

版权声明

本次分享来源自中国MySQL用户组（ACMUG）

南京站活动，版权归演讲嘉宾所有。

关注ACMUG公众号，参与社区活动，
交流开源技术，分享学习心得，一起共同进步。



公众号ID: A_CMUG

AliSQL Open Source Intro

Lixun Peng
Staff Database Engineer

中国MySQL用户组

Agenda



- p AliSQL Introduction
- p Background Bussiness Requirement
- p Important Features

中国MySQL用户组

p Background

MySQL branch originated from Alibaba Group, based on Oracle MySQL

p Philosophy

From the open source community, we get more

For the open source community, we hope to contribute more

p Open Source

Opened on August 2016

Github Address: <https://github.com/alibaba/AliSQL>

p Released

Five Releases (34 Features, 8 Performance Improvement)

AliSQL Releases

- [Changes-in-AliSQL-5.6.32-\(2017-02-14\)](#)
- [Changes-in-AliSQL-5.6.32-\(2016-12-25\)](#)
- [Changes-in-AliSQL-5.6.32-\(2016-11-11\)](#)
- [Changes-in-AliSQL-5.6.32-\(2016-10-14\)](#)
- [Changes-in-AliSQL-5.6.32-\(2016-09-15\)](#)

p Related Tools

AliSQLBackup Integrated with TokuDB Engine based on Xtrabackup

Github Address: <https://github.com/alibaba/AliSQLBackup>

Background Bussiness Requirement

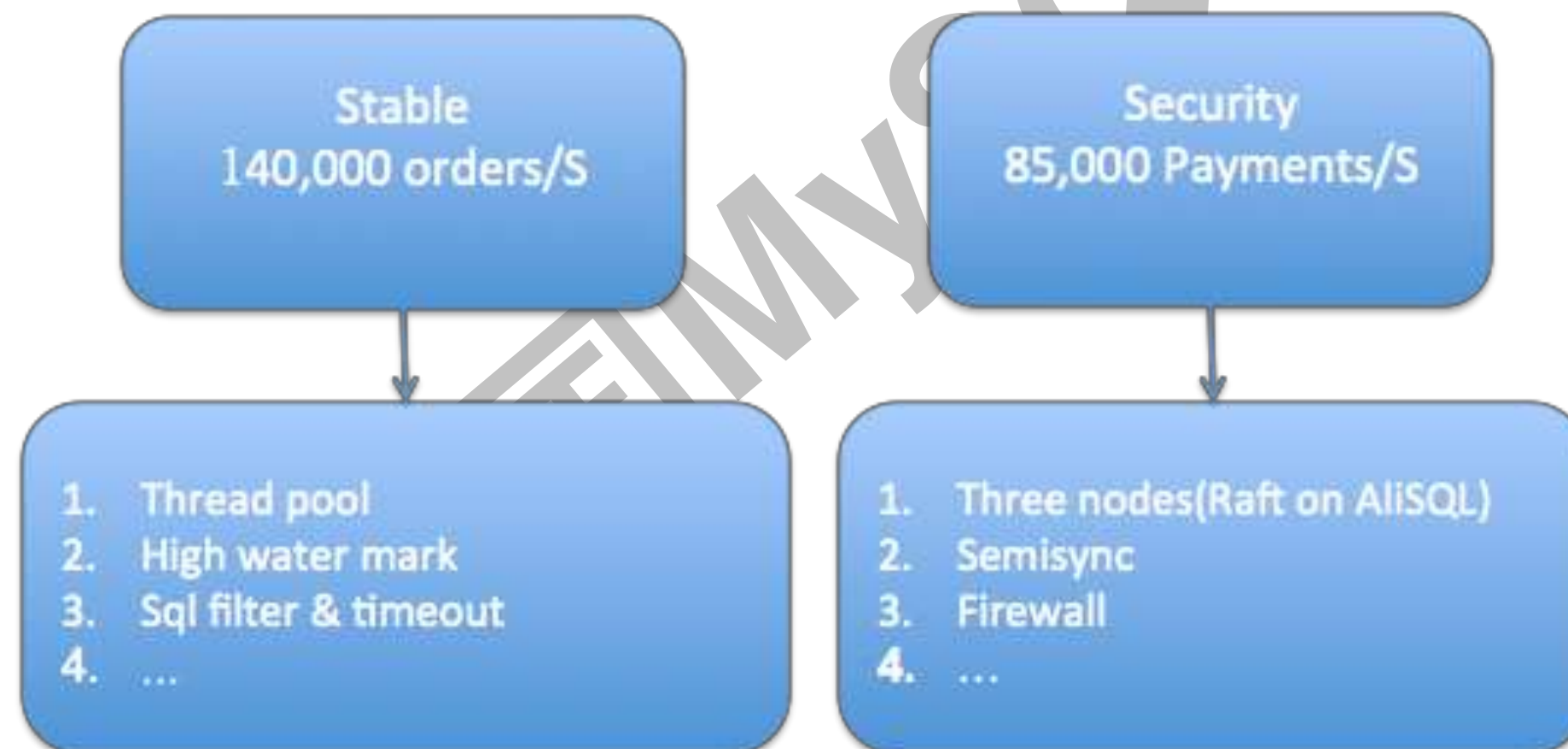


p Mainly two scenes

– Electronic Commerce

More stable for 11.11 sales activity

More strict on data protection for finance



Background Bussiness Requirement

p Mainly two scenes

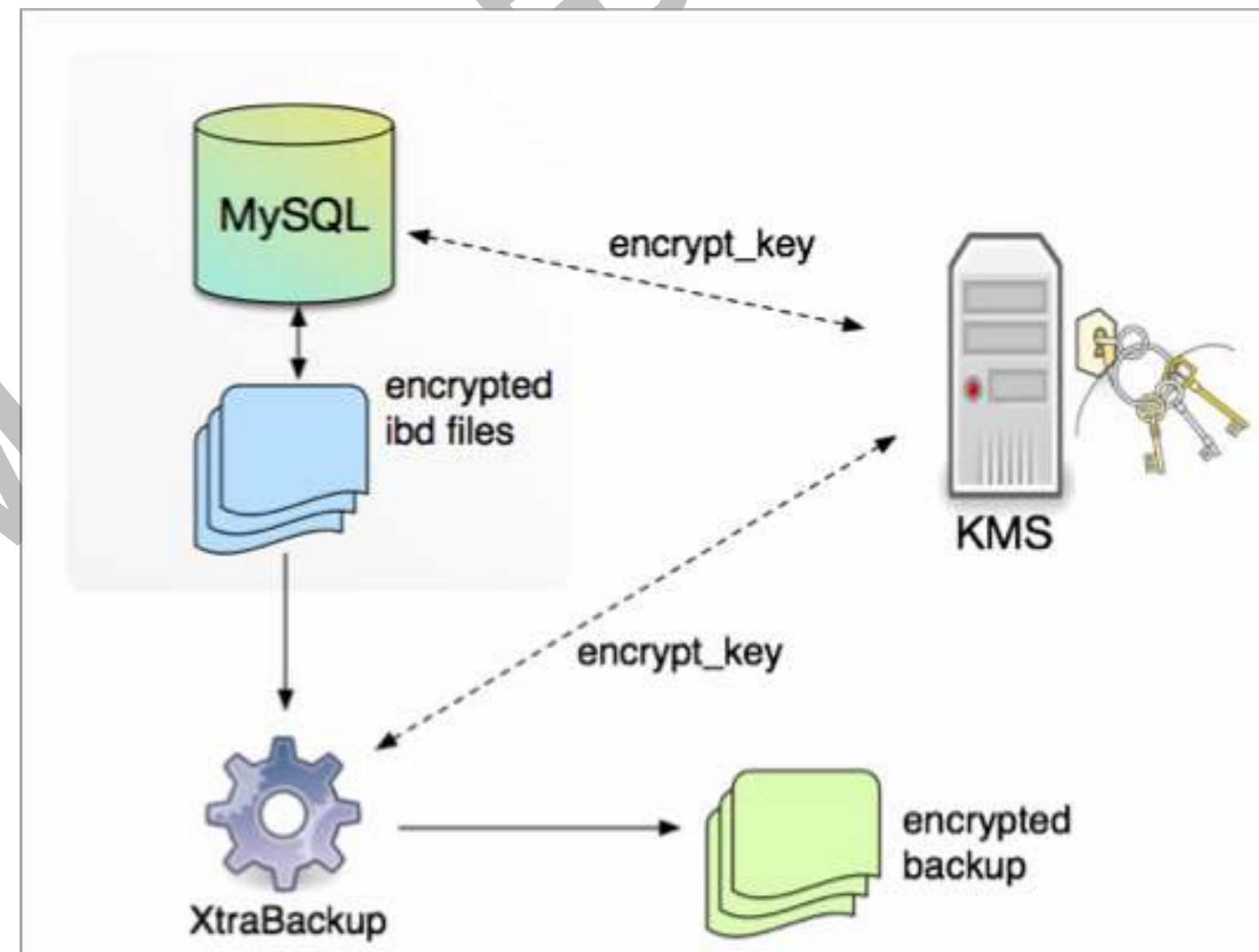
p Cloud Environment

High availability on instance

Absolute safety on data

Transparent Data Encrypt

Full page encrypted only once



Background Bussiness Requirement - stability



p Stability

– Scenarios

100 thousand concurrent update on the same row which is hot commodity high overhead on deadlock

