Vitess 调研报告

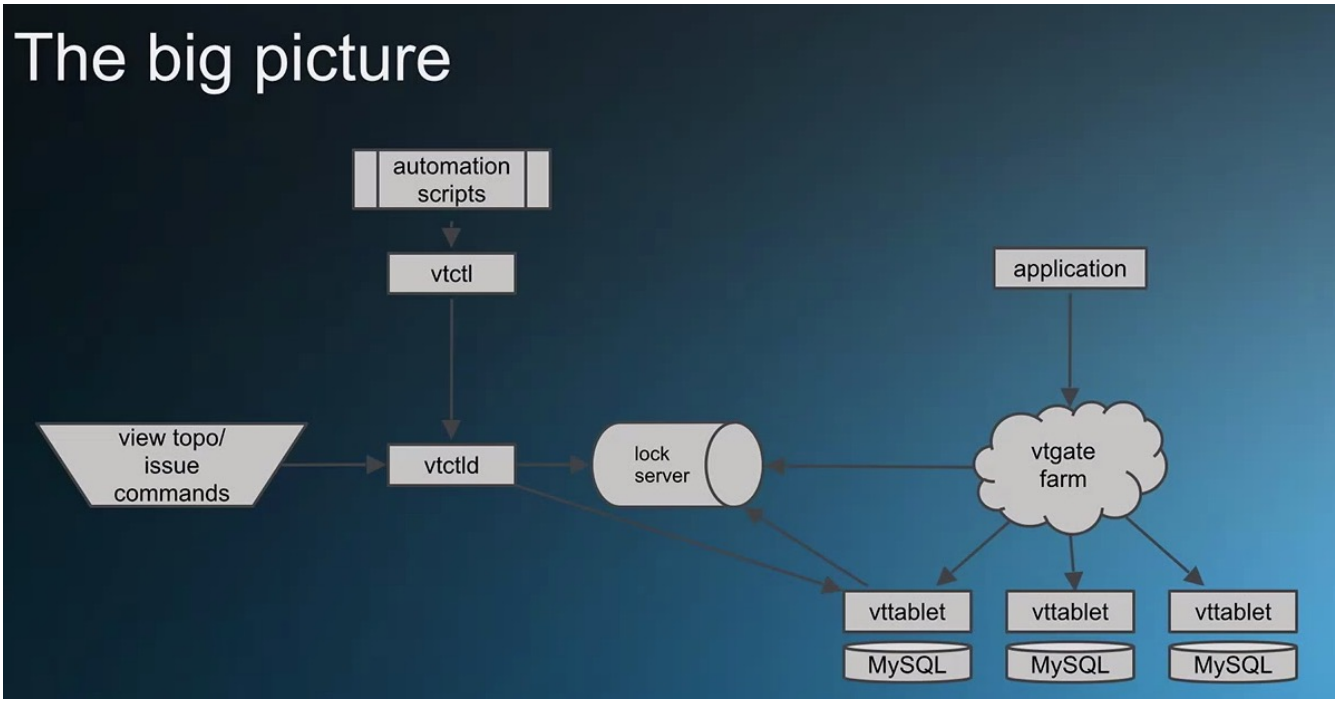
需求

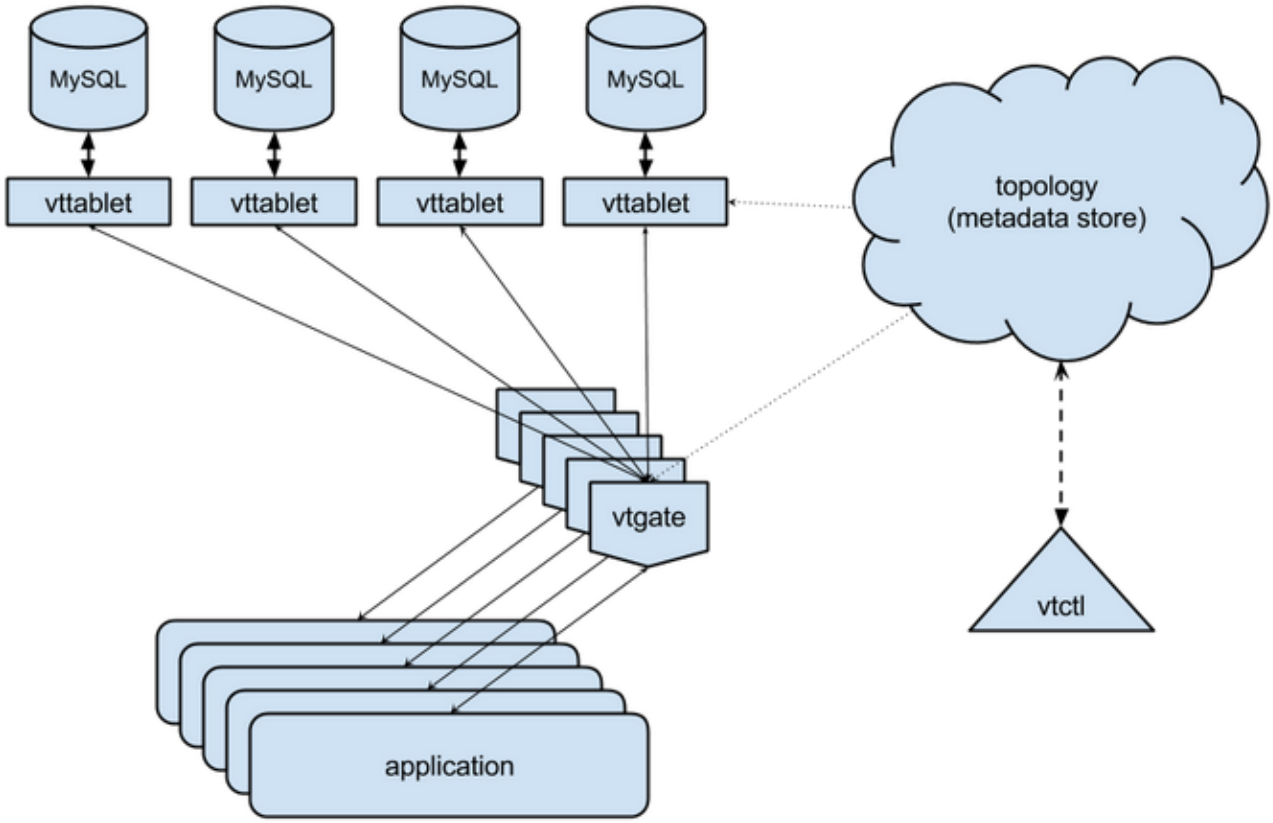
- 截止2014-1126,开源社区对MySQL Sharding方案,国内外的产品有几个候选,我们无论是采用,还是自主开发(借鉴)都需要去调研.
- Vitess是youtube采用多语言(核心golang)开发的MySQL Sharding方案,从2011年开发9个月,后来上线在youtube核心的视频业务上.

