

Lest-PB级持久化缓存系统

大嘴

xuhaifeng@yuewen.com

说起缓存.....



副作用

- ❑ 内存式缓存，断电就清空数据
- ❑ 缓存down机，无法马上恢复，系统被击穿
- ❑ 缓存大小受限于内存，成本提高
- ❑ 缓存分布一般由客户端完成，更改、迁移算法成本大
- ❑ 缓存无同步、无容灾
- ❑ 只增不减、关键是不敢减、不敢下
- ❑ 机器越来越多，管理成本越来越高
- ❑

拒绝缓存



拒绝

决绝



DTCC
2018

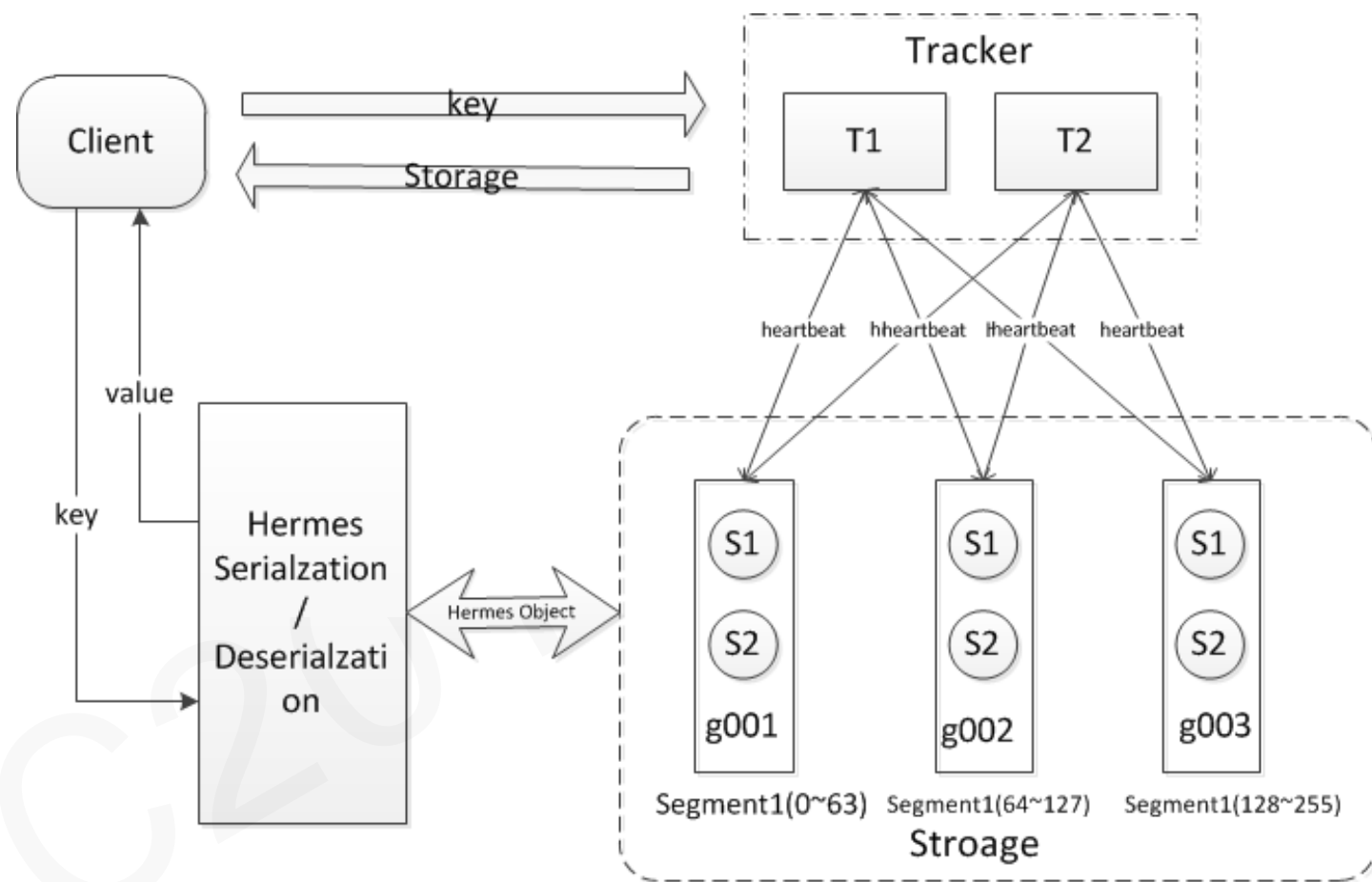
数领先机 智赢未来 (9)

IT168.com

ChinaUnix

ITPUB

强身健体



- 杜绝缓存易丢失：持久化缓存内容
- 缓存的必备条件：支持海量数据、高并发访问
- 解决数据恢复慢：使用gossip协议sync
- 缓存分布策略：单独组件完成均匀sharing
- 缓存内容：String、List、Map

负载均衡

DTCC
2018

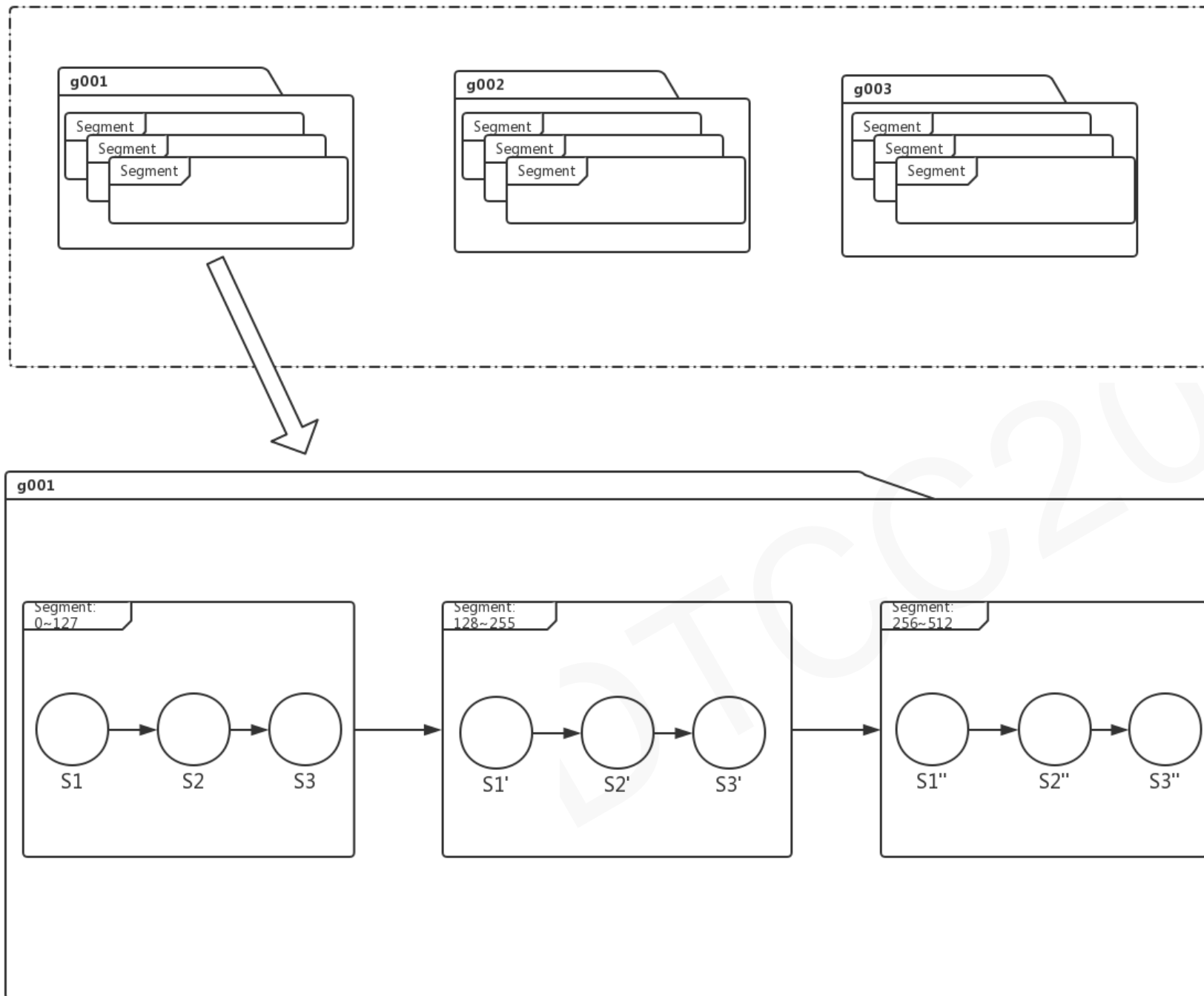
数领先机 智赢未来 (9)

