_list, order_list</pre> | GROUP BY, ORDER BY 中的内容             |
| item_1ist                         | SELECT 列表中的内容                       |
| join, join_list                   | Join 信息和 join 的表                    |

sq1\_1ex.h 中有详细的描述,搜索" The state of the 1ex parsing for selects"



\* LEX(sq1/sq1\_1ex.h)

解析后的 SQL 语句信息,每个语句解析后生成1个 LEX 对象,记录在THD::1ex中.

| 变量名  | 描述  |
|--|---|
| unit select_lex current_select all_selects_lists | SELECT_LEX_UNIT, SELECT_LEX. 这 2 个结构的变量用来表达多层次的 SELECT 语句。 Unit 是最外层的单元.select_lex 是语句中的第一个 select_lex. |
| create_info<br>alter_info                        | CREATE TABLE, ALTER TABLE 的信息   |

MySQL 中用 Item 代表 SQL 语句中的一个子单元.

Field、表达式、条件、函数、WHERE、HAVING、Limit都是表达式



#### \* 选择一个适合的阅读工具

```
vim + ctags,cscope
emacs + cscope
```

- 可以查找函数的定义和调用
- 可以查找符号(变量名,宏定义,结构体等)的定义和引用
- 可以列出文件中的函数、符号表
- 可以列出一个函数的函数调用列表

#### \*提示\*

将.yy和.ic文件加入源码文件中

.yy -> sql\_yacc.yy 语法解析文件

.ic -> innodb 的 include 文件



#### \*利用 bzr 来获取信息

patch message, bugs.mysq1.com,
https://forge.mysq1.com/worklog

#### \* 通过错误信息查找代码

sq1/share/errmsg-utf8.txt 包含所有错误信息

通过错误信息,找到错误标签,然后通过错误标签找到代码。

例:"PROCEDURE spl can't return a result set in the given context"

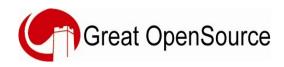
errmsg-utf8.txt → ER\_SP\_BADSELECT

ER\_SP\_BADSELECT → sq1\_parse.cc::SQLCOM\_CALL 和 MULTI\_RESULT 相

关



- \*快读代码,跳过不相干的代码
- 注意宏定义, debug 的代码可以直接跳过。
- 注意宏定义,某些代码只属于特定的功能,如果和问题无关可以直接 跳过。如,
  - WITH\_PARTITION\_STORAGE\_ENGINE, HAVE\_REPLICATION
- 跨平台的代码可以忽略掉。MySQL代码中对很多系统函数 做了封装,功能没有任何变化,只是为了更好的跨平台。 如:my\_malloc, my\_free, my\_write, my\_read等. 只需 要理解其功能,不用关注实现细节。



\* 通过客户端程序的代码来查找 server 的代码

客户端程序一般都比较简单,更容易理解.在有些时候可以帮助更快的定位代码。

接上页的例子, MySQL 如何发送多结果集给客户端?

从mysq1 命令一定会处理这种情况。

mysql\_next\_result() -->

server\_status → SERVER\_MORE\_RESULTS\_EXISTS

mysq1 → mysq1 通讯相关的代码

mysqlbinlog → binary log 相关的代码



## \* 通过阅读 patch 来理解代码

http://lists.mysql.com/commits

可以选择一些自己关注的代码和功能的 patch 来阅读。比如想了解, MySQL 变量相关的代码,找某个变量实现的 patch 来看。整个模块的代码就很清楚了。

对于希望从事 mysql server 代码开发的人,可以选择自己关注的 bug 或 feature,自己现实现一个,然后和官方的 patch 对比。



# Q&A



# 谢谢!