lower and lower throughput, until unavailable

– Solution

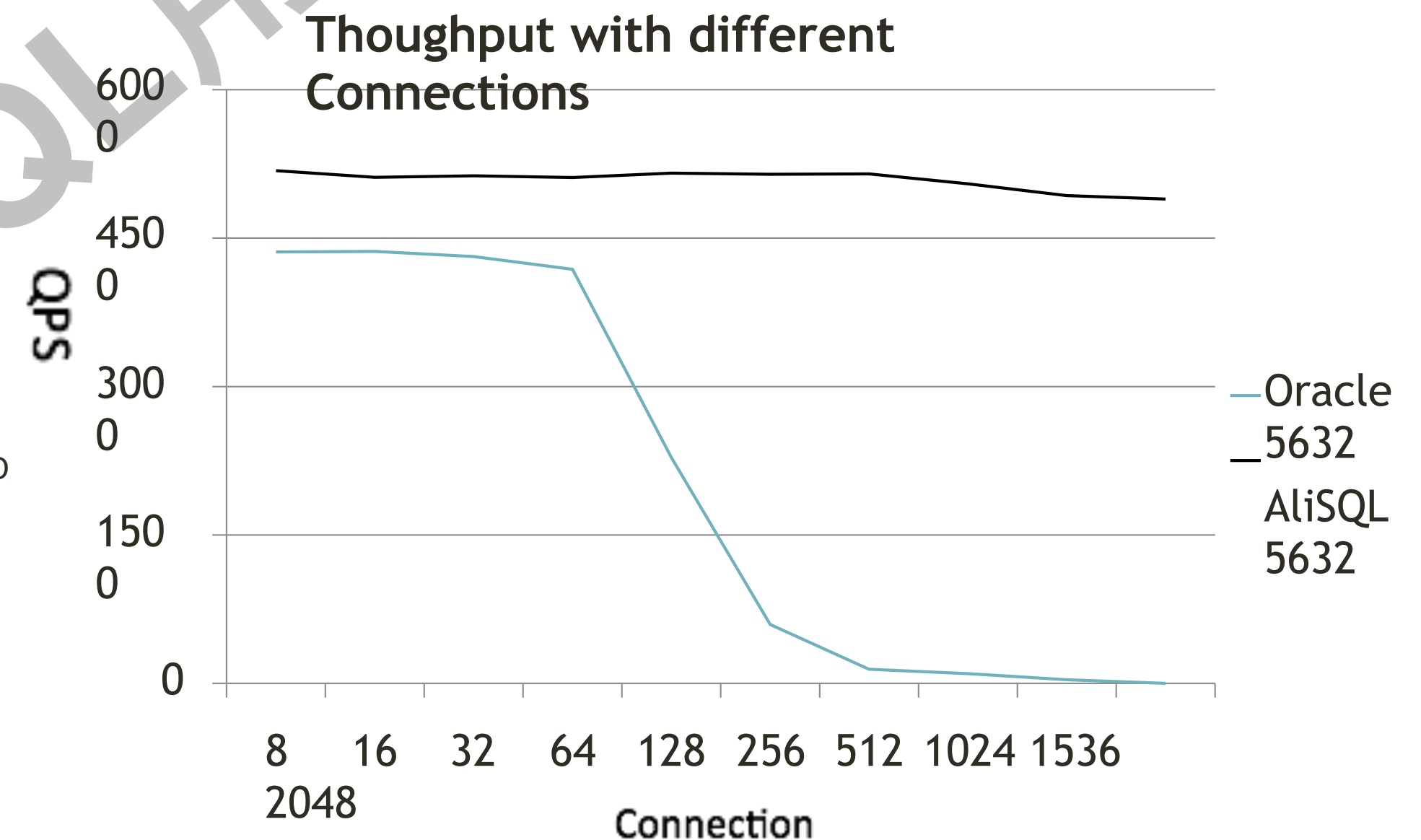
Queue by commodity ID

Server layer queue by HINT:

```
COMMIT_ON_SUCCESS ROLLBACK_ON_FAIL QUEUE_ON_PK $ID TARGET_AFFECT_ROW $ID
```

InnoDB layer queue by innodb_thread_concurrency :