目的

- 参考候选方案, 为减少后续工作提供尽可能多的依据

架构





核心算法

- Range Based Sharding 唯一支持

1. 数据分片与算法

a. resharding / split : both horizontal sharding andvertical sharding
b. resharding:::
    i. The process to achieve this goal is composed of the following steps:
        - pick the original shard(s)
        - pick the destination shard(s) coverage
        - create the destination shard(s) tablets (in a mode where they are not used to serve traffic yet)
        - bring up the destination shard(s) tablets, with read-only masters.
        - backup and split the data from the original shard(s)
        - merge and import the data on the destination shard(s)
        - start and run filtered replication from original to destination shard(s), catch up
        - move the read-only traffic to the destination shard(s), stop serving read-only traffic from original shard(s). This transition can take a few hours. We might want to move rdonly separately from replica traffic.
        - in quick succession:
            - make original master(s) read-only
            - flush filtered replication on all filtered replication source servers (after making sure they were caught up with their masters)
            - wait until replication is caught up on all destination shard(s) masters
            - move the write traffic to the destination shard(s)
            - make destination master(s) read-write
        - scrap the original shard(s)

2. 数据迁移与算法

a. 见上面的描述
b. 使用 mysqldump + vttable(bin files)过滤处理

一些建议/考虑点

- 程序核心用 golang 开发，代码量少/开发快，性能接近于C
- 集成测试/命令工具行使用 python 开发，开发快
- 暂时缺少Web管理工具，相反cli工具集较全
- 目前,分支版本众多，Features不断加入
- 从代码上看：默认支持MariaDB 有测试， MySQL 公版没有完全测试，Google改版MySQL是第二个支持
- 各个工具的特性和MySQL本身结合很紧密，对MySQL自身的原理很清楚. 并且有很大的改进
- vtgate 很轻量的路由 + vttablet(重要功能并且耗时SQL Query处理) -- 对于mysql特性 ， 从整体架构上有更大的优势

功能点与限制

功能\产品	Vitess		
————————————————	Supported?	相应工具	command
SQL parser,保护MySQL	Y	vttablet / vtocc	
Query rewrite and sanitation	Y	vttablet / vtocc	
Query blacklisting	Y	vtocc	
Table ACLs	Y	vtocc	
Query killer	Y	vtocc	
Transaction management	Y	vtocc	
monitoring features	Y	vtocc	
horizontal/vertical sharding	Y	Vitess toolchain	
resharding	Y	Vitess toolchain	
Reparenting	Y	Vitess toolchain	vtctl ReparentShard
different replicas	Y	vtctl+ vttablet	
connection pooling	Y	vtocc	
workflow management	Y	Vitess toolchain	
indexs. create secondary indexes on your tables	N		
joins	N		
RowCache	Y	vttablet(2014done,计划明年上线)	
CI Testing	Y	Python code	
Configuration	Y	vtctl + zookeeper	
track shards	Y	vtctl	
replication graphs	Y	vtctl	
db categories	Y	vtctl	
initiate failovers	Y	vtctl	
init vttablet	Y	vtctl	
filtered replication	Y	vttablet	
data export	Y	vttablet	
unified view of the entire fleet	Y	vtgate	
vtctl manager	Y	vtctld	
resharding differ jobs	Y	vtworker	
vertical split differ jobs	Y	vtworker	
mysql instance manager	Y	mysqlctl	

zk manager	Y	zkctl	
Schema Management	Y	vtctl	
Replication Graph	Y	vtctl	vtctl InitTablet
Serving Graph	Y	vtctl	
Cell (Data Center)	N		
ACL	N	有部分代码	