

MySQL最佳实践32条

知数堂 吴炳锡
2018年5月



大纲

硬件方面

MySQL版本

开发方面

安全

操作系统

MySQL配置

运维事项

总结

开篇

从事MySQL DBA工作十几年，遇到很多低级类错误，一直有一个心愿，把这些低级类的错误整理分享出来，以方便使用MySQL的人员能快查看一下，避免少掉很多坑。

----- 知数堂

硬件方面

1

01

BIOS

关闭NUMA
关闭节能模式
Raid卡强制Force WriteBack (WB)

02

CPU

高主频VS多核

03

MEM

内存越大越好

04

HDD VS SSD

15000转的SAS 145个IOPS
PCI-E SSD 10万IOPS起步

BIOS

- BIOS配置优化
 - CPU设置最大性能模式，关闭C1E，C-stats
 - 内存设置最大性能模式
 - 关闭NUMA，5.6.27以上的版本，新增选项innodb_numa_interleave
- RAID配置优化
 - RAID-10
 - CACHE & BBU
 - WB & FORCE WB
- 使用PCIe-SSD等高速I/O设备

BIOS

Dell Inc. (www.dell.com) - PowerEdge R710 BIOS Version 3.0.0	
Service Tag:	Asset Tag:

System Time	15:18:19
Sy	Power Management Maximum Performance
Me	
Pr	CPU Power and Performance Management Maximum Performance
	Fan Power and Performance Management Maximum Performance
SA	Memory Power and Performance Management ... Maximum Performance
Bo	
Integrated Devices	<ENTER>
PCI IRQ Assignment	<ENTER>
Serial Communication	<ENTER>
Embedded Server Management	<ENTER>
Power Management	<ENTER>

Up,Down Arrow to select	SPACE,+, - to change	ESC to exit	F1=Help
-------------------------	----------------------	-------------	---------

- sysstat工具集
 - sar、pidstat、mpstat
- perf top -p \$pid
 - 实时显示系统/进程的性能统计信息
- pt-ioprofile
 - mysql中的iotop
- pstack
 - 当mysql里有线程hang住时，利用pstack排查由于哪些函数调用存在问题

NUMA

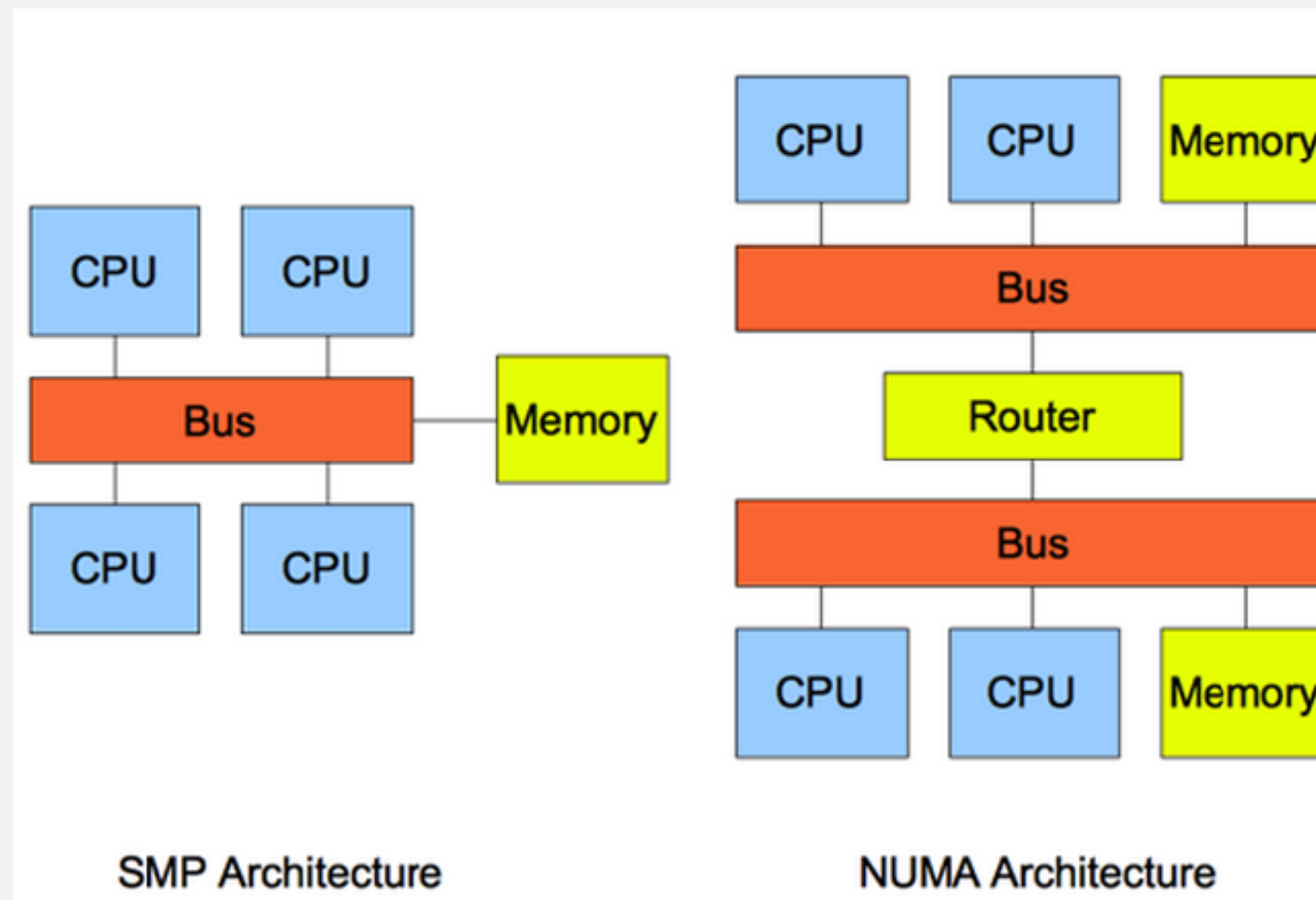
System Time 14:41:21
System Date ..

