



# 云HBase之时序引擎

OpenTSDB介绍&压缩优化

—— 郭泽晖 阿里云技术专家



# 目录 / Contents

01

OpenTSDB介绍

02

OpenTSDB常见问题

03

OpenTSDB压缩优化

04

云OpenTSDB使用模式



# 01

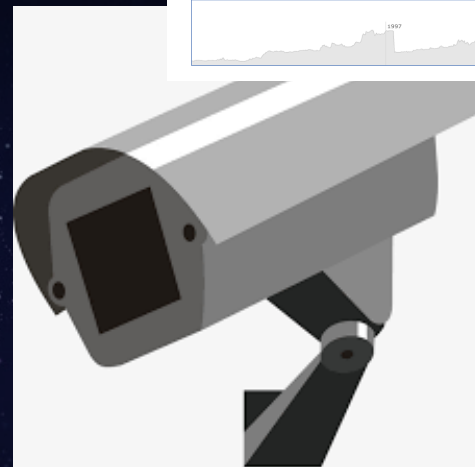
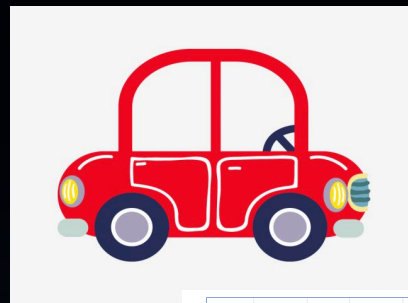
## OpenTSDB介绍