innodb_thread_concurrency degenerated to transaction

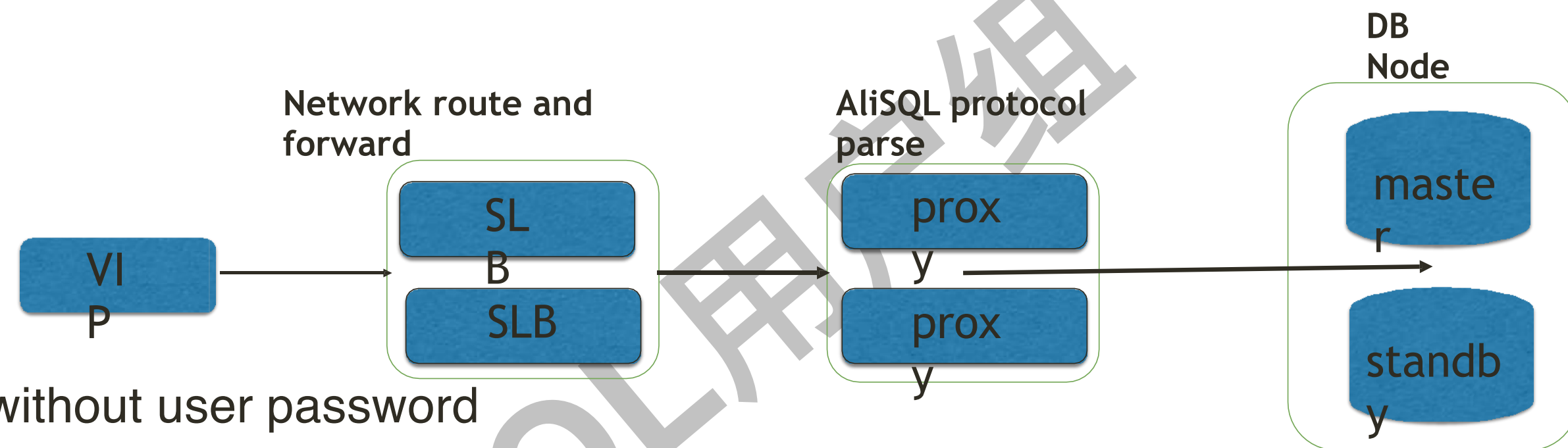


Background Bussiness Requirement - availability



p Availability

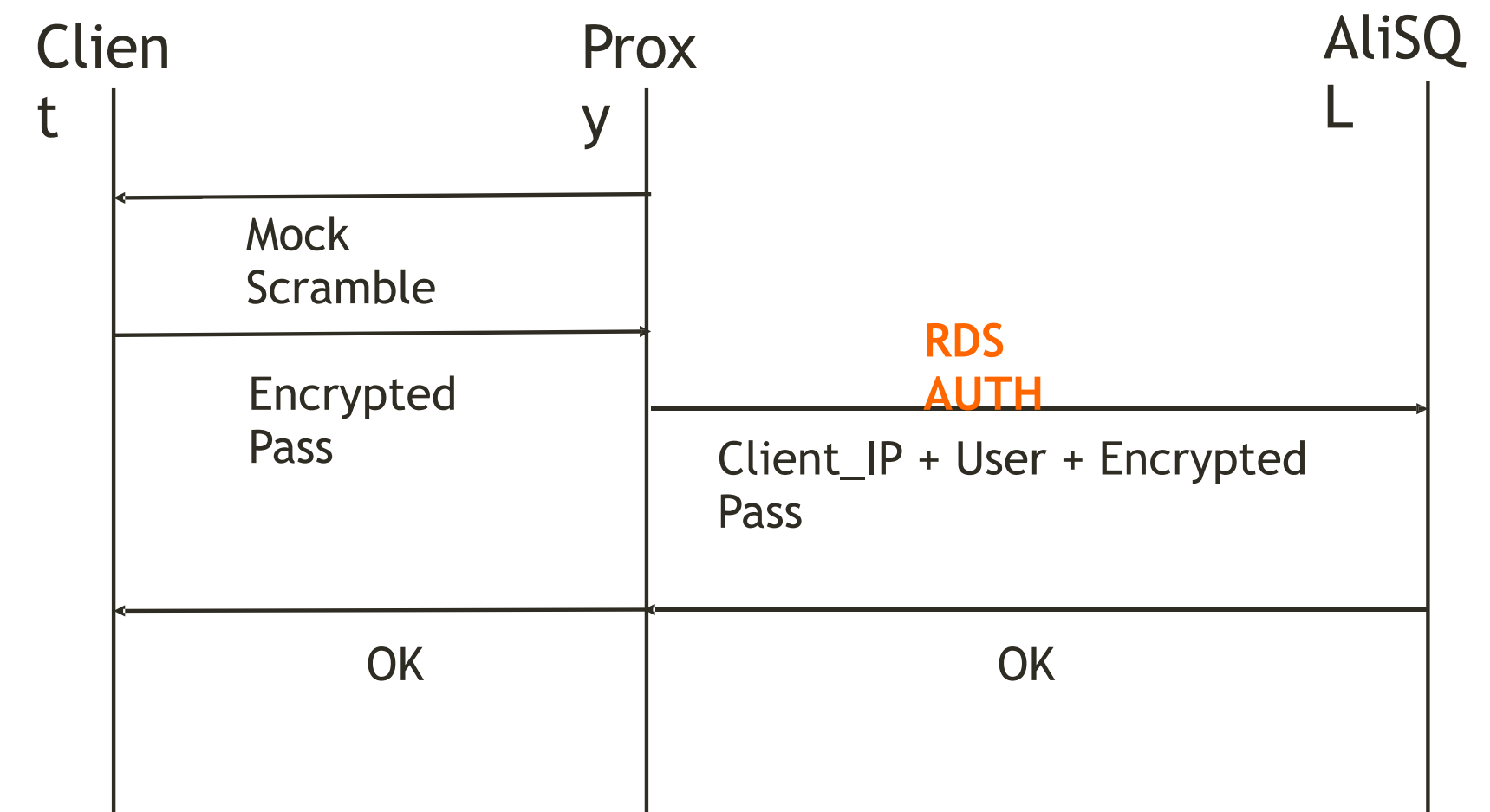
– RDS Architecture



1. Proxy cannot connect instance without user password
