SACC

卓越 5周年 变迁

2013中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2013

大数据下的IT架构变迁

Atlas技术实践

构 建高性能与高可用的MySQL中间层之路

内容提纲

- 背景
- 架构与原理
- 改进点
- 新功能
- 未来发展

应用程序员需要关注DB的细节

- 配置主库与多个从库的IP和端口
- 自己实现读写分离
- 自己实现分表

DBA运维工作繁重

- DB宕机与上下线对应用造成影响
- 协调业务修改配置文件

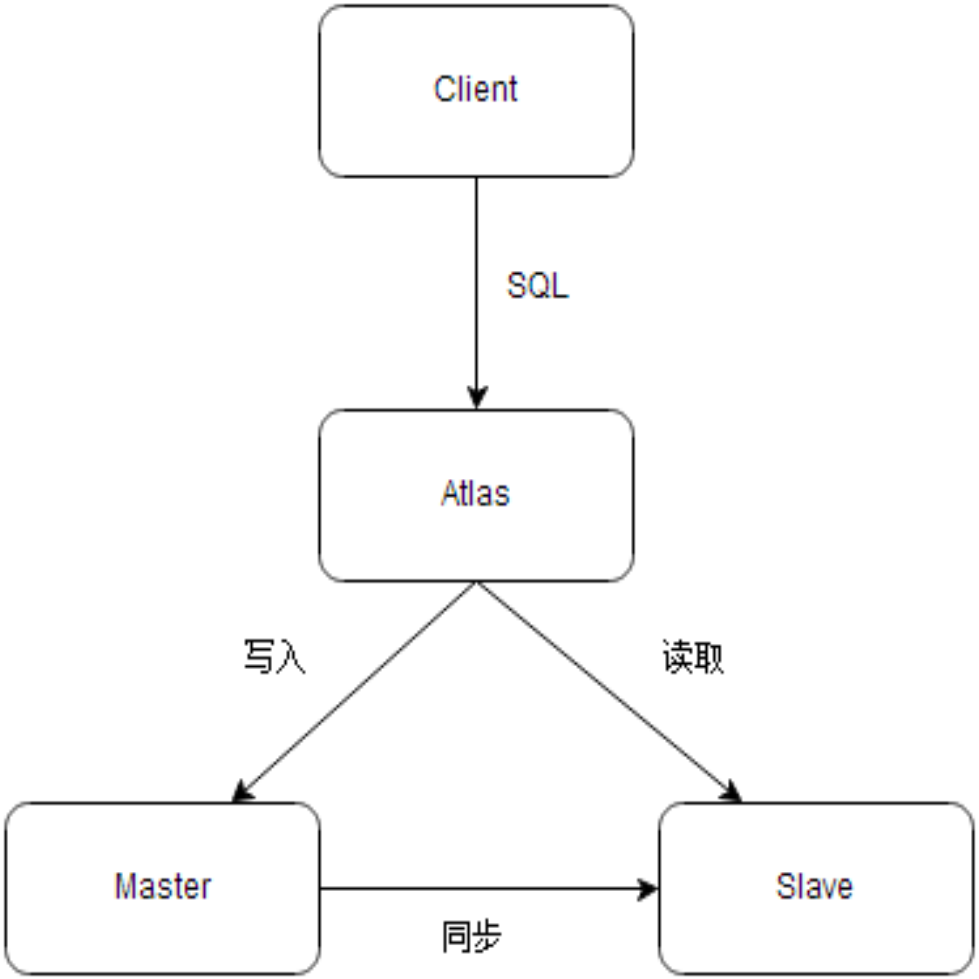
中间层的目标

- 应用与DB相对隔离
- 应用程序员可专注于编写业务逻辑
- DBA工作量降低
- 后端DB的更改对应用的影响降到最低

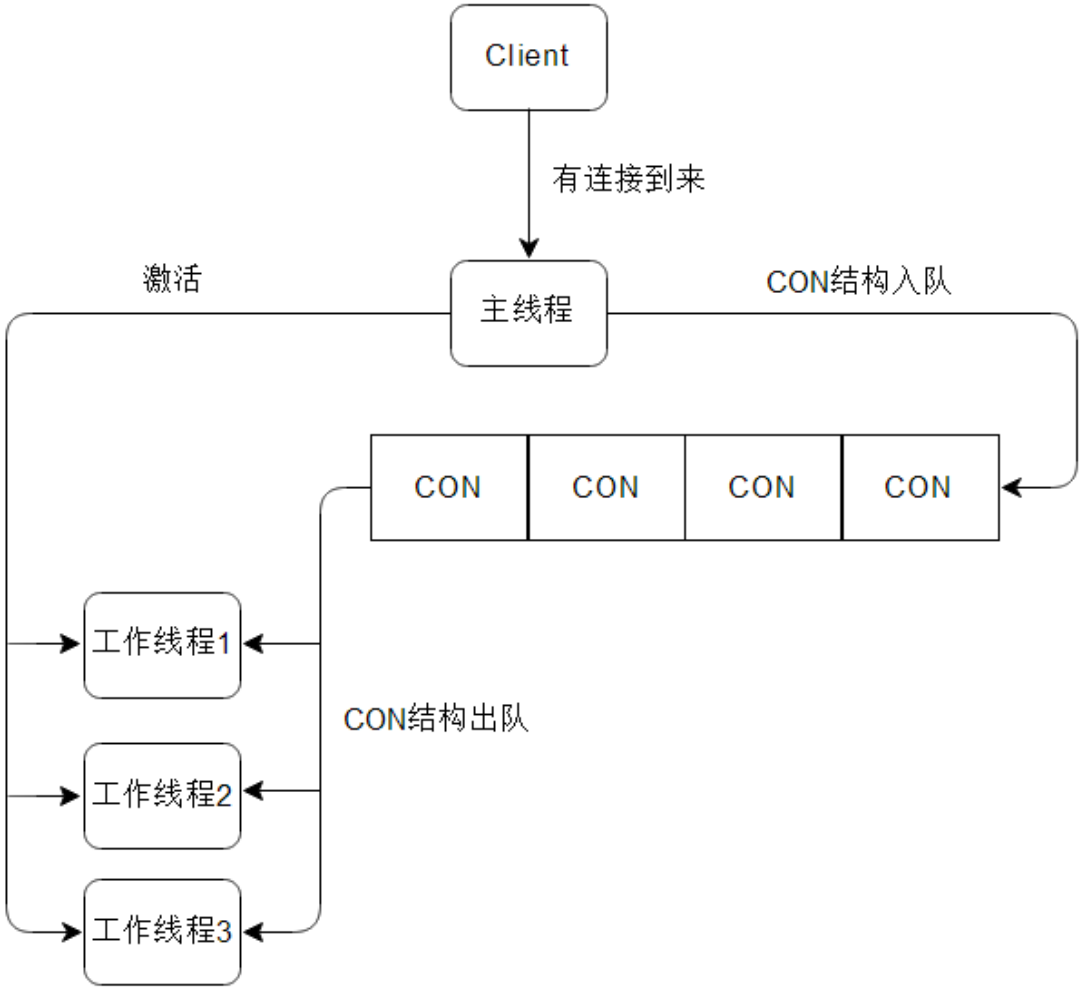
内容提纲

- 背景
- 架构与原理
- 改进点
- 新功能
- 未来发展

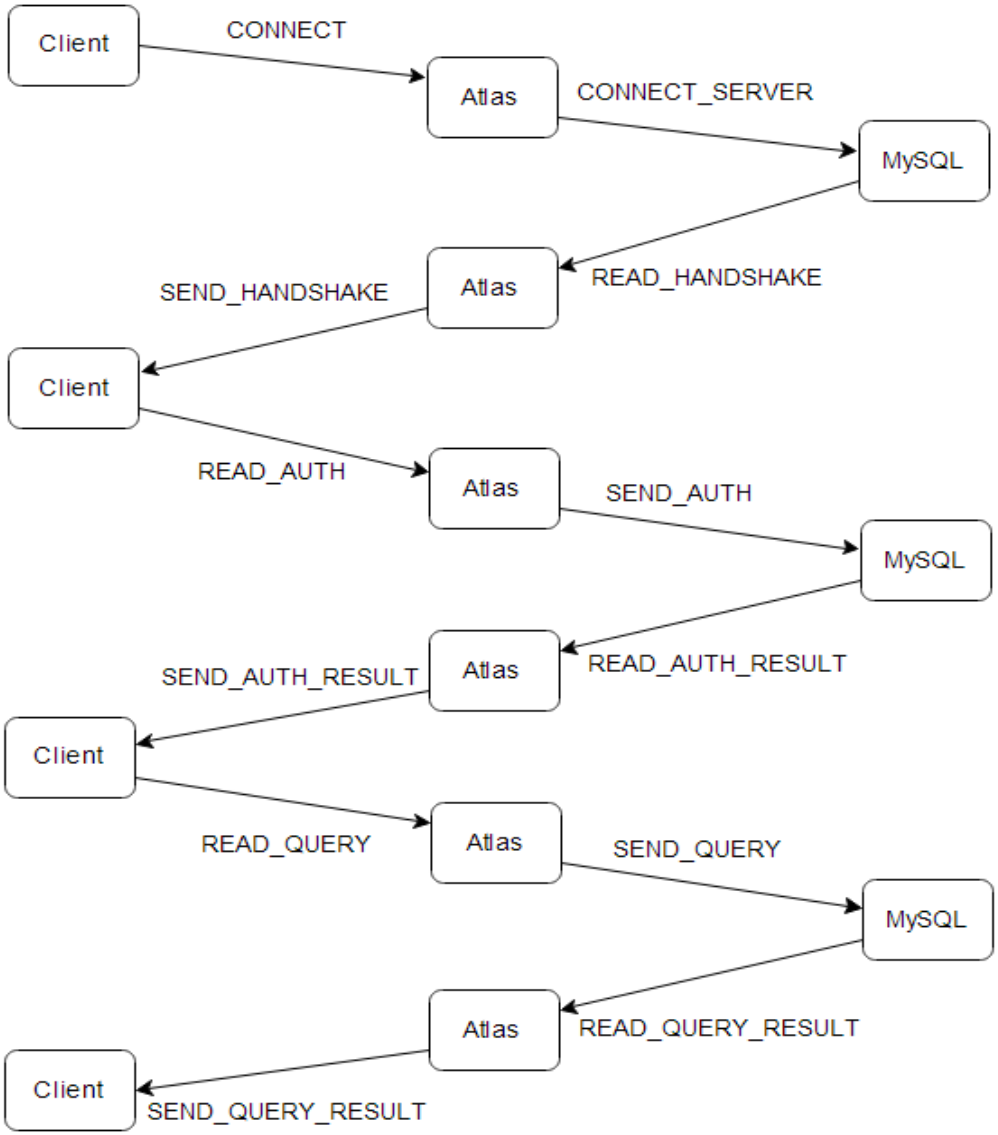
总体架构



线程模型


Client
有连接到来
激活
主线程
CON结构入队
CON
CON
CON
CON
工作线程1
工作线程2
工作线程3
CON结构出队

状态转换


Client
CONNECT
Atlas
CONNECT_SERVER
MySQL
READ_HANDSHAKE
Atlas
SEND_HANDSHAKE
Client
READ_AUTH
Atlas
SEND_AUTH
MySQL
READ_AUTH_RESULT
Atlas
SEND_AUTH_RESULT
Client
READ_QUERY
Atlas
SEND_QUERY
MySQL
READ_QUERY_RESULT
Atlas
Client
SEND_QUERY_RESULT