IT168.com

ChinaUnix

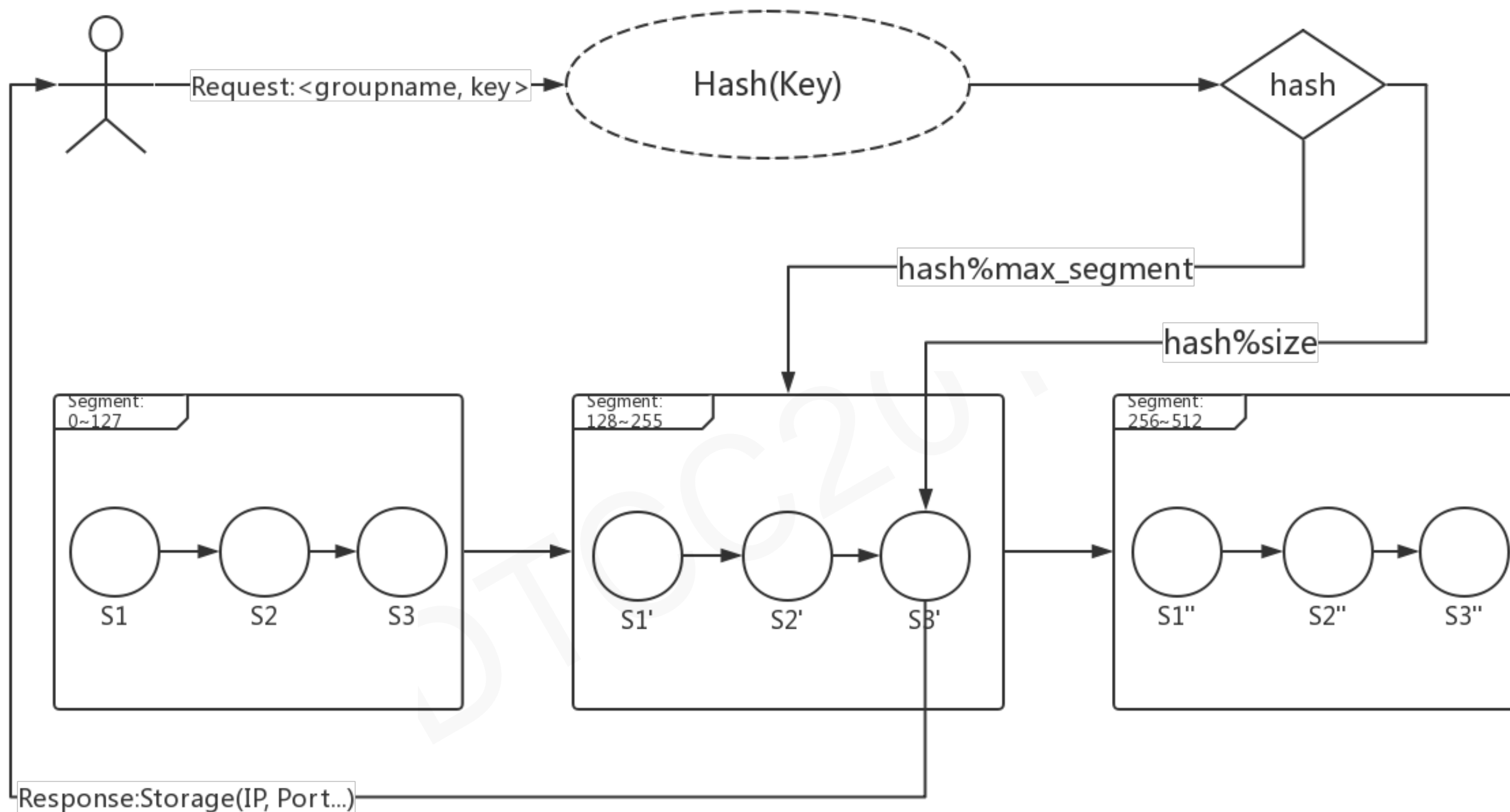
ITPUB

tracker



- **业务组**
按groupname索引
- **段区间组**
组内按segment分组
- **段同步组**
在段组内，主主同步

负载均衡算法-二次hash到加权两次hash的演变



存储

DTCC
2018

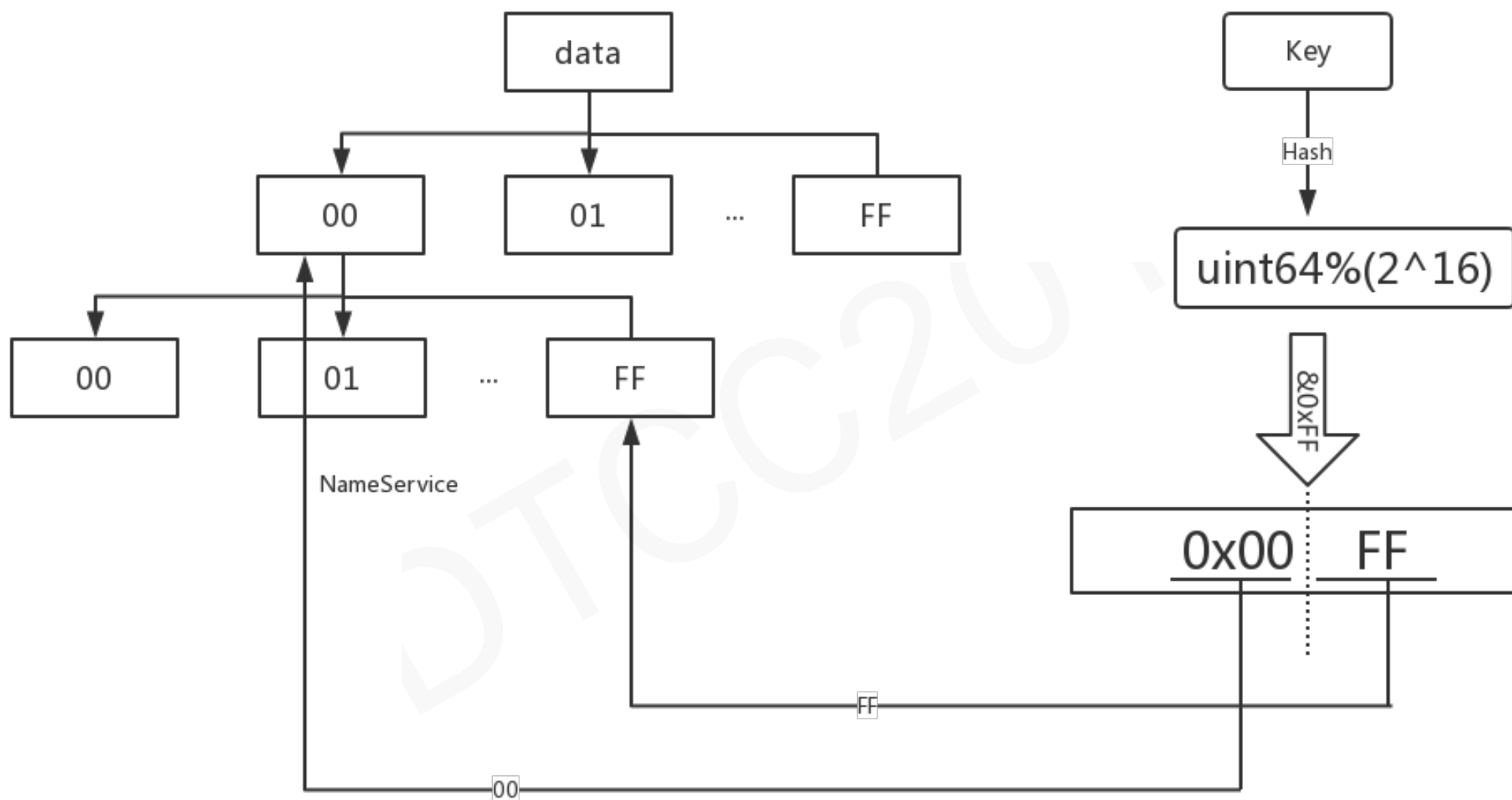
数领先机 智赢未来 (9)