2. Connection will break if DB switchover

– Open privileges

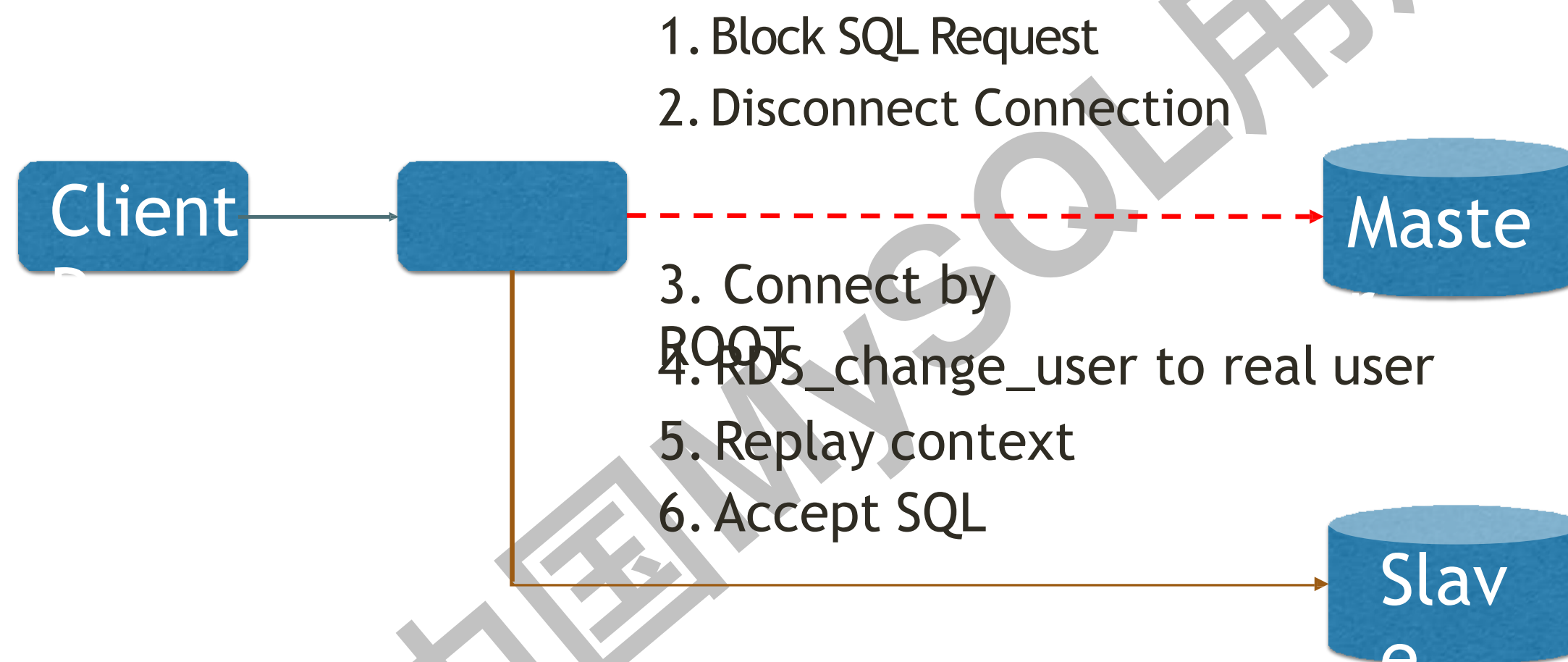
Proxy bridge connection with new AUTH protocol



Background Bussiness Requirement- availability

– Keep Connection

Proxy keep connection when switch over



Replay the connection context(DB, Charset, PS.....)