内容提纲

- 背景
- 架构与原理
- 改进点
- 新功能
- 未来发展

改进点

- 主库宕机不影响读
- 连接池
- 多线程
- Lua VS C
- 字符集修正
- 加解锁语句
- 存活检测
- 消除死等
- 新协议兼容
- 线程模型

改进点之1：主库宕机不影响读

- 官方：主库宕机从库亦不可用
- 阶段1：主库宕机时可读不可写

改进点之2：连接池

- 官方：连接池形同虚设，连接数不断上涨
- 阶段1：实现了连接复用
- 阶段2：各线程连接池独立(QPS提高2倍)
- 阶段3：各用户连接池统一

改进点之3：多线程

- 官方：多线程下频繁崩溃
- 阶段1：设置独立的回收线程
- 阶段2：各线程连接池独立

改进点之4：Lua VS C

- 官方：主要的功能逻辑使用Lua脚本编写，效率低
- 阶段1：C改写，QPS提高3倍，latency降低80%
- 阶段2：线程锁粒度细化，QPS提高50%

改进点之5：字符集修正

- 官方：多个客户端分别set不同的字符集，会导致字符集混乱
- 阶段1：自动将服务端的字符集修正为客户端的字符集

改进点之6：加解锁语句

- 官方：不支持get_lock和release_lock，锁权限混乱
- 阶段1：加锁过程中保持当前连接

改进点之7：存活检测

- 官方：利用正常请求的执行结果判断DB状态，对应用有影响
- 阶段1：利用独立的检测线程判断DB状态

改进点之8：消除死等

- 官方：某台DB无法连接时会僵死
- 阶段1：重新添加事件时指定超时时间

改进点之9：新协议兼容

- 官方：5.5.7以上版本MySQL出现Unknown Command错误
- 阶段1：连接时伪装成5.5.7以下版本的客户端

改进点之10：线程模型

- 官方：所有线程监听同一个fd，惊群问题
- 阶段1：每个线程监听自己的fd
- 阶段2：同一个SQL请求由单个线程完成

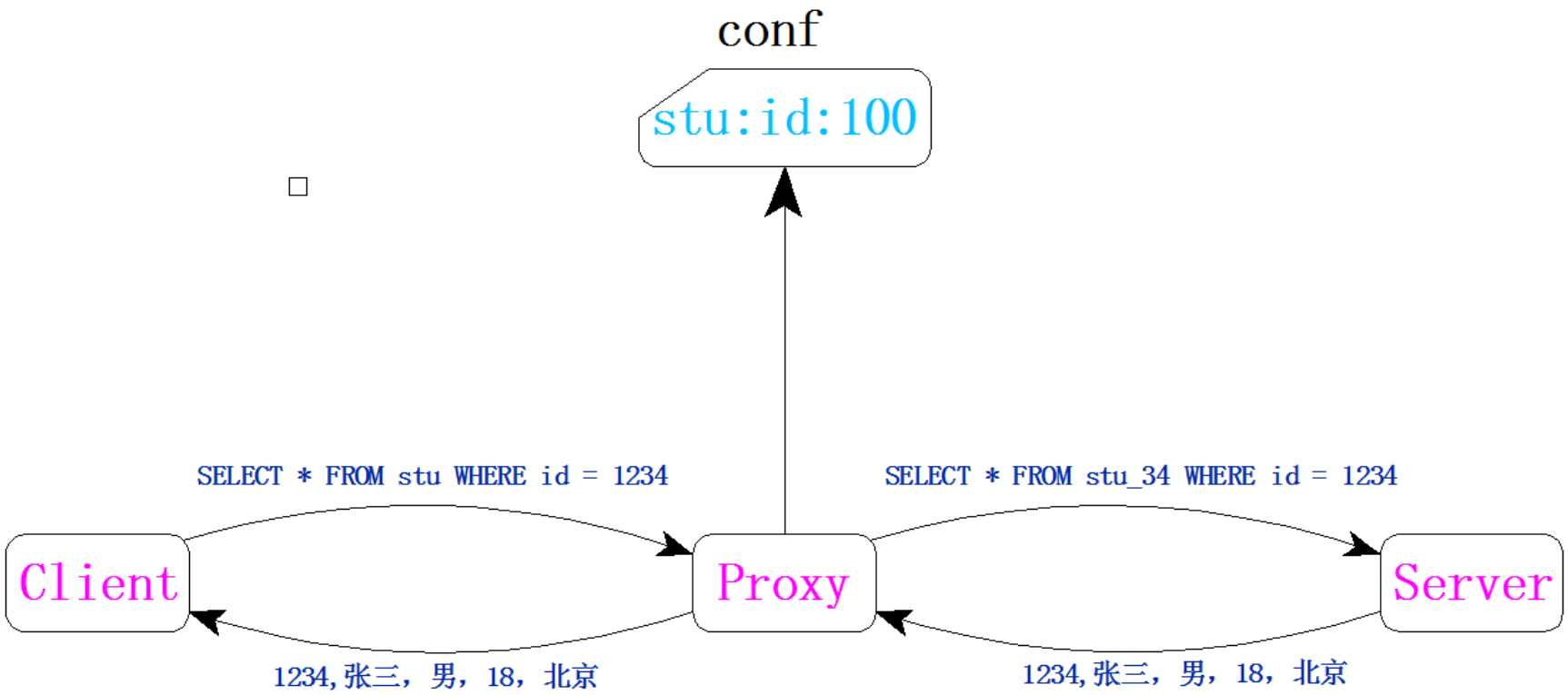
内容提纲

- 背景
- 架构与原理
- 改进点
- 新功能
- 未来发展

新功能

- 分表
- 强制读主库
- 负载均衡
- 在线增减与上下线DB
- 平滑重启
- SQL过滤
- IP过滤
- 查询日志

新功能之1：分表


conf
stu:id:100