Memory Setting
Processor Sett

SATA Settings

Boot Settings

System Memory Size 64.0 GB
System Memory Type ECC DDR3
System Memory Speed 1333 MHz
Video Memory 8 MB
System Memory Testing Disabled
Redundant Memory Disabled
Node Interleaving Enabled



Raid

Virtual Disk Management

Create New VD

RAID Level : RAID-0
RAID-1
RAID-5
RAID-6
RAID-10
RAID-50

PD per Span :
Physical D
Disk ID

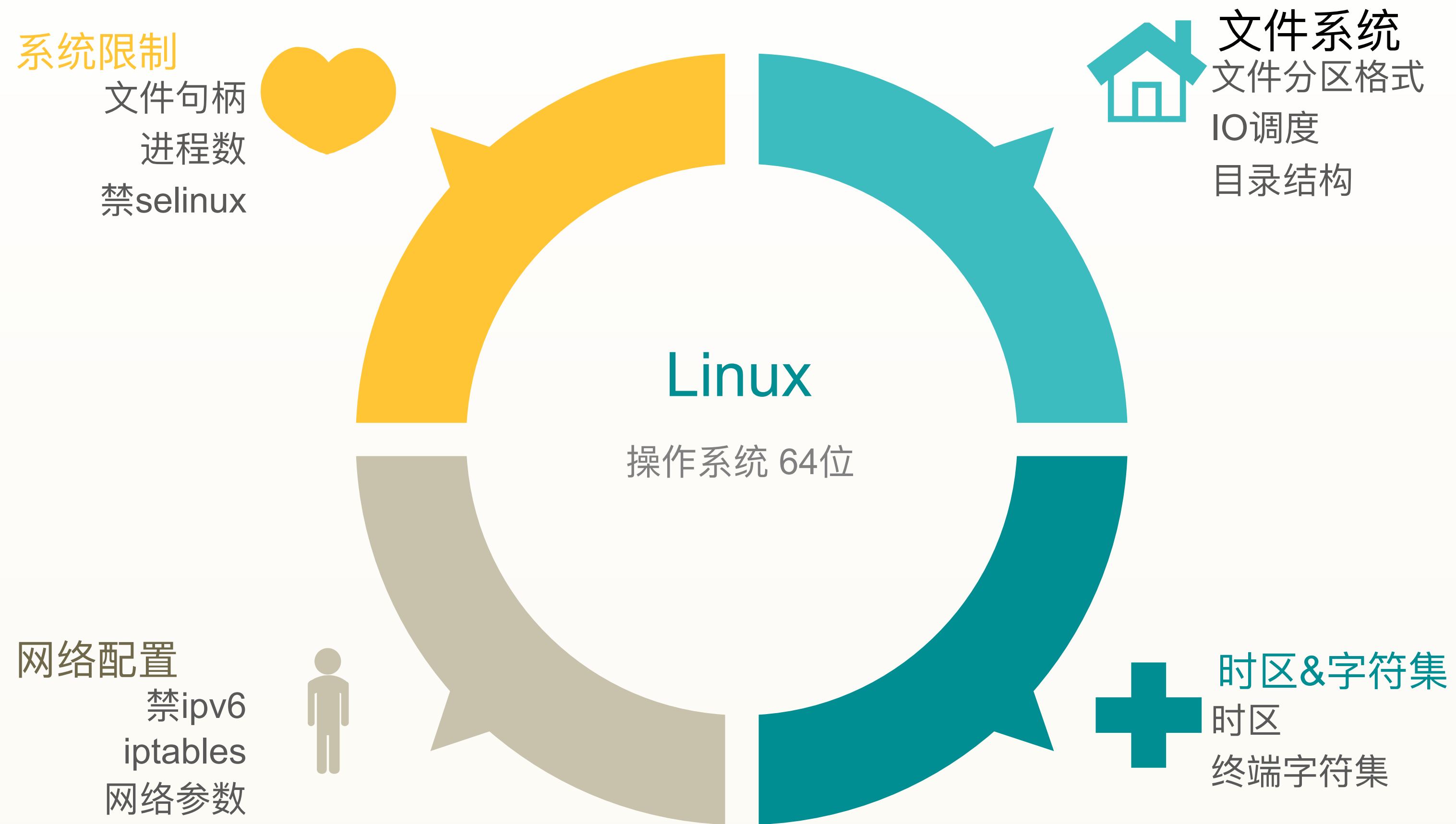
[X] 101:00:00	B	# 00
[X] 101:00:01	157.12 GB	01
[] 101:00:02	476.37 GB	--
[] 101:00:03	476.37 GB	--
[] 101:00:04	476.37 GB	--
[] 101:00:05	476.37 GB	--