Wait for some time if within transaction

Important Features



- p Contribute to MariaDB
- p Port from MariaDB
- p Joint Development

中国MySQL用户组

Important Features - Contribute to MariaDB



p Multi-Source Replication

– Design

Independent IO/SQL thread, relay_log for every channel

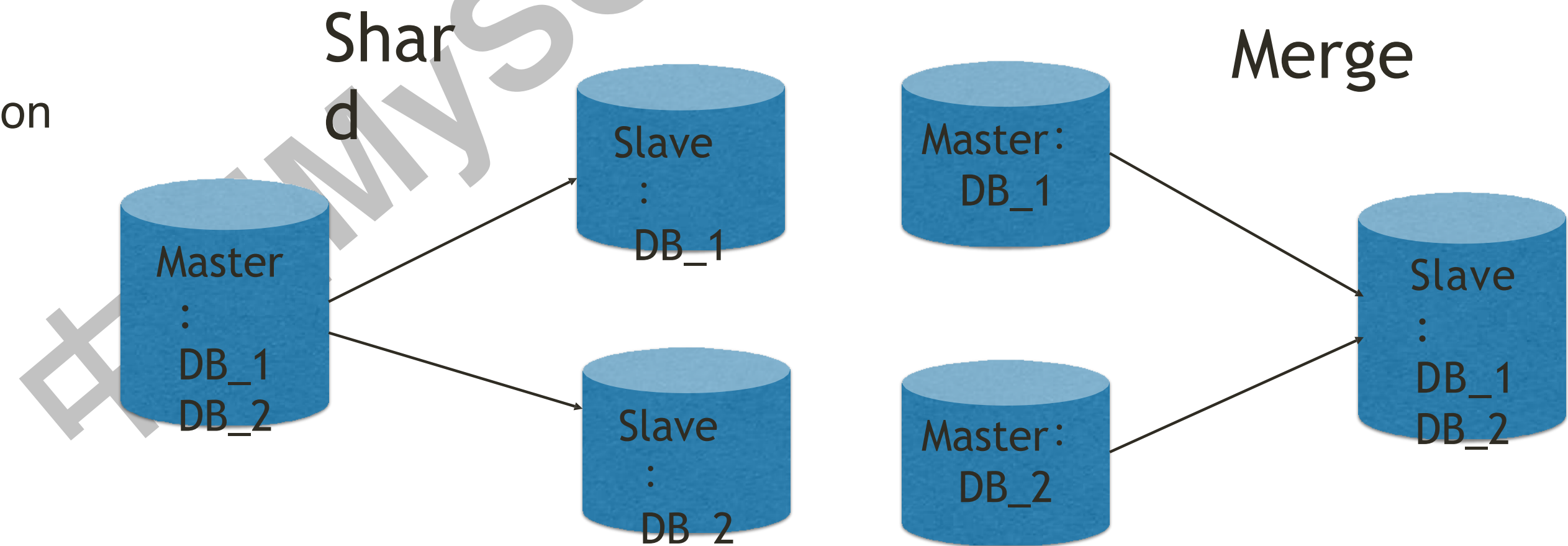
Without guarantee data conflict

All slave-related commands add 'for channel="channel_name"'

– Usage

Elastic capacity expansion

Elastic capacity shrink



Important Features - Contribute to MariaDB



p Flashback

- MySQL Binlog
 - Total order by transaction
 - Row event including pre-image, post-image
- Reverse Row Event
 - DML (INSERT+ UPDATE+ DELETE)
 - Unsupport DDL (Recyclebin instead)
- Usage
 - mysqlbinlog -B --start_positioin=xx --stop_posit

```
### INSERT INTO `test`.`xp`
### SET
###   @1=1 /* INT meta=0 nullable=0 is_null=0 */
###   @2='pre' /* VARSTRING(100) meta=100 nullable=1 is_null=0 */

### UPDATE `test`.`xp`
### WHERE
###   @1=1 /* INT meta=0 nullable=0 is_null=0 */
###   @2='pre' /* VARSTRING(100) meta=100 nullable=1 is_null=0 */
### SET
###   @1=1 /* INT meta=0 nullable=0 is_null=0 */
###   @2='post' /* VARSTRING(100) meta=100 nullable=1 is_null=0 */

### DELETE FROM `test`.`xp`
### WHERE
###   @1=1 /* INT meta=0 nullable=0 is_null=0 */
###   @2='post' /* VARSTRING(100) meta=100 nullable=1 is_null=0 */
```

Important Features - Contribute to MariaDB



p Auto Increment Persistence

– Design

InnoDB auto_increment value can be persisted into IBD data file. It will occupy the Clustered Index root page PAGE_MAX_TRX_ID. PAGE_MAX_TRX_ID makes sense only within secondary index.

– Parameters

INNODB_AUTOINC_PERSISTENT

This persistent behaviour is enabled or not

INNODB_AUTOINC_PERSISTENT_INTERVAL

The interval of persist max auto_increment value

Important Features - Port from MariaDB



p Max Statement Time

Kill statement that takes over the specified number of milliseconds

It is very useful to keep stability

– Usage

SET GLOBAL/SESSION max_statement_time= constant;

Hint: /* max_statement_time= constant */

– Expansion

Max Transaction Time:

IBDATA file extend Horribly if transaction didn't commit for long time

DDL Fast Fail

MDL will block user query if DDL blocked

Important Features - Port from MariaDB



p InnoDB Defragmentation

new ways to free the unusable space that didn't need to create new table and move records.

