DTCC2011

基亍MySQL的分布式实践

杨海朝

Senior MySQL DBA@SINA

jackbillow@gmail.com

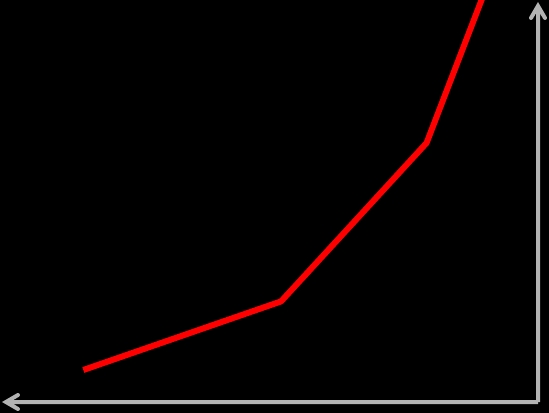
DTCC 2011

2011.4.15

业务增长

DTCC2011

usage



· Sharding策略

· Cache应用

· NoSQL应用

· 多IDC部署

议题

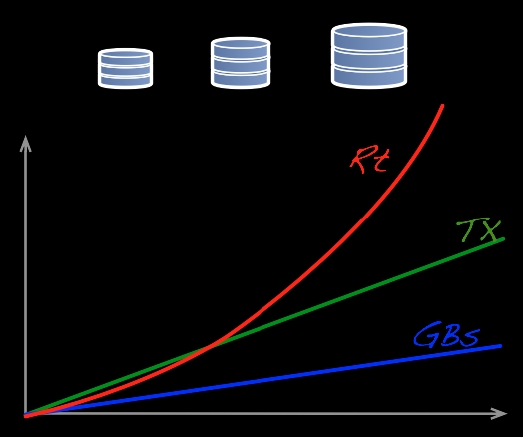
DTCC2011

Database Sharding

DTCC2011

what drives the need for database sharding?

http://www.codefutures.com/database-sharding/

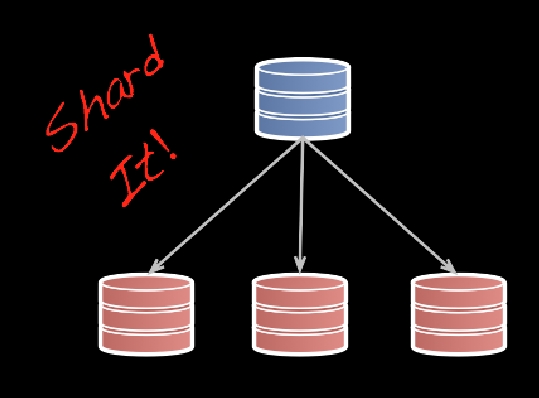


Database Sharding

what is database sharding?

http://www.codefutures.com/database-sharding/

DTCC2011



前期

· 一开始就考虑CAP & BASE理论

· 异步所有能异步的操作

· 避免分布式事务

DTCC2011

前期

按功能分割：

· 不同的业务不同的DB

· Master/Slaves

· MPSM

· 多个DB一个port

DTCC2011

前期

水平切分：

· 索引和数据在不同的DB

· Table partitioning(256张tables)