Secure VD:
No

Basic Settings
VD Size: 334.25 GB
VD Name:

[] Advanced Settings
Strip
Element Size: 64KB
Read Policy : Adaptive R
Write Policy: Write Back
[] Force WB with no battery
[] Initialize
[] Configure HotSpare

OK
CANCEL



推荐： CentOS Ubuntu

先择通用性强的，内核高版本
性能好，别外Oracle UEK内核
也是一个不错的选择。

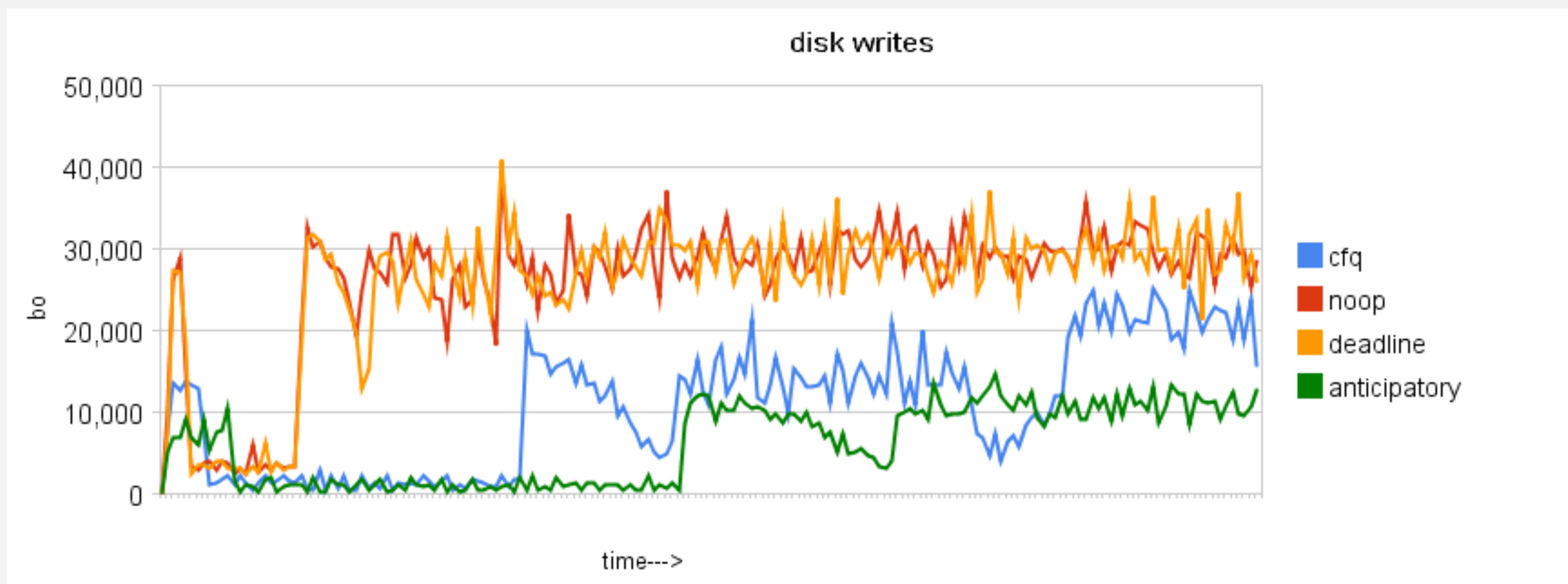
系统限制

- 默认： 1024 open files, 4096 max user processes
- /etc/security/limits.conf
增大这些值
- 禁止selinux
 - /etc/sysconfig/selinux
SELINUX=disabled