IT168.com

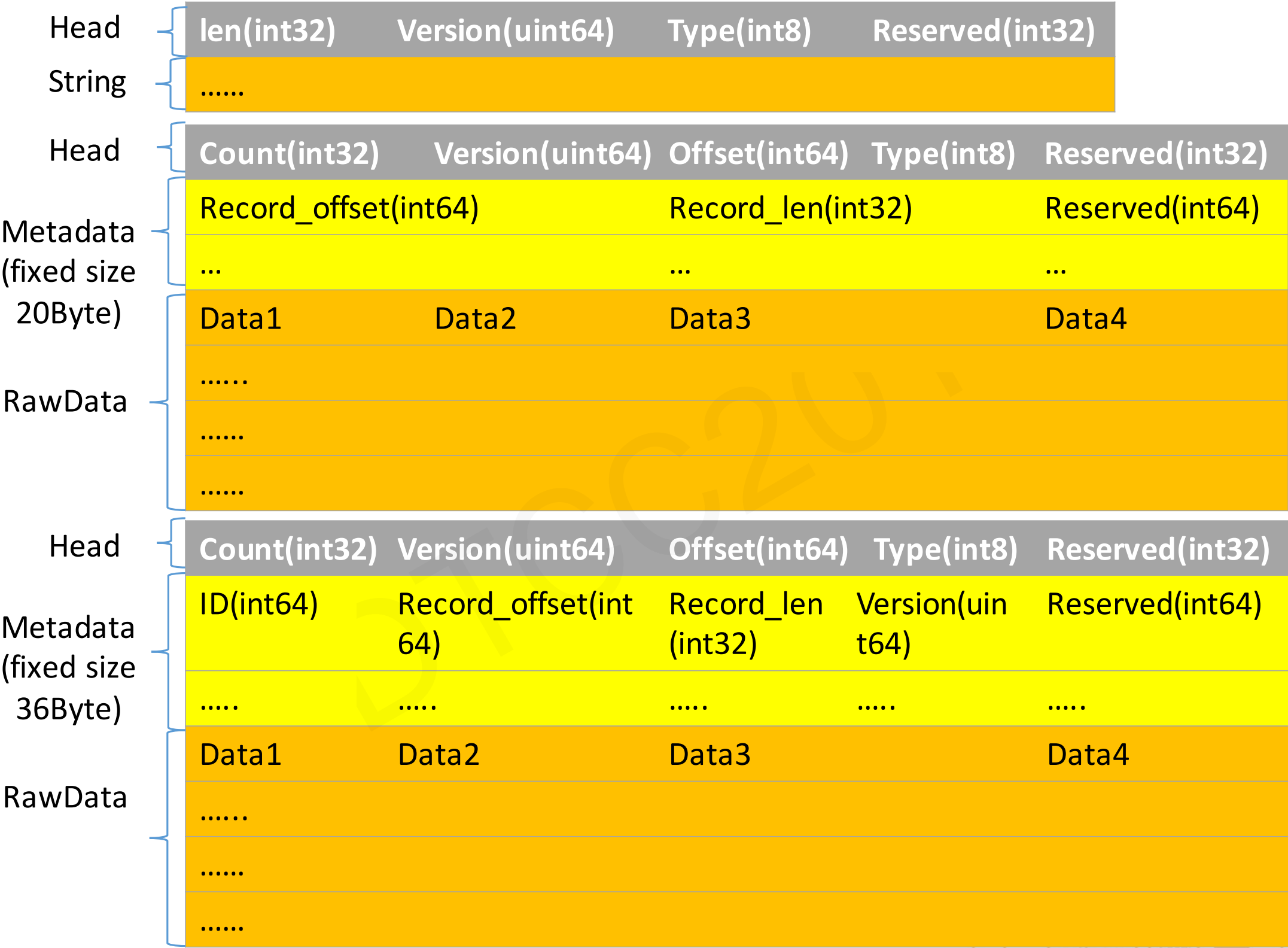
ChinaUnix

ITPUB

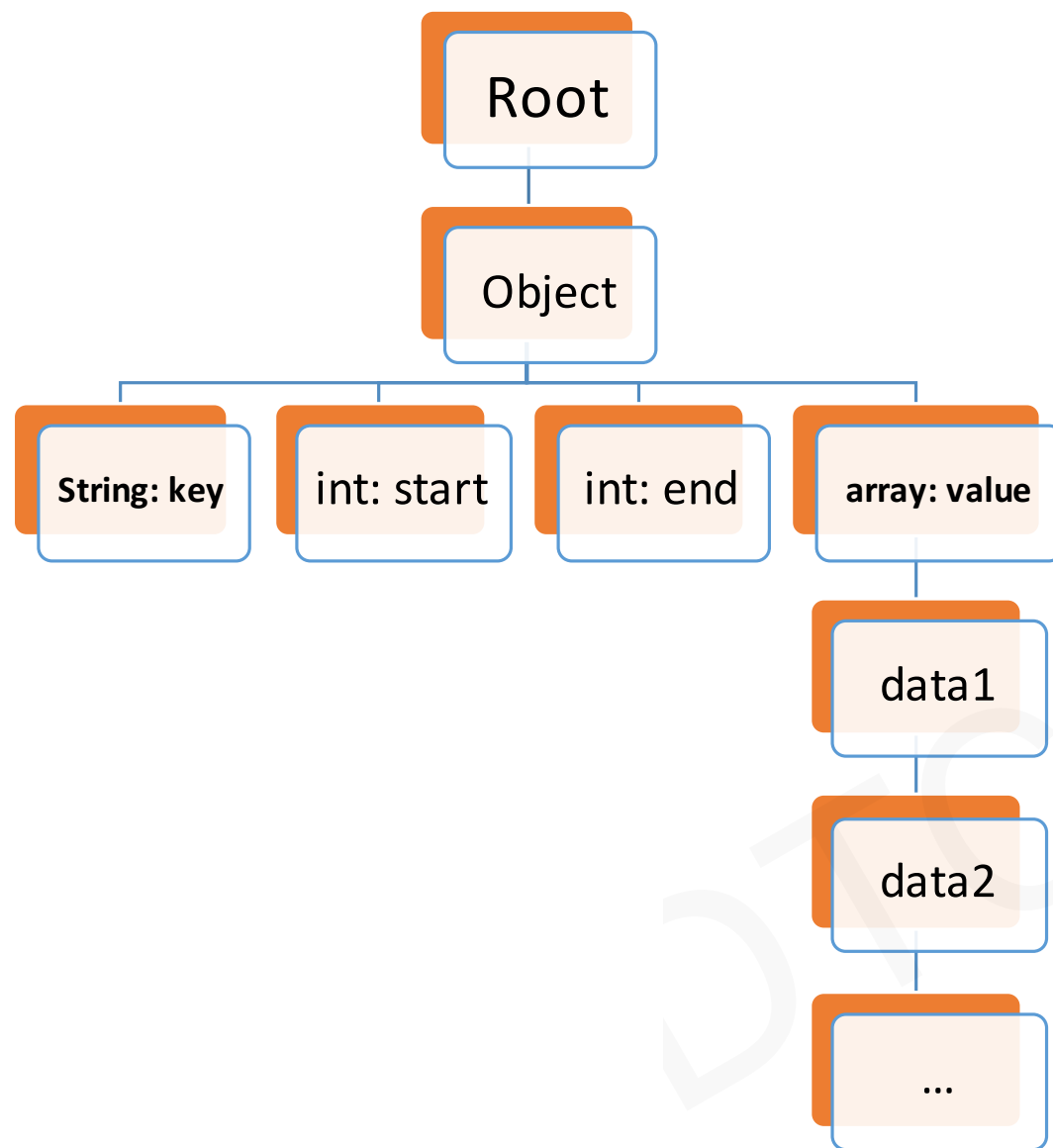
数据存储-路径映射



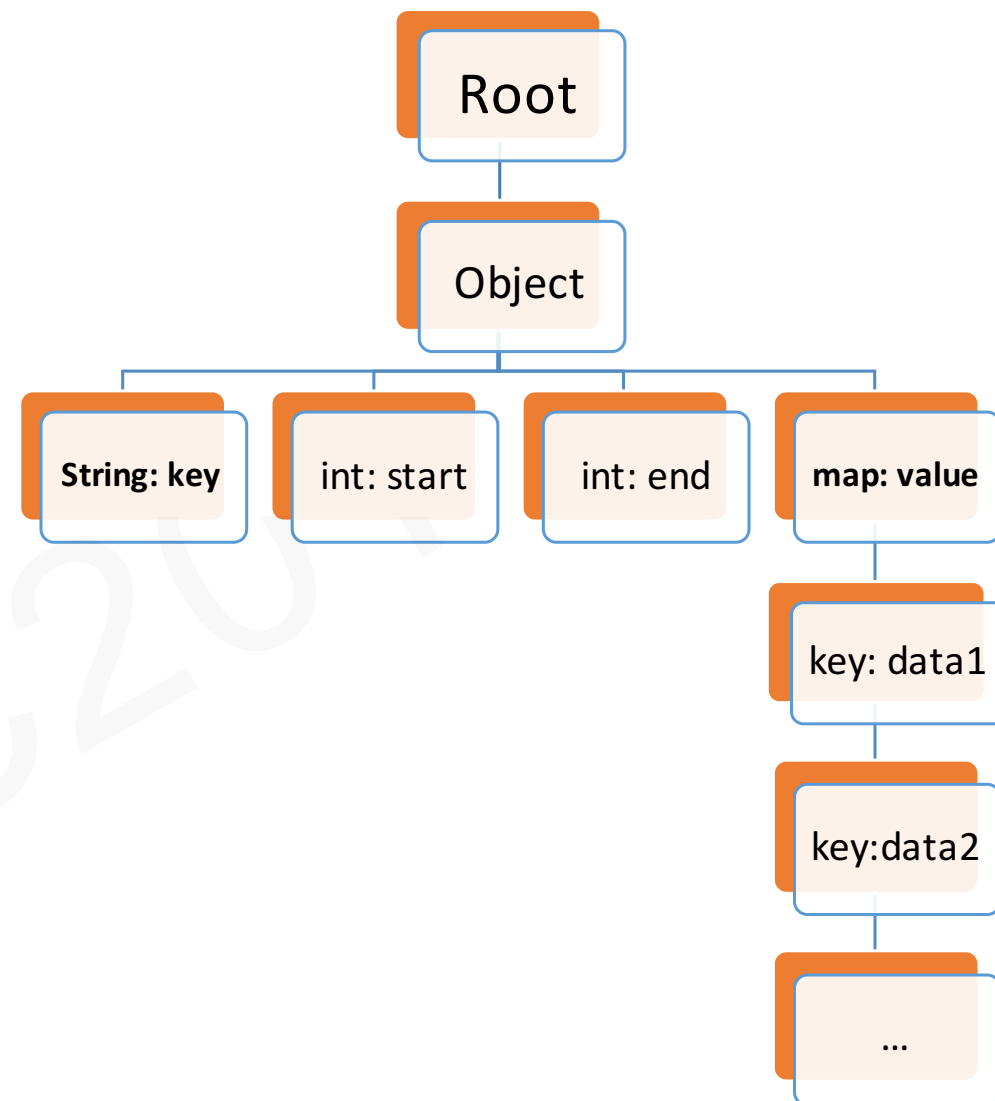
数据存储-文件存储



HMS对象



List



Map

API

操作类型	参数	返回状态	方法说明	支持类型
Put	Key, String	成功或失败	根据Key存储	String
Get	Key	List对象或失败	根据Key获取	String