DTCC2011

Scaling实践#1

· SPSM

· 提高slave数量

DTCC2011

Scaling实践#2

· 索引和数据物理分开

· 一个DB一个端口

DTCC2011

Scaling实践#3

提高单机的性能：

· SSD

· IODrive

· Cachecade

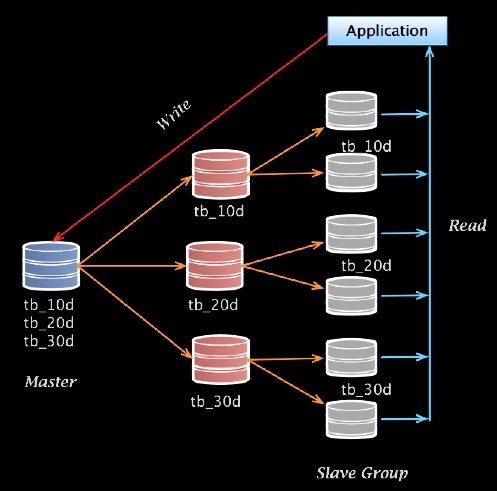
· Flashcache

DTCC2011

Scaling实践#4.1

Master不拆分Slave分成多组

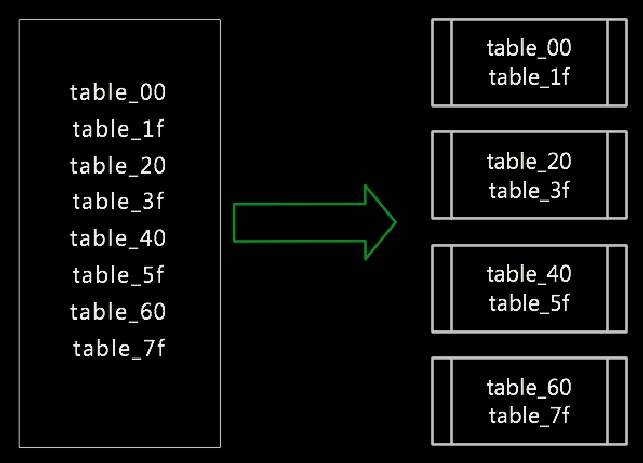
DTCC2011



Scaling实践#4.2

以表对象进行拆分

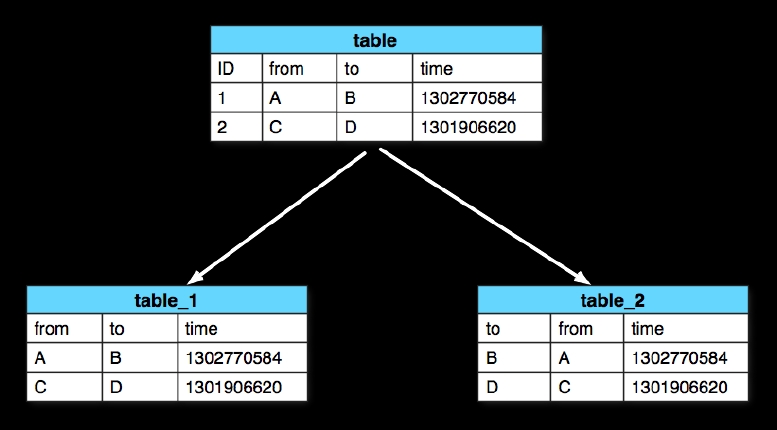
DTCC2011



Scaling实践#4.3

按不同partitioning key进行拆分

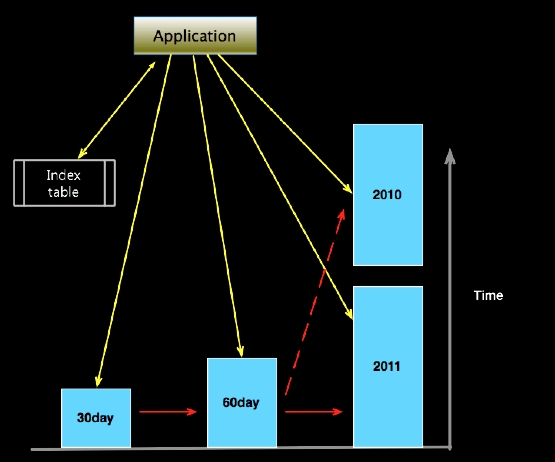
DTCC2011



Scaling实践#5

时间维度来分，持续归档

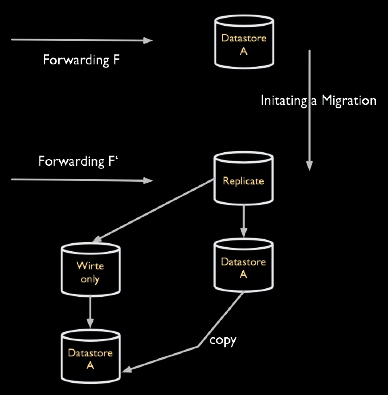
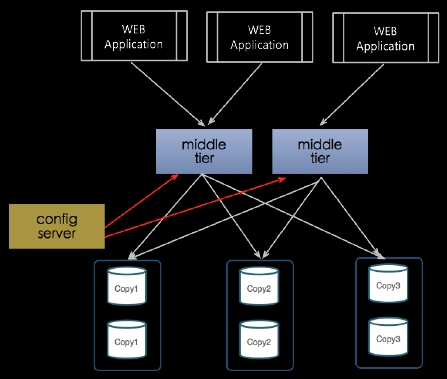
DTCC2011



Scaling实践#6

中间件来减少开发和运维的成本

DTCC2011



Sharding原则

· 一开始就关注架构设计

· Scale-up Scale-out Scale-up

· 成本可控下硬件是首选

· 逐步解决拆分中成本问题

DTCC2011

挑战

· 如何减少TCO？

· 如何权衡业务后期的功能扩展?