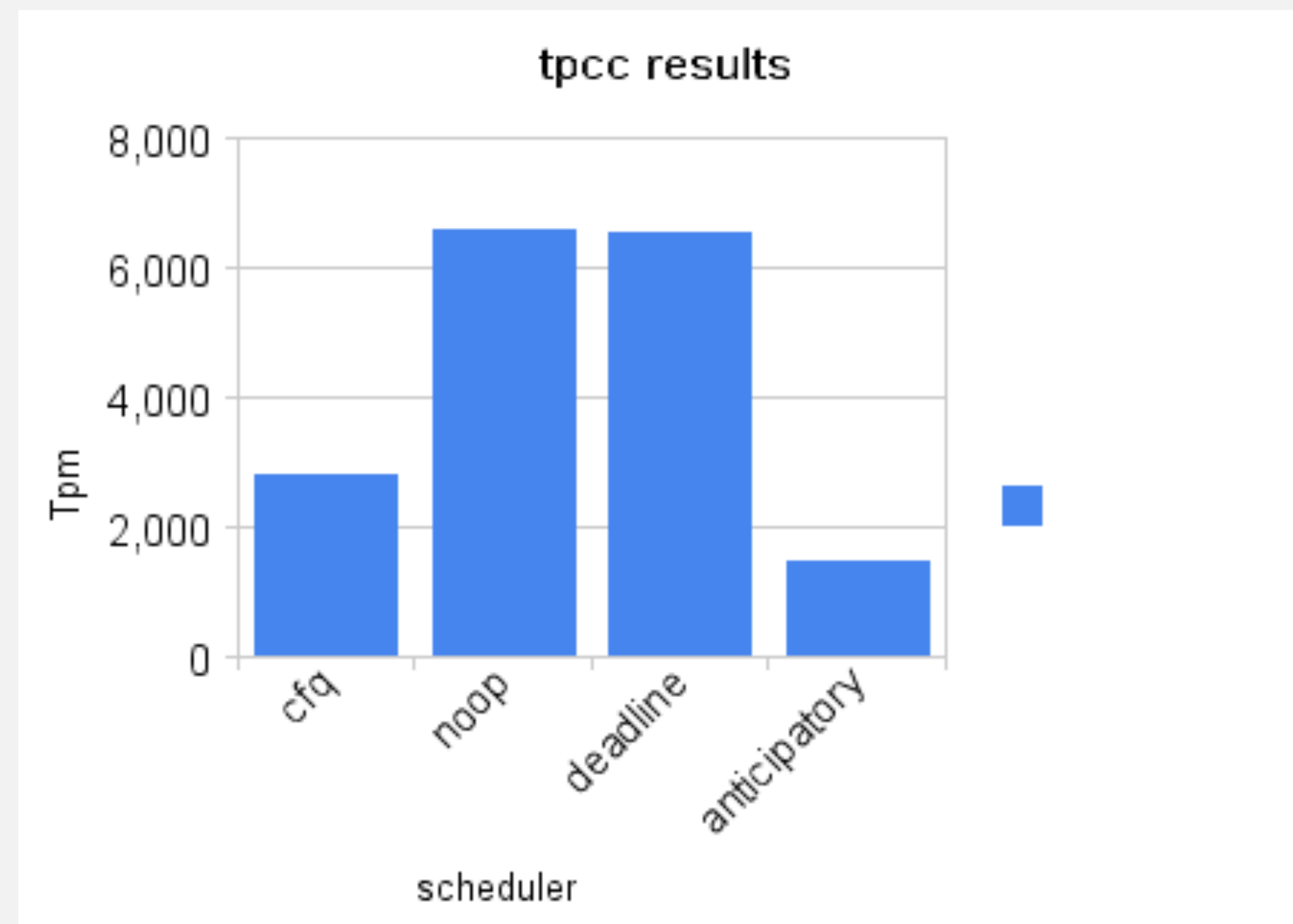
文件系统

- io scheduler
 - deadline/noop
- filesystems
 - xfs
 - ext4
- 别用NFS放置datadir
- kernel
 - vm.swappiness = 1 - 5 (rhel7以后, 设置0要慎重)
 - vm.dirty_background_ratio = 5
 - vm.dirty_ratio = 10

文件系统



文件系统



对于短连接应用

- `/etc/sysctl.conf`
`net.ipv4.tcp_max_syn_backlog = 819200`
`net.core.netdev_max_backlog = 400000`
`net.core.somaxconn = 4096`
`net.ipv4.tcp_tw_reuse=1`
`net.ipv4.tcp_tw_recycle=0`
- `hostname`
`/etc/hosts`

时区&字符集

- MySQL使用系统的timezone设置产生时间
- 主从系统时区必须一致
- 建议改更mysql的配置: `time_zone='+8:00'`
- 系统字符集一致
 - locale

```
#locale
LANG=en_US.UTF-8
LC_CTYPE="en_US.UTF-8"
LC_NUMERIC="en_US.UTF-8"
LC_TIME="en_US.UTF-8"
LC_COLLATE="en_US.UTF-8"
LC_MONETARY="en_US.UTF-8"
LC_MESSAGES="en_US.UTF-8"
LC_PAPER="en_US.UTF-8"
LC_NAME="en_US.UTF-8"
LC_ADDRESS="en_US.UTF-8"