操作类型	参数	返回状态	方法说明	支持类型
Put	Key, List	成功或失败	根据Key存储	List
Get	Key	List对象或失	根据Key获取	List
GetRange	Key, Start, End	[start, end)范围	根据Key获取	List

操作类型	参数	返回状态	方法说明	支持类型
MapPut	Key, Map	成功或失败	根据Key存储	Map
MapGet	Key	Map对象或失	根据Key获取	Map
MapUpdate	Key, Map	成功或失败	根据Key更新	Map
MapPutOrUpdate	Key, Map	成功或失败	根据Key存储	Map
MapGetById	Key, ID	指定对象或失	根据Key和ID	Map

同步

DTCC
2018

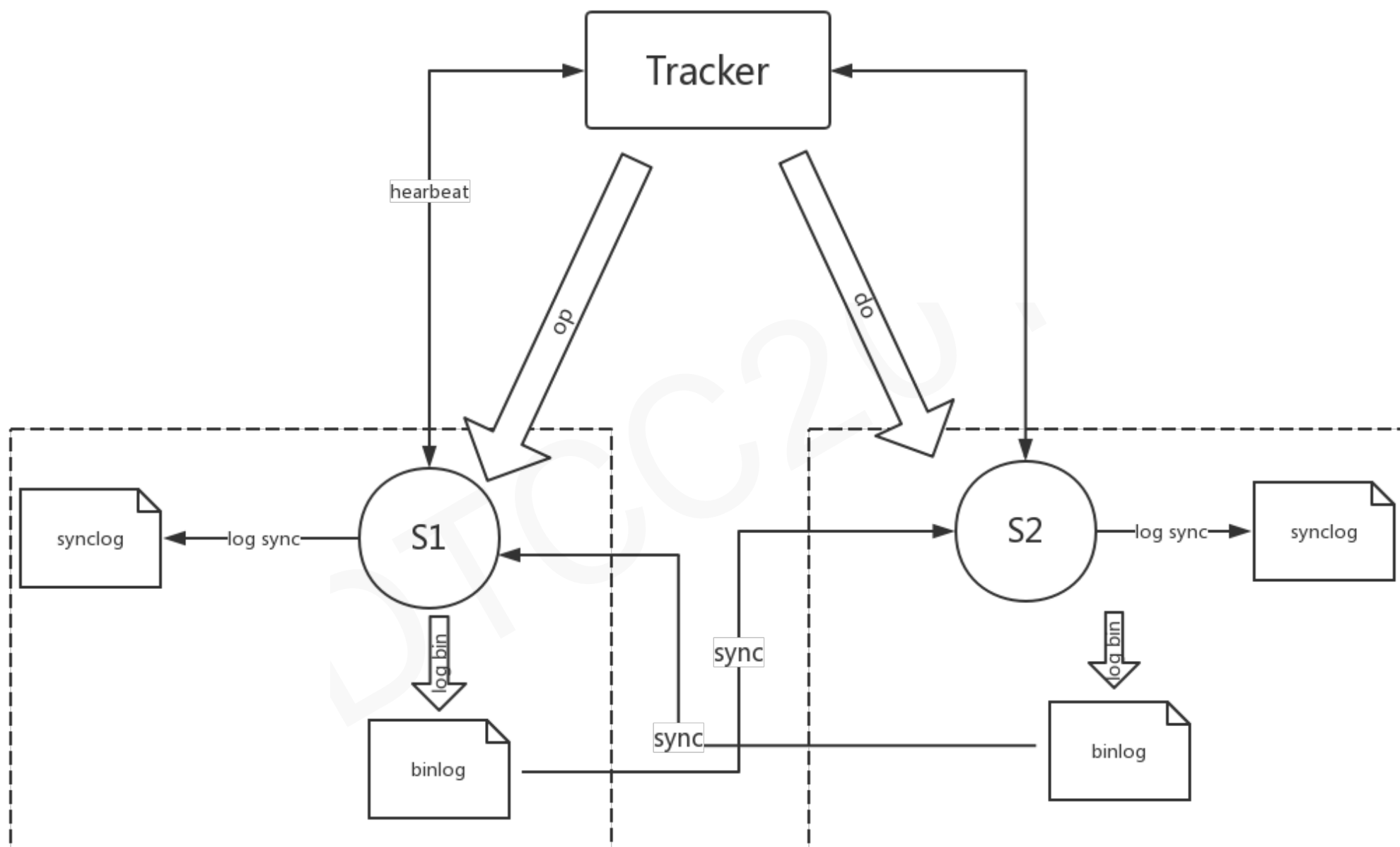
数领先机 智赢未来 (9)

IT168.com

ChinaUnix

ITPUB

同步架构



同步-binlog

- 操作类型
ADD
UPDATE
- 文件类型
L -> List
M -> Map
- Key
- ID(仅用于UPDATE)

操作	文件类型	Key	ID
ADD	L	Hello_List	
ADD	M	Hello_Map	
UPDATE	M	Hello_Map	0
UPDATE	M	Hello_Map	1
UPDATE	M	Hello_Map	2

同步复制状态

- 状态转移（控制简单，传输数据量大）

$$1 + 1 = 2 \Rightarrow 2$$

$$2 + 1 = 3 \Rightarrow 3$$

- 复制状态机（控制复杂，传输数据小）

$$1 + 1 = 2 \Rightarrow +1$$

$$2 + 1 = 3 \Rightarrow +1$$

性能测试

DTCC
2018

数领先机 智赢未来 (9)