– Usage

```
SET GLOBAL INNODB_DEFRACTIONATION= 1;  
OPTIMIZE TABLE table_name;
```

– Expansion

Tools to shrink IBDATA undo space.

中国MySQL用户组

Important Features - Joint Development



p Sequence Engine

– Syntax

```
CREATE SEQUENCE [IF NOT EXISTS] schema.sequence_name  
  [START WITH <constant>]  
  [MINVALUE <constant>]  
  [MAXVALUE <constant>]  
  [INCREMENT BY <constant>]  
  [CACHE <constant> | NOCACHE]  
  [CYCLE | NOCYCLE]  
;
```

```
SELECT nextval For schema.sequence_name;  
SHOW CREATE TABLE/SEQUENCE schema.sequence_name;
```

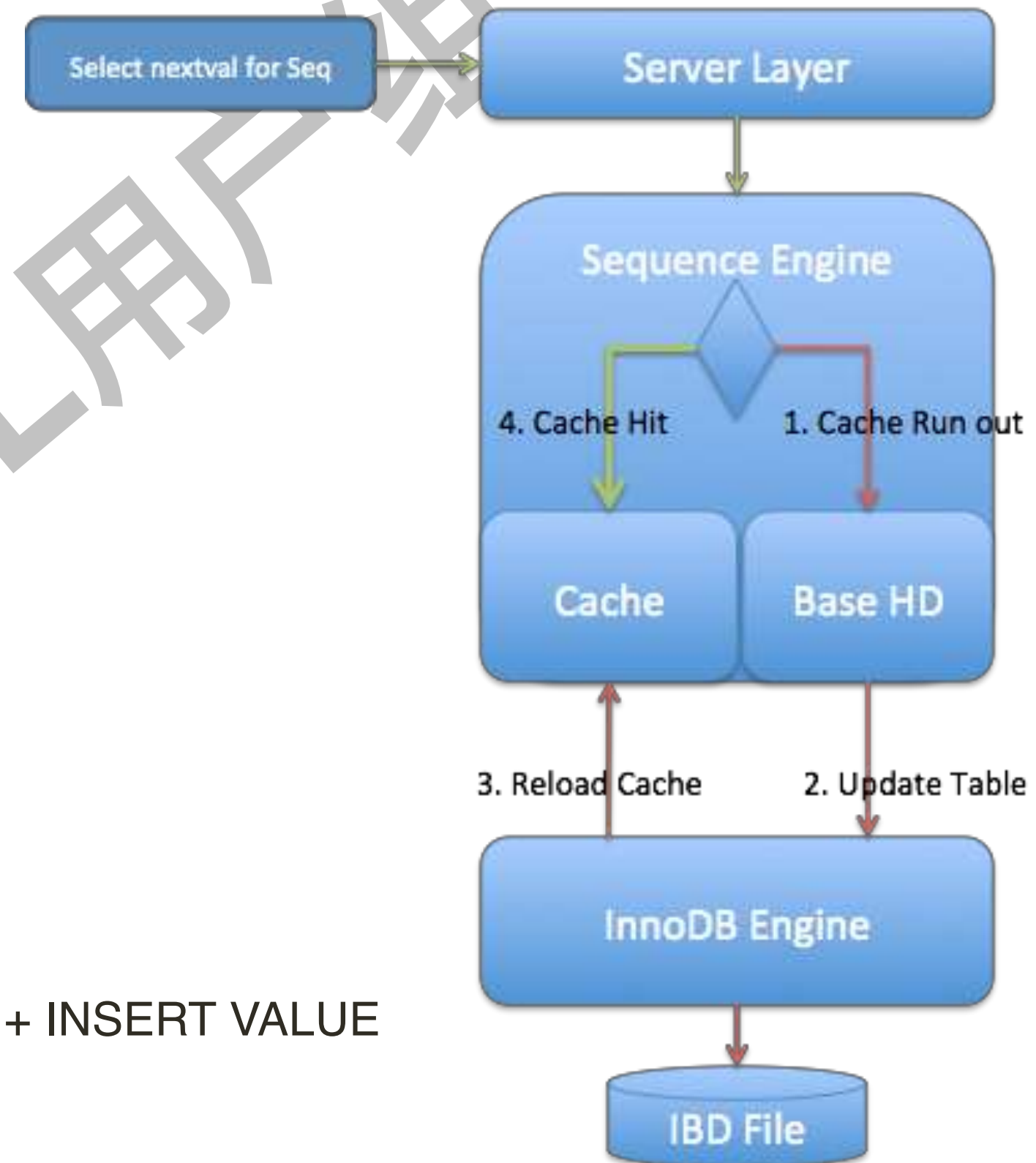
– High Architecture

Sequence is logical engine

Sequence attributes are saved on based table

Handler interface transparent pass through mostly

Logical backup will be change to CREATE SEQUENCE TABLE + INSERT VALUE



Important Features - Joint Development



- Sequence initialization
 - Create Sequence

CREATE SEQUENCE will be changed into:
[Table Creation + Record Insert]

```
create sequence test.seq_t1
  start with 1
  minvalue 1
  maxvalue 9999999
  increment by 1
  cache 20
  cycle;
```

```
CREATE SEQUENCE schema.sequence_name (
  `currval` bigint(21) NOT NULL COMMENT 'current value',
  `nextval` bigint(21) NOT NULL COMMENT 'next value',
  `minvalue` bigint(21) NOT NULL COMMENT 'min value',
  `maxvalue` bigint(21) NOT NULL COMMENT 'max value',
  `start` bigint(21) NOT NULL COMMENT 'start value',
  `increment` bigint(21) NOT NULL COMMENT 'increment value',
  `cache` bigint(21) NOT NULL COMMENT 'cache size',
  `cycle` bigint(21) NOT NULL COMMENT 'cycle state',
  `round` bigint(21) NOT NULL COMMENT 'already how many round'
) ENGINE=InnoDB DEFAULT CHARSET=latin1

INSERT INTO schema.sequence_name VALUES(0,0,1,9223372036854775807,1,1,10000,1,0);
COMMIT;
```

Important Features - Joint Development



- Sequence Cache Management
 - Sequence update steps when cache is run out
 1. Upgrade SHARE_READ MDL to SHARE_WRITE
 2. Hold GLOBAL INTENTION_EXCLUSIVE
 3. Begin autonomous transaction(Binlog, InnoDB)
 4. Update InnoDB table, generate redo & binlog event
 5. Hold COMMIT MDL
 6. Commit autonomous transaction
 7. End autonomous transaction(restore parent ctx)
 8. Release(GLOBAL, COMMIT) MDL
 9. Release table MDL(transaction->statement)

Important Features - Joint Development



– Sequence Autonomous Transaction