```

MySQL版本

线上环最好统一版本，统一操作系统，减少环境差异带来的不必要的麻烦。

目录结构

版本选择

单实例VS多实例

目录结构

- basedir : /usr/local/mysql
- 其它目录 /data/mysql/mysql\$port/{data,tmp, logs, my\$port.cnf}
- datadir: /data/mysql/mysql\$port/data

```
[root@zst2 08:11:18 /data]
#ls -l /usr/local/mysql
lrwxrwxrwx 1 mysql mysql 46 Jan 10 21:10 /usr/local/mysql -> /opt/mysql/mysql-5.7.20-linux-glibc2.12-x86_64
```

- 标准化
- 方便自动化
- /data挂载点独立分区

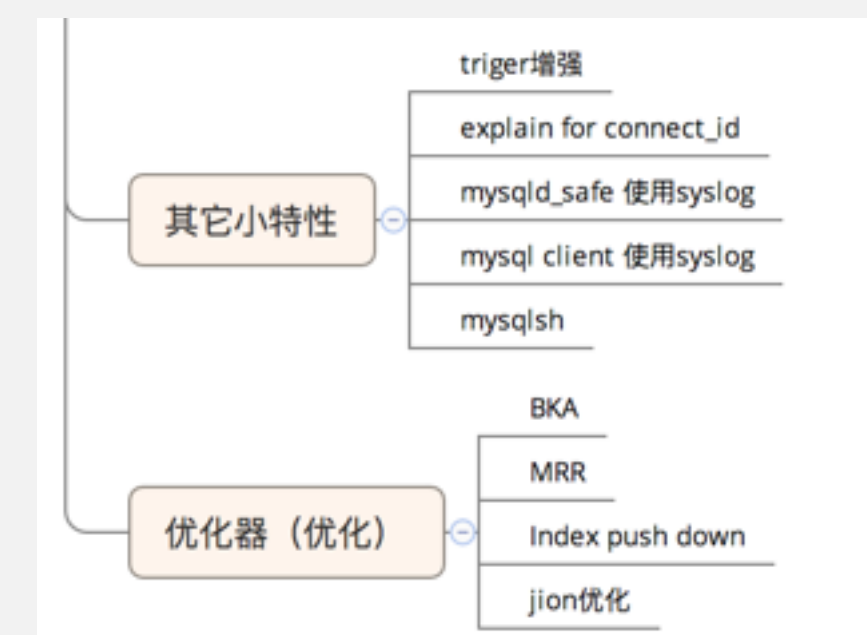
版本选择

- CentOS/Redhat:
 - `yum install mysql-server`
 - `yum install mysql-client`
 - `[root@zst1 ~]#mysql --verson`
mysql Ver 14.14 Distrib 5.7.20, for linux-glibc2.12 (x86_64) using EditLine wrapper
 - `[root@zst2 ~]#mysql --verson`
mysql Ver 14.14 Distrib 5.6.34, for linux-glibc2.5 (x86_64) using EditLine wrapper
- **强烈推荐:**
 - 下载二进制安装, 统一版本
 - 构建自己的yum安装, 统一版本