IT168.com

ChinaUnix

ITPUB

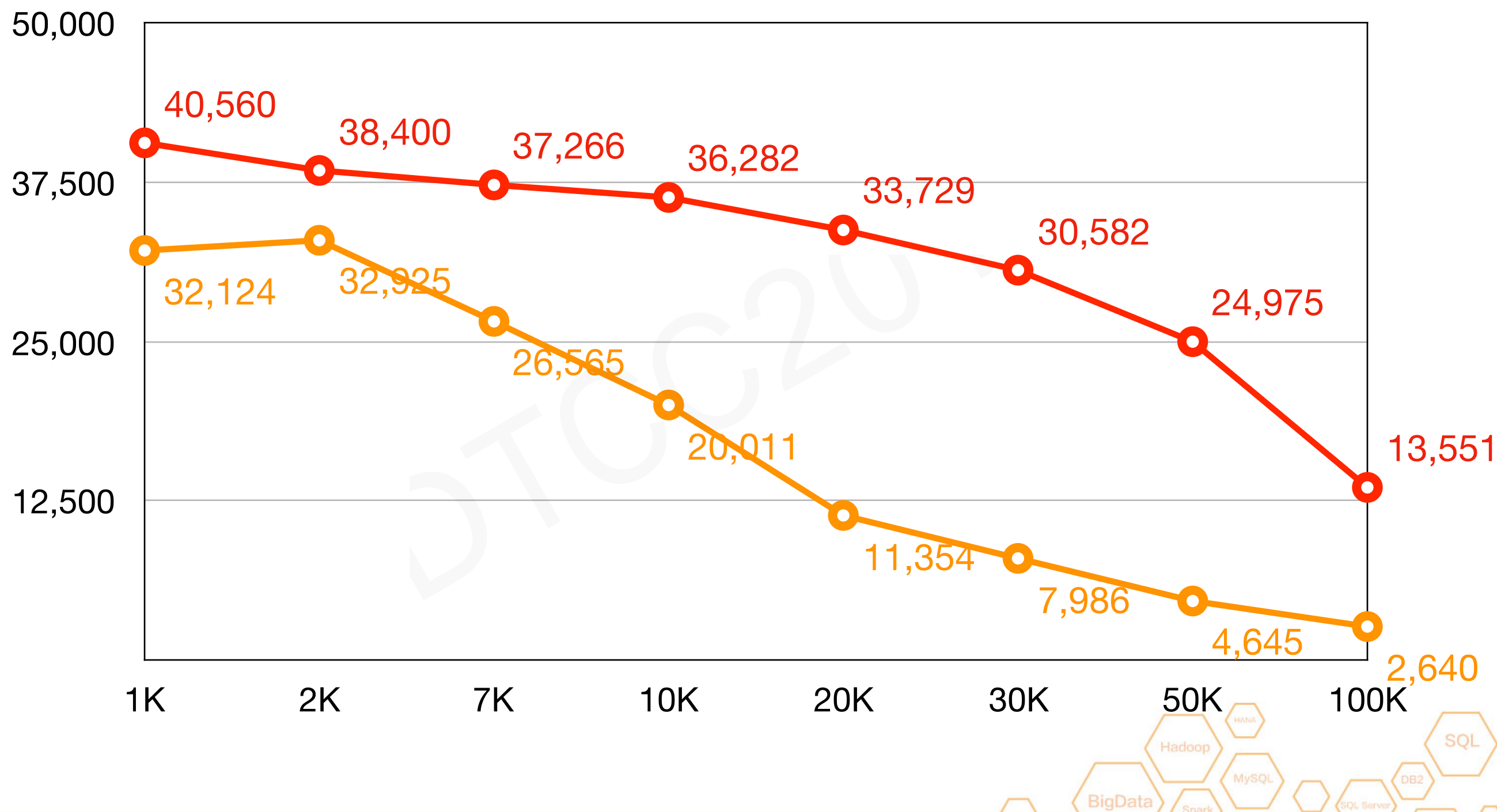
Lest性能测试

操作	平均长度/10k	平均耗时/10k	平均大小/10k	内存占用
Put	1007	84ms	245KB	20M
Get	1007	8ms	245KB	20M
MapPut	995	38ms	242KB	20M
MapGet	995	8ms	242KB	20M
MapGetByID	1	1ms	249Byte	20M



万兆+SSD 千兆+SATA

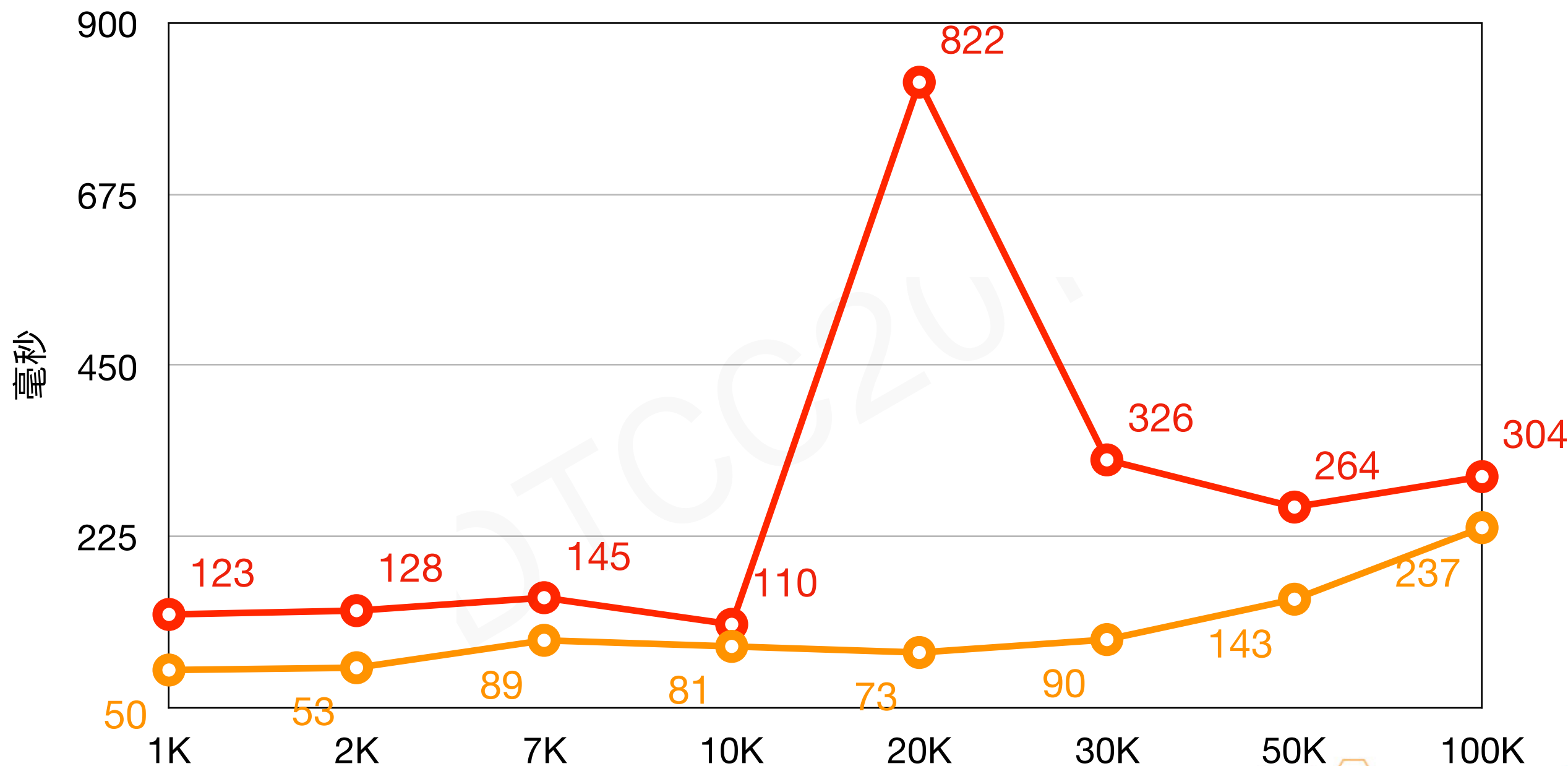
处理请求正确响应数



万兆+SSD (最大)