· 故障率如何降低？

· 大批量的快速部署？

DTCC2011

Caching

DTCC2011

Everything runs from memory in Web 2.0

--Evan Weaver

Caching

DTCC2011

· Delay is the fault

--Me

DB的延时越来越不可以接受，进行了并发

复制的开发，降低响应时间还是没有解决

Caching

第一阶段: MySQL + Memcached

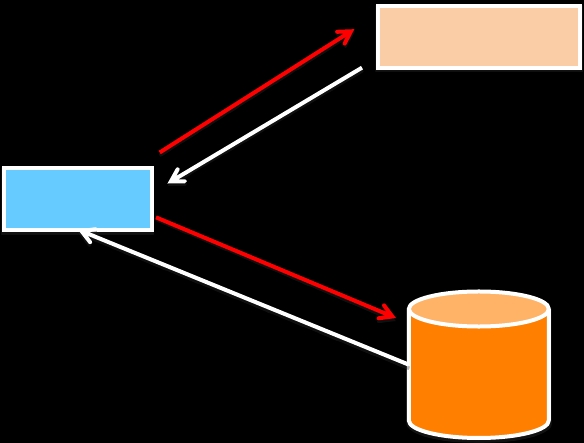
Memcache

Server

Client

Database

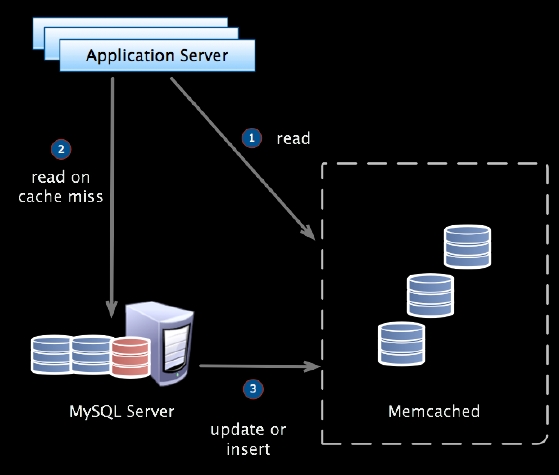
DTCC2011



Caching

第二阶段: MySQL + UDF

DTCC2011

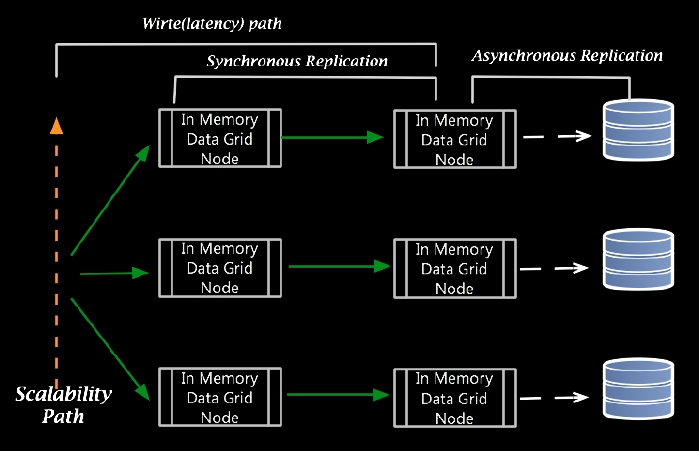


Caching

第三阶段: 按冷热度分层

http://natishalom.typepad.com/nati\_shaloms\_blog/2010/10/nocap.html

DTCC2011



Caching原则

· 采用一致性Hash部署

· Cache按照冷热分层

· 所有热数据都放入Cache

· 双写来避免雪崩问题

· 队列方式持久化落入MySQL

DTCC2011

挑战

· 双写导致应用复杂？

· 实现locality RAM?

· 多IDC部署Cache的复杂度问题？

DTCC2011

NoSQL

· No use only MySQL

· 高性能

· 丰富的数据类型

DTCC2011

独立redis：

· 关系

· 计数

· 通知

……

NoSQL

DTCC2011

NoSQL

DTCC2011

MySQL Redis复制

· 通过RBR解析BINLOG同步到redis

· Redis提供特定数据结构的读访问

· 实现关系型数据转变成队列数据

NoSQL

DTCC2011

Redis MySQL复制

· Redis提供特定数据结构的读写

· 通过replication接口同时写入到MySQL

NoSQL

对redis进行完善

· 解决replicaton问题(rdb+aof)

· 容灾问题(加入position思想)