版本选择

- 版本选择
 - MySQL/MariaDB/Percona
 - 选择新版本
 - changelog 阅读
 - 新特性引入

MySQL 5.7 特性



单实例 VS多实例

- 私有业务 & 云平台RDS化业务
- 虚拟化和非虚拟化
- 资源分配对齐

虚拟的Min型：

2000 IOPS
CPU： 2-4
内存大最4-16G
DB大小最大： 100G
磁盘最大： 500G

中型多实例

6000 IOPS
CPU： 4-16
内存大最8-32G
DB大小最大： 500G
磁盘最大： 1T

Big型： 单实例

3W IOPS
CPU： 16-32
内存大最32-128G
DB大小最大： 1T
磁盘最大： 3T

MySQL配置

对于配置的优化是不是有作用，最好的办法就是压力测试跑一下，当然调优的基础也需要理论来支撑。

----- 知数堂 叶大师

Innodb Buffer Pool

连接数

放弃MyISAM

复制类型

Innodb buffer Pool

- 推荐 50%-80%的内存
- Innodb Buffer Pool中含: data+index
- 可以通过增大Buffer Pool减小磁盘的IO
- show engine innodb status\G
- Buffer pool hit rate : 999/1000
- innodb_buffer_pool_size=50%-80%内存
- innodb_buffer_pool_instances=4-8

连接数

- max_connections
- max_user_connections
- MySQL默认给Super权限的帐号保留一个连接，支持的最大连接数：max_connections+1
- `gdb -p `pidof mysqld` -ex "set max_connections=2048" --batch`
- 提示：
 - wait_timeout default : 28800
 - interactive_timeout default: 28800

放弃MyISAM

- 只能使用一个Core CPU
 - 内存最大支持4G，只缓存索引
 - 不支持同时并发读和更新
 - 无事务支持，表容易Crash，可修复，丢数据
 - 并发能力弱
-
- 需要高速写入环境，试一下TokuDB
优化版本： <https://github.com/xelabs/tokudb>

复制类型

复制类型	说明	适用
传统复制 <i>binlog row+position</i>	版本MySQL 5.5以下	有点老了，可以考虑退出
GTID复制 <i>binlog row+gtid</i>	版本MySQL 5.6以上	跨IDC，对复制要求不严格环境
半同步 (GTID)	MySQL 5.5 不支持GTID MySQL 5.6 支持GTID	不推荐（这两个版本存在性能问题） MySQL 5.7 可用
增强半同步 (GTID)	MySQL 5.7	推荐，金融环境，同IDC，同城
多源复制 (GTID)	MySQL 5.7	OLAP节点，延迟复制

开发方面

开发以功能为主，项目比较紧，经常是怎么方便怎么来，所以造成考虑的东西比较少。DBA变成了保姆或是抗雷的。

----- DBA

索引/主建选择

Many Tables

容量规划

Bad Query

索引&主键选择

- InnoDB以主键排序存储
 - 普通索引包含主键存储
 - UUID 目前不是顺序增长，做为主键写入导致，随机IO严重
-
- 联合索引 性能优于 独立索引
 - 区别度低的字段，在联合索引中效果不错

显式主键

第一个非NULL的唯一索引

6字节的RowId

Many Tables(5.7)

- 每个DB是一个数据库目录
- Innodb每个表包含两个文件(.frm+.ibd)
- 某论坛将近200张表，共享系统安装1000个，将近200000表
- 访问information_schema 相当于打开了一次所有的表
- 物理文件系统访问非常慢

推荐：

每个实例最多1万张表

最佳实践单实例小于500张表

容量规划

- 资源对齐分配!!!
 - 单实例最大1T
 - 单机多实例，规范管理
-
- 不是DB放不下，只是数据库大了，对于核心库的维护备份恢复上太难处理，开发的同学，也要想想怎么维护。