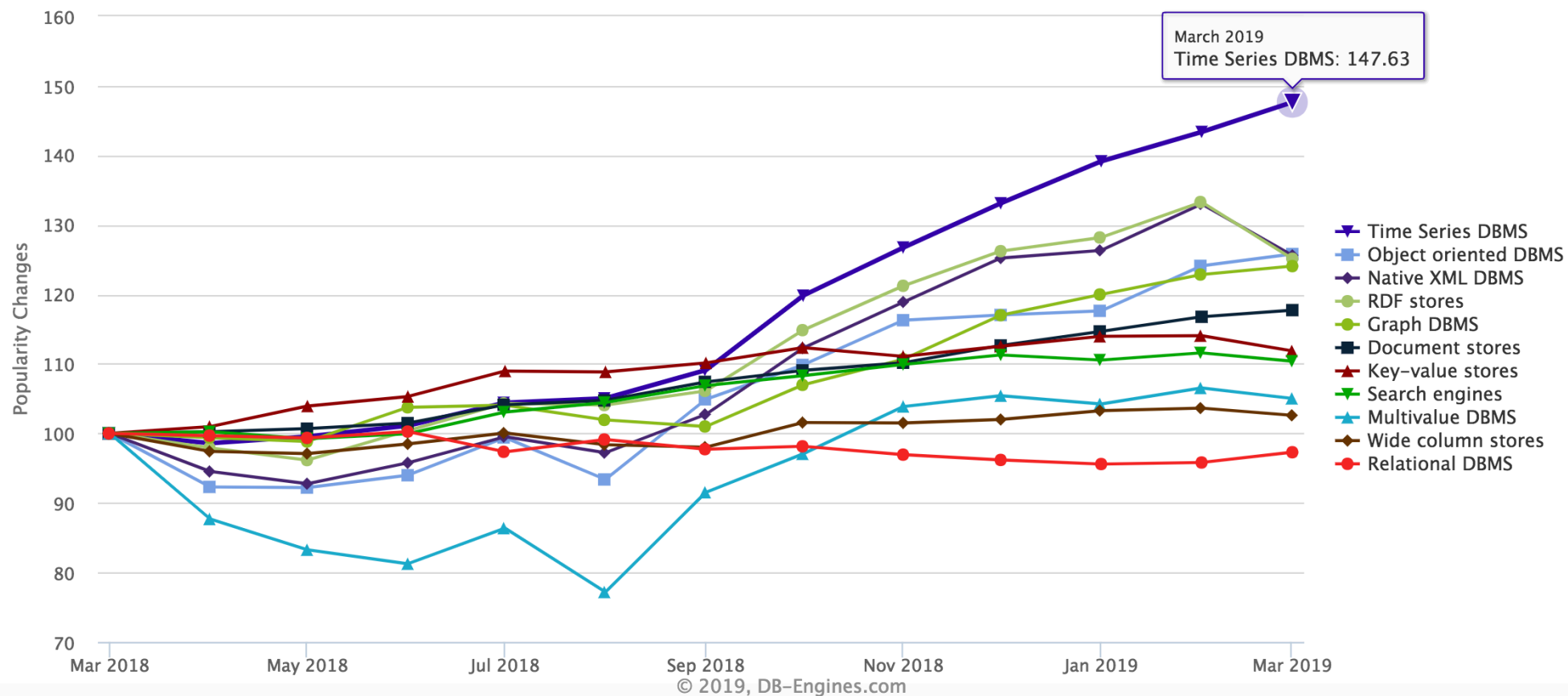
OpenTSDB是一款基于HBase构建的时序数据库

### 时序数据特点：

- 1.数值类型（整数，浮点）
- 2.数据通常按照时间顺序抵达，源源不断
- 3.基本不会更新数据
- 4.写多读少



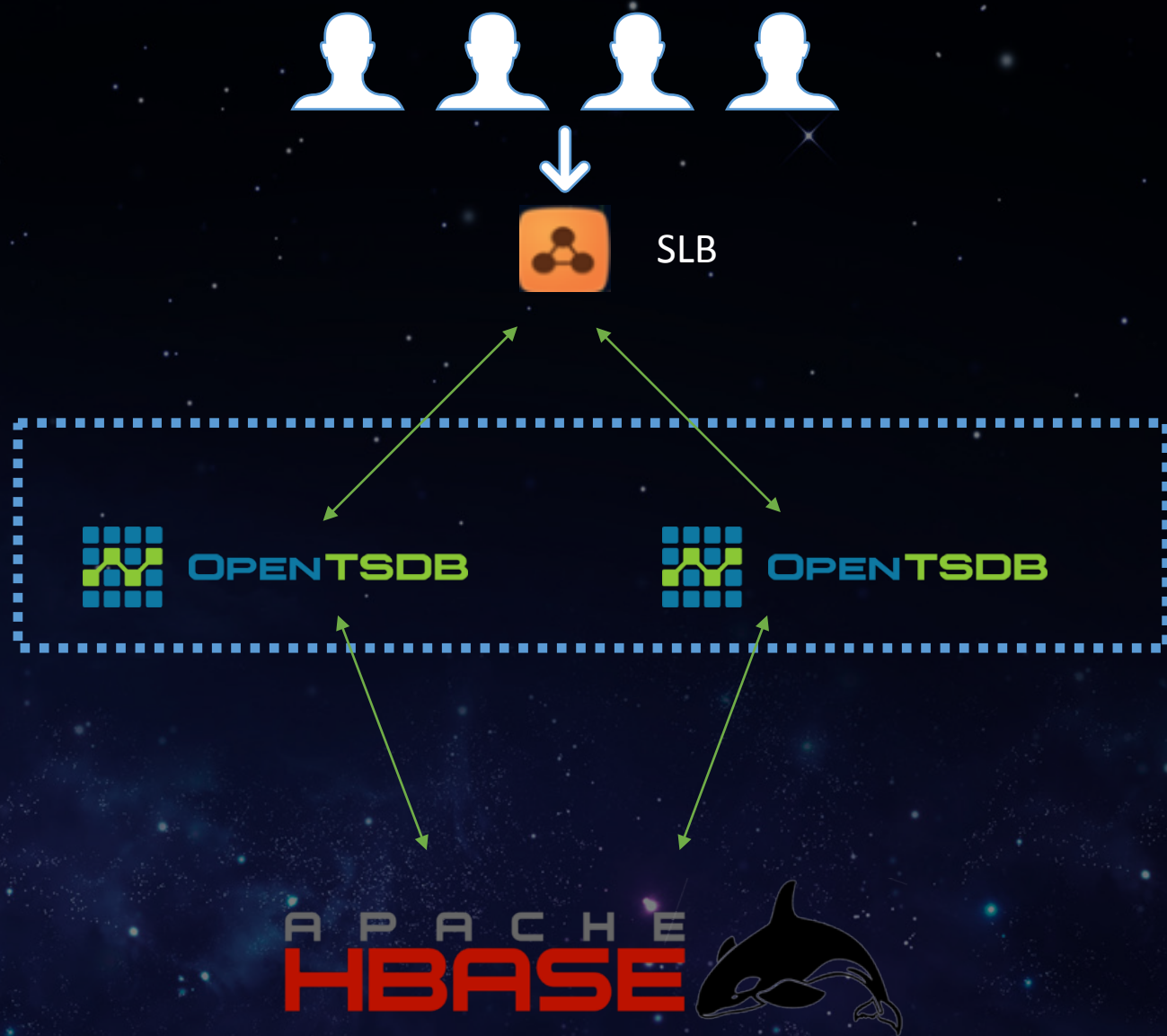
Trend of the last 12 months





# 1 OpenTSDB介绍

## 部署架构

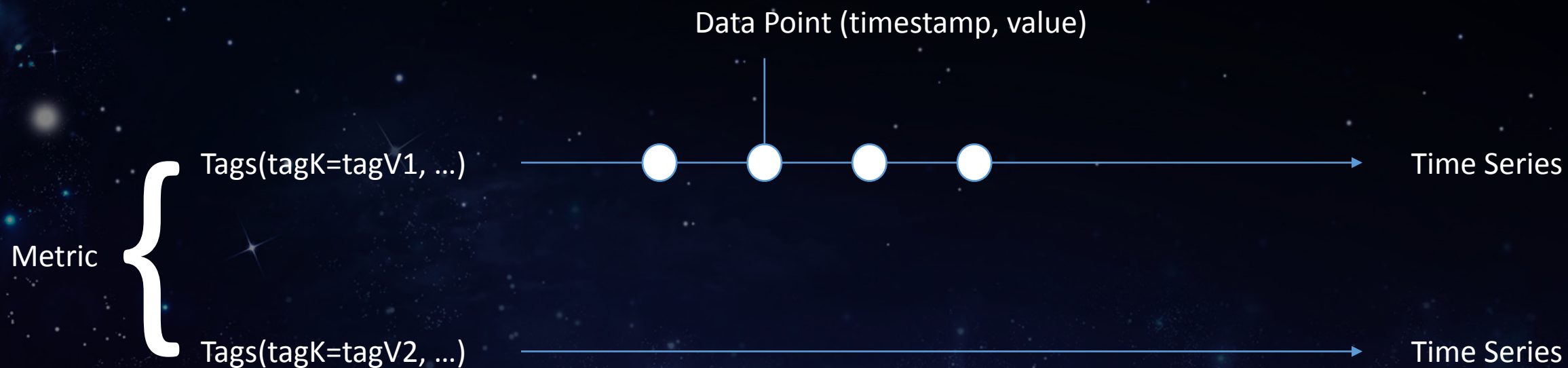