千兆+SATA (最大)

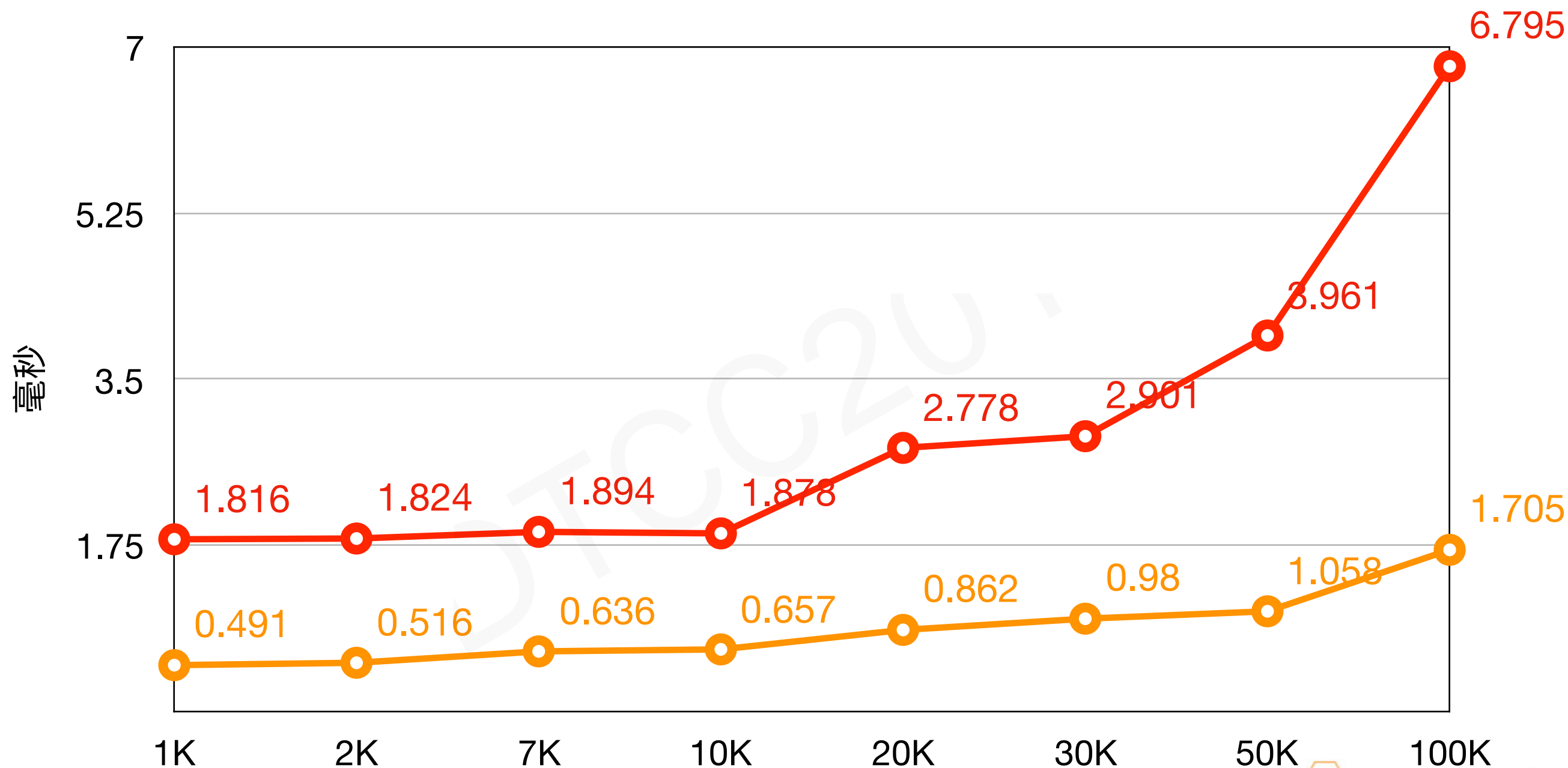
最大响应时间



万兆+SSD (最小)

千兆+SATA (最小)