Sequence NEXTVAL does not support ROLLBACK and reuse.

– Autonomous Transaction steps

1. Back up the transaction context of the current storage engines.
2. Back up the transaction context of the current BINLOG engine
3. SEQUENCE and BINLOG were registered autonomous transaction
4. When the update is complete, XA submits an autonomous transaction
5. Restore the current transaction context

Important Features - Joint Development



p Add Column Dynamically

Design a new record format ROW_FORMAT= COMFORT, allow to add column dynamically, unnecessary to copy data, only change the Dictionary.

– High architecture

- COMFORT Record Format



N_Fields: n_fields will occupy 1 or 2 bytes to save the column number.

Extra_Bytes: info_bits will occupy 1 spare bit to flag comfort.

• Limitation

The column only can added at last, and nullable, non-default value

Important Features - Joint Development



- Low Design
- ALTER steps
 1. Hold exclusive MDL
 2. Downgrade to shared MDL
 3. Allow the table DML
 4. Upgrade to exclusive MDL
 5. Update dictionary (sys_tables, sys_columns)
 6. Update dictionary cache(dict_table_t, dict_index_t)
 7. Release MDL

Important Features - Joint Development



– Low Design

• ALTER Contrast

Compact VS Comfort

```
mysql> show create table t1\G;
***** 1. row *****
      Table: t1
Create Table: CREATE TABLE `t1` (
  `id` int(11) NOT NULL AUTO_INCREMENT,
  `col1` varchar(1000) DEFAULT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=262126 DEFAULT CHARSET=latin1 ROW_FORMAT=COMPACT
1 row in set (0.00 sec)

ERROR:
No query specified

mysql> alter table t1 add col2 int;
Query OK, 0 rows affected (25.98 sec)
Records: 0 Duplicates: 0 Warnings: 0
```

```
mysql> show create table t2\G;
***** 1. row *****
      Table: t2
Create Table: CREATE TABLE `t2` (
  `id` int(11) NOT NULL AUTO_INCREMENT,
  `col1` varchar(1000) DEFAULT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=262126 DEFAULT CHARSET=latin1 ROW_FORMAT=COMFORT
1 row in set (0.00 sec)

ERROR:
No query specified

mysql> alter table t2 add col2 int;
Query OK, 0 rows affected (0.01 sec)
Records: 0 Duplicates: 0 Warnings: 0
```

Continue to Open Source



- 1. AliSQL 开源 (Code)
- <https://github.com/alibaba/AliSQL>
- <https://github.com/alibaba/AliSQLBackup>
- 2. 内核月报 (技术干货)
- <http://mysql.taobao.org/monthly/>

为了无法计算的价值 |  阿里云