Bad Query

- 用不到索引
- 一次读取过多记录
- like "%zst%"

mysql> explain select * from text where data like '%something';

id	select_type	table	type	possible_keys	key	key_len	ref	rows	Extra
1	SIMPLE	text	ALL	NULL	NULL	NULL	NULL	2092230	Using where

▪ JOIN列无索引

mysql> explain select text.* from text join text2 on text.data = text2.data where text2.data = 'something';

id	select_type	table	type	possible_keys	key	key_len	ref	rows	Extra
1	SIMPLE	text2	ALL	NULL	NULL	NULL	NULL	2091815	Using where
1	SIMPLE	text	ALL	NULL	NULL	NULL	NULL	2092230	Using where

Bad Query

- 监控慢查询
 - pt-query-digest
 - Anemometer
- 干掉垃圾查询
 - pt-kill
 - 优化

Cache
MySQL+NoSQL

运维注意事项

alter table

MySQL 大表

复制延迟

变量动态修改

其它

Alter table

- 直接alter table容易锁表
- 高并发系统直接alter table易出现拿不到锁，导致server Crash
- MySQL 5.6 支持在线的online ddl，但容易造成复制延迟
- 对于alter table操作，请查看每个版本，确认只是更改字典信息，无copy可以在线操作
- 拿不准情况推荐： pt-osc

MySQL大表

- delete, update大表, 分批段进行
- 更改Session会话binlog_format='statement'
- 大表删除

```
#ls stock.* -l
-rw-rw-- 1 MySQL mysql 9196 Apr 14 23:03 stock.frm
-rw-r--r- 1 mysql mysql 19096666112 Apr 15 09:55 stock.ibd

#ln stock.ibd stock.ibd.hdlk
#ls stock.* -l
-rw-rw-- 1 MySQL mysql 9196 Apr 14 23:03 stock.frm
-rw-r--r- 2 mysql mysql 19096666112 Apr 15 09:55 stock.ibd
-rw-r--r- 2 mysql mysql 19096666112 Apr 15 09:55 stock.id.hdlk

mysql>drop table stock;
Query Ok, 0 rows affected(0.92 sec)
```

复制延迟

- 复制中sql_thread单线程工作
- 大的delete/update/alter会让复制延迟比较长
- MySQL 5.6 支持库级别并行复制
- MySQL 5.7 支持基于事务级的并行复制(group commit)
- show slave status\G ;
- mysqlbinlog -v --base64-output=decode-rows --start-position=exec_master_log_pos relay_master_log_file

变量动态修改

- set global innodb_buffer_pool_size=N;
- set join_buffer_size=4*1024*1024;
- set

重启，更改丢失

会话变量，连接不Close会影响其它会话

其它注意-过多的Slave

- Master 在Binlog中记录所有的变更
- 所有的Slave需要拉取全量的Binlog
- 过多的Slave对主的网络及IO压力较大
- 如： 每分100M写入， 20个Slave

$20 * 100 \rightarrow 2\text{GB/分} \rightarrow 33.33\text{M/S}$

MySQL安全

用户名&密码

权限分配

网络监听

load data

备份&恢复

用户名和密码

权限分配

- 低于MySQL 5.7的版本删除匿名帐号和泛授权
 - `delete from mysql.user where user="" or host!='localhost';`
 - `drop database test; truncate table mysql.db; flush privileges;`
- 不要使用Root做业务帐号
- 不要使用无密码帐号做业务帐号
- 合理控制权限, 不要把用户名和密码写到程序中
- 创建用户名 长度8-16个字符, 密码8-32个字符
 - 避免使用关键词做用户名和密码

用户名和密码

权限分配

- DB系统包含：
 - 业务帐户
 - Delete, Update, Insert, Select, Execute
 - 监控帐号
 - replication client, show
 - 业务平台更新帐号
 - create, drop, 含业务帐户
 - 管理帐号 (with grant option)
 - 创建用户, 权限管理相关

网络监听

- 禁ipv6, iptables
- 只监听内网

load data

- MySQL 5.6及以下版本

```
mysql> LOAD DATA LOCAL INFILE '/etc/shadow' INTO TABLE `text` (data);  
Query OK, 27 rows affected (0.00 sec)
```

- MySQL 5.7及以上版本

```
root@localhost:mysql3306.sock [wubx]>load data infile '/etc/passwd' into table t1(t);  
ERROR 1290 (HY000): The MySQL server is running with the --secure-file-priv option so it cannot execute this statement
```