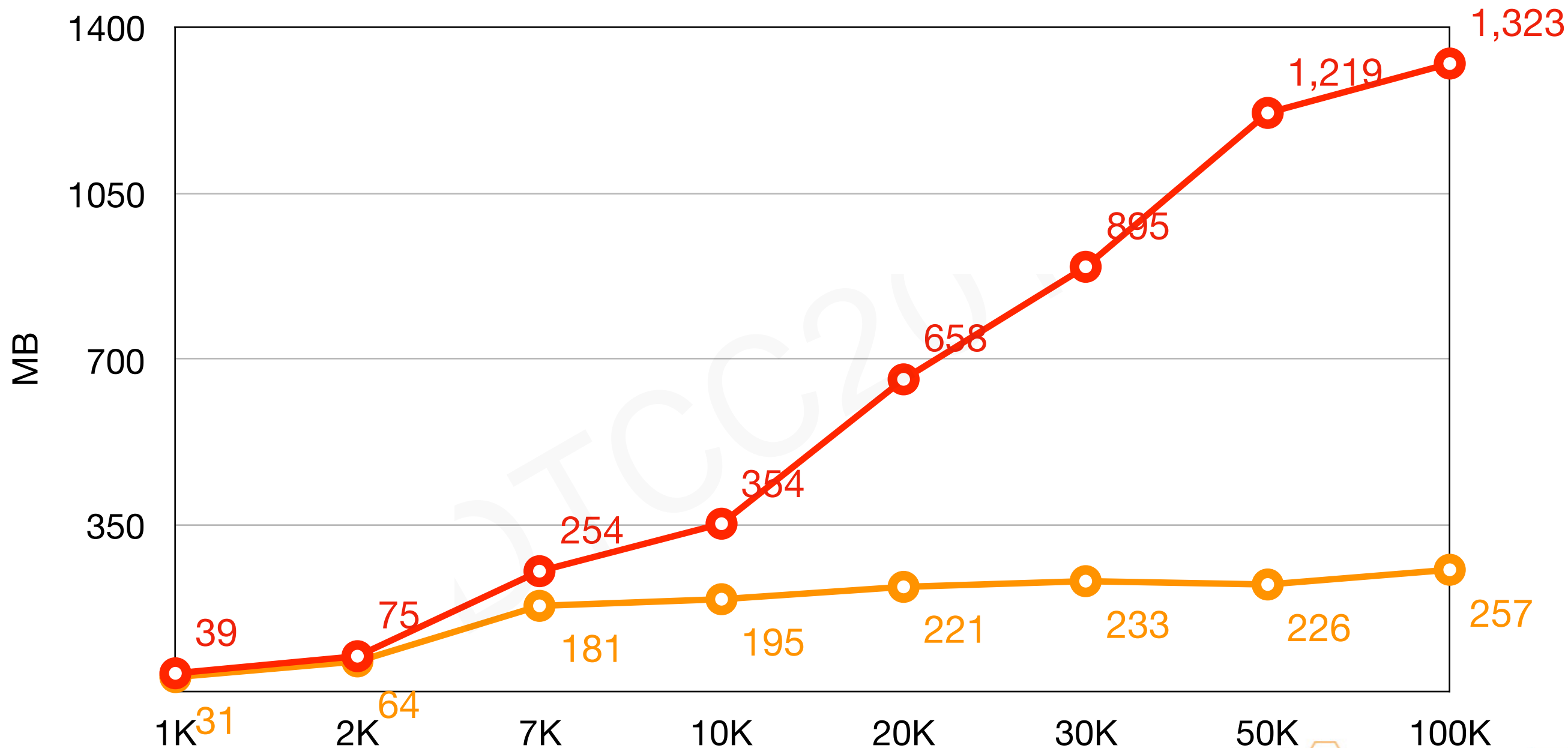
最小响应时间



万兆+SSD

千兆+SATA

传输量

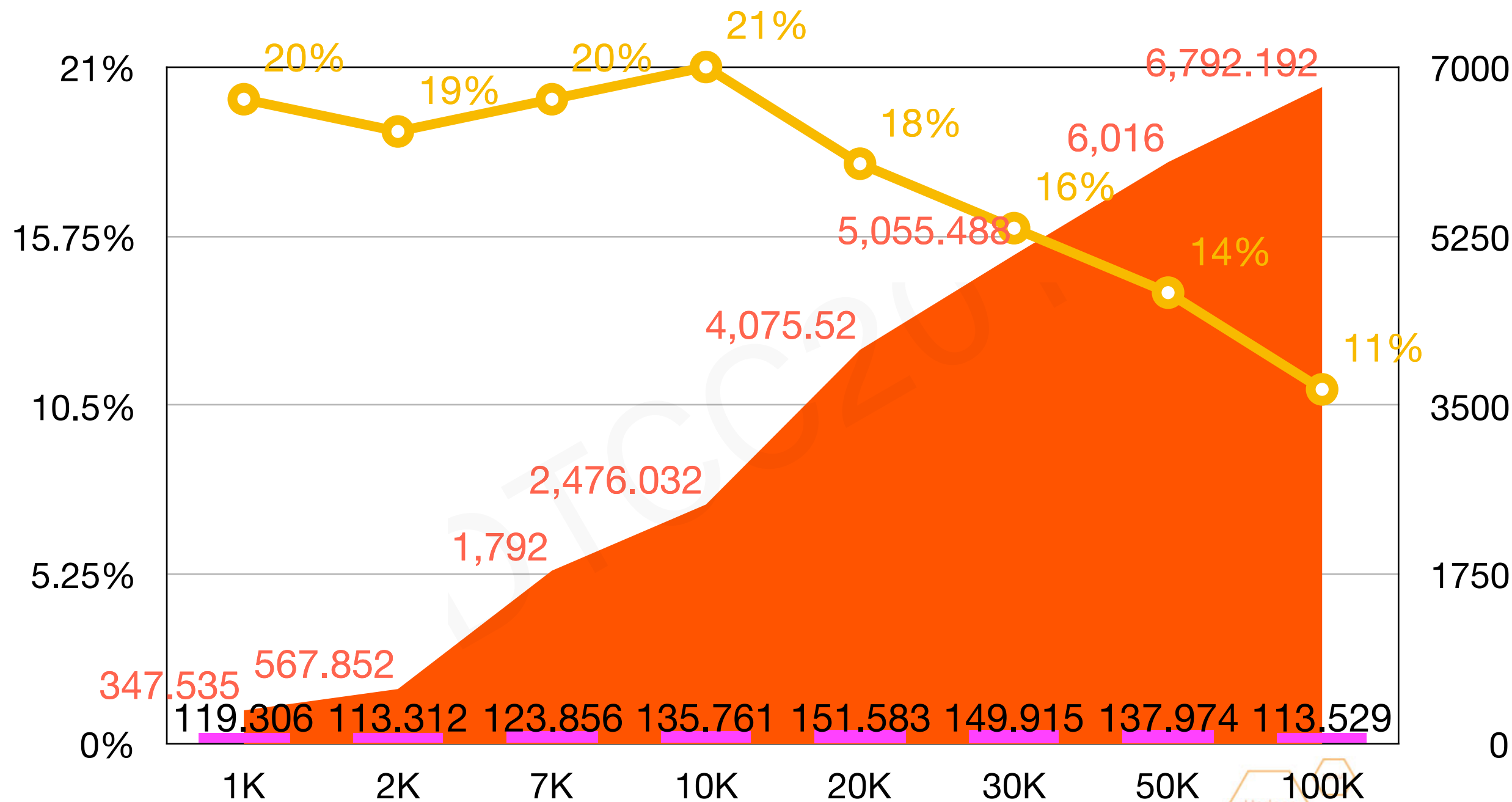


CPU

网卡入

网卡出

SSD服务器服务器压力



Lest优劣势

- 吃磁盘、不吃内存
- 支持快速恢复
- 主主同步有利于负载均衡
- 轻量二进制存储与传输协议
- 对象树方便、快速
- 精确版本控制
- 负载均衡不受客户端控制
- 随机读写问题、推荐SSD
- 头部元信息扩展难
- 数据单条同步可改为批量
- 增加热数据LRU内存缓存

THANKS



欢迎关注我的公众号：94geek的大嘴

DTCC
2018

数领先机 智赢未来 (9)

IT168.com

ChinaUnix

ITPUB



讲师申请

联系电话（微信号）：18612470168

关注“ITPUB”更多
技术干货等你来拿~

与百度外卖、京东、魅族等先后合作系列分享活动



让学习更简单

微学堂是以ChinaUnix、ITPUB所组建的微信群为载体，定期邀请嘉宾对热点话题、技术难题、新产品发布等进行移动端的在线直播活动。

截至目前，累计举办活动期数60+，参与人次40000+。

ITPUB学院

ITPUB学院是盛拓传媒IT168企业事业部（ITPUB）旗下
企业级在线学习咨询平台
历经18年技术社区平台发展
汇聚5000万技术用户
紧随企业一线IT技术需求
打造全方式技术培训与技术咨询咨询服务
提供包括企业应用方案培训咨询（包括企业内训）
个人实战技能培训（包括认证培训）
在内的全方位IT技术培训咨询服务

ITPUB学院讲师均来自于企业
一些工程师、架构师、技术经理和CTO
大会演讲专家1800+
社区版主和博客专家500+

培训特色

无限次免费播放
随时随地在线观看
碎片化时间集中学习
聚焦知识点详细解读
讲师在线答疑
强大的技术人脉圈

八大课程体系

基础架构设计与建设
大数据平台
应用架构设计与开发
系统运维与数据库
传统企业数字化转型
人工智能
区块链
移动开发与SEO



联系我们

联系人：黄老师
电话：010-59127187
邮箱：edu@itpub.net
网址：edu.itpub.net
培训微信号：18500940168