DTCC2011

挑战

· 如何批量管理？

· 后期容量扩展？

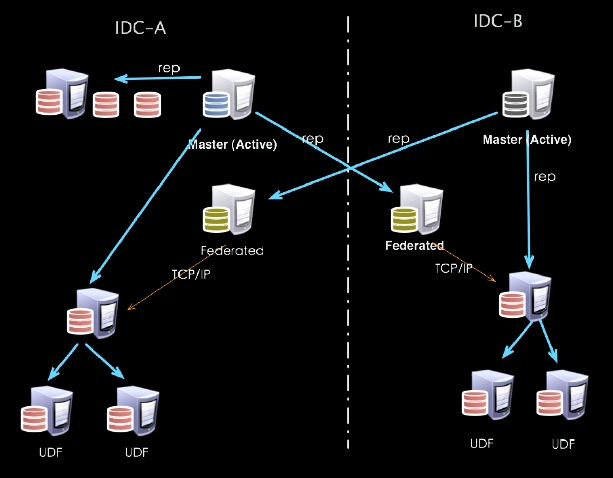
· 复杂数据的load速度问题？

DTCC2011

Federated table

多IDC部署

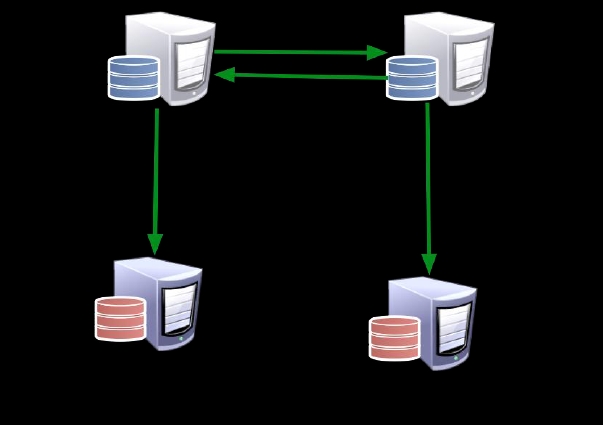
DTCC2011



Master/Master

多IDC部署

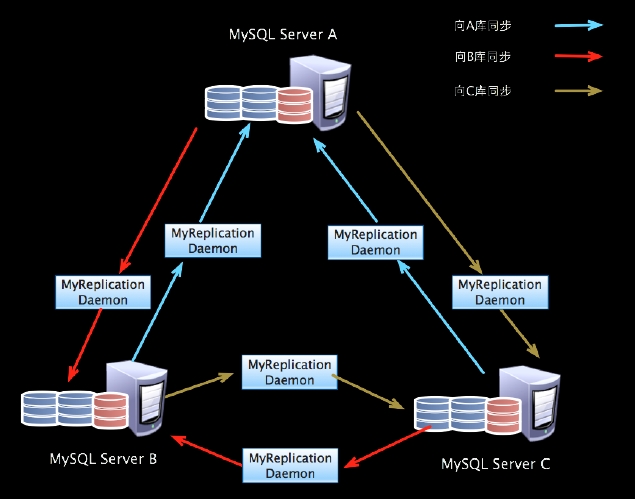
DTCC2011



开发多主复制

多IDC部署

DTCC2011

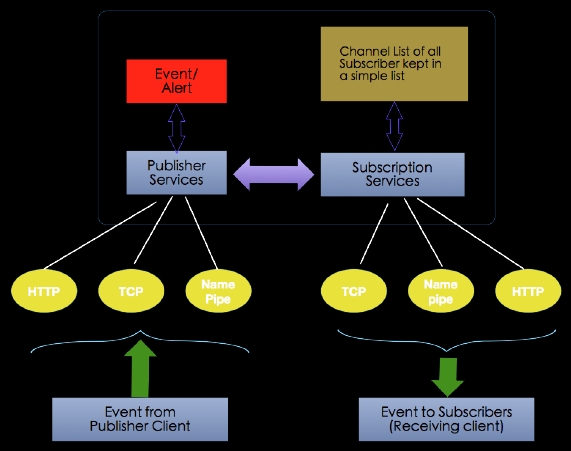


多IDC部署

Pub/sub + MySQL replication

http://www.codeproject.com/KB/WCF/ASENS.aspx?msg=2937455

DTCC2011



挑战

· 延时带来的应用时序问题？

· 部署datacenter的数量有限？

DTCC2011

总结

· 持续的database sharding

· 硬件解决是首选

· 数据库简化为存储

· 关系型和非关系结合

· 谨慎考虑多机房部署

DTCC2011

DTCC2011

Thank you for coming!

@jackbillow

jackbillow@gmail.com

Q&A

长期招聘，欢迎加入！