备份恢复

- 逻辑备份
 - mysqldump
 - mydumper
- 在线热备
 - xtrabackup
- Point-to-Point备份恢复
 - binlog + 复制
- 快速单表恢复
- 快速增量应用

备份安全

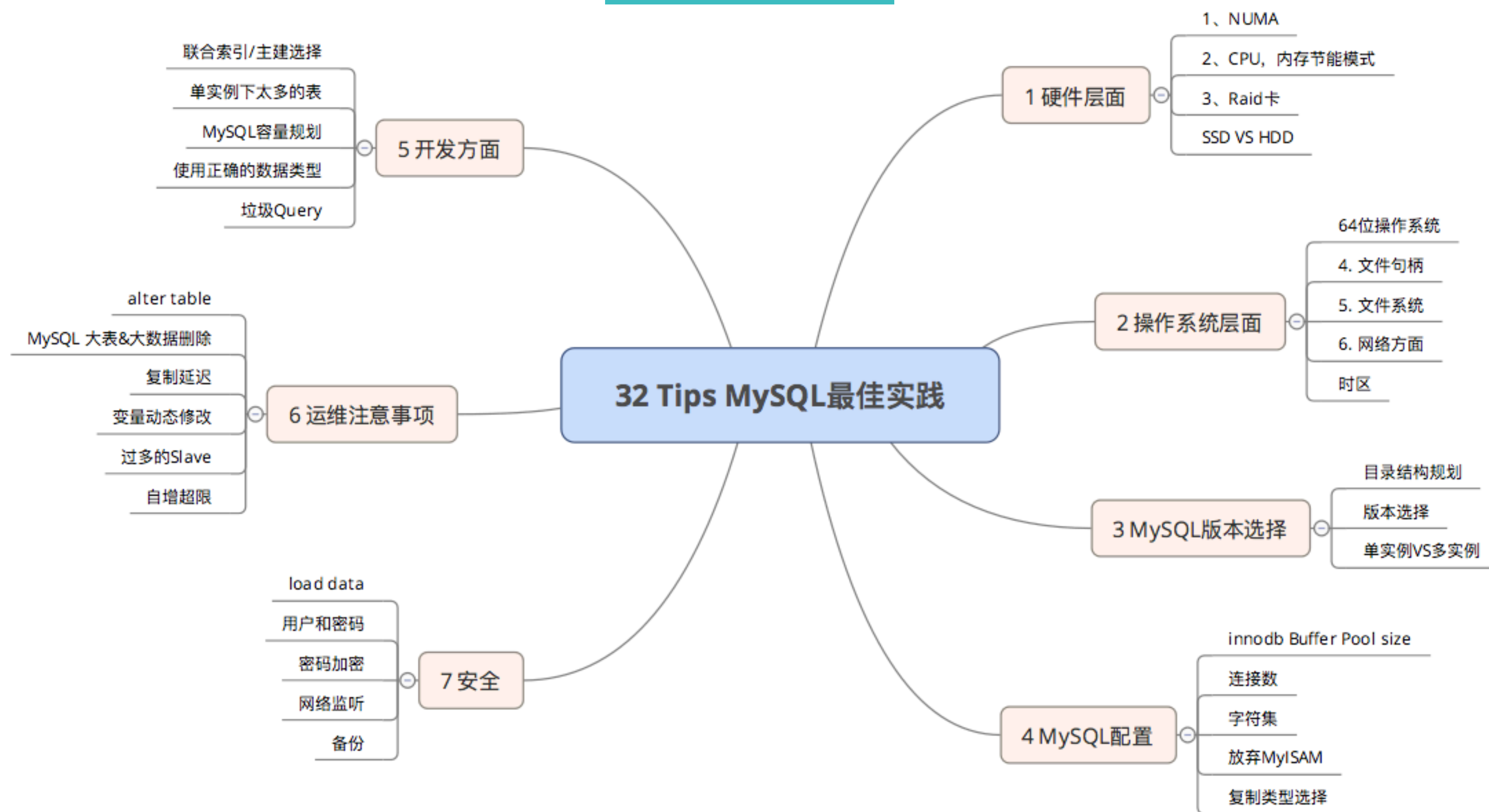
备份可用校验

脱库->备份

备份单点存储

异存备份还是放从库？

希望你帮助 QA环节



<https://github.com/zhishutech/mysql-sql-standard>

知数堂MySQL开发规范

等你来丰富