SELECT * FROM stu WHERE id = 1234
SELECT * FROM stu_34 WHERE id = 1234
Client
Proxy
Server
1234,张三，男，18，北京
1234,张三，男，18，北京

新功能之2：强制读主库

- 避免从库的同步延迟
- /*master*/ SELECT * FROM mytable

新功能之3：负载均衡

- 精确到每个SQL请求
- 每台从库被赋予一个权重

新功能之4：在线增减DB

- select * from backends(官方)
- add master ip:port
- add slave ip:port@weight
- remove backend i
- set offline i
- set online i

新功能之5：平滑重启

- 阶段1：修改配置文件中的online标志
- 阶段2：发信号

新功能之6：SQL过滤

- 不带WHERE子句的DELETE
- SLEEP函数

新功能之7：IP过滤

- 精确IP
- IP段

新功能之8：SQL日志

- 记录所有处理的SQL语句，包括客户端IP、实际执行该语句的DB、执行成功与否、执行所耗费的时间

- [02/14/2013 16:21:41] C:192.168.1.2 S:192.168.1.3 OK 21.807 "SELECT * FROM person.mt WHERE id = 1025189561"

内容提纲

- 背景
- 架构与原理
- 改进点
- 新功能
- 未来发展

数据分片

- 将数据分布在多台机器上
- 支持跨机器分库分表

自建连接

- 可在任意阶段主动向DB发起连接请求
- 不再依赖客户端的连接动作来被动建立连接
- 不再需要参数控制连接数量

引入Zookeeper

- Atlas在Zookeeper上注册一个临时结点
- 应用从Zookeeper上查询Atlas的IP端口
- 不再依赖LVS实现failover，解决了LVS存在的固有问题，网络减少一跳

项目信息

- 开源地址：https://github.com/qihoo360/atlas

Thanks!


SequeMedia
盛拓传媒
IT168.com
www.it168.com
ChinaUnix.net
ITPUB