## 1 OpenTSDB介绍

### 定义



时间精度支持：秒/毫秒



## 1 OpenTSDB介绍

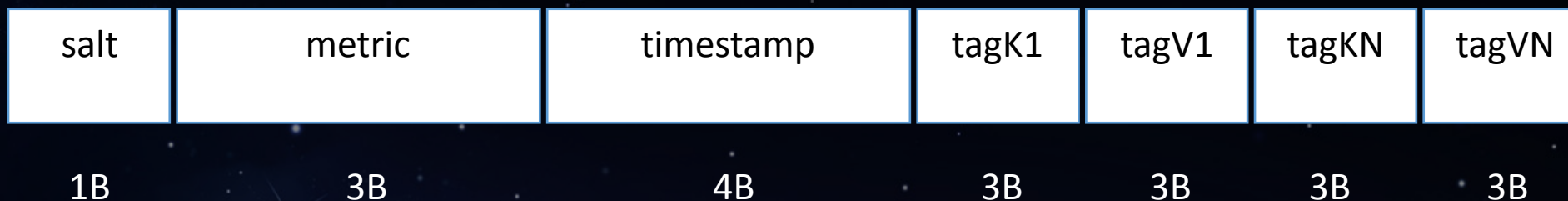
# 数据表设计



每条时间线每小时数据存一行，数据点只记录和这个小时起始时间戳偏移量，以减少存储空间。



## RowKey数据格式



salt: 打散同一metric不同时间线的热点

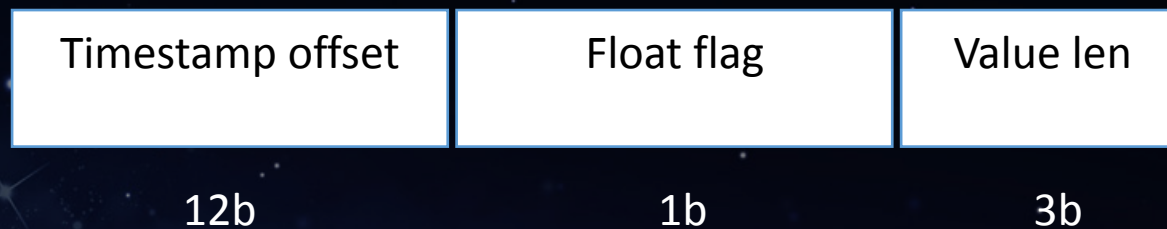
metric, tagK, tagV: 实际存储的是字符串对应的UID (在tsdb-uid表中)

timestamp: 每小时数据存在一行, 记录的是每小时整点秒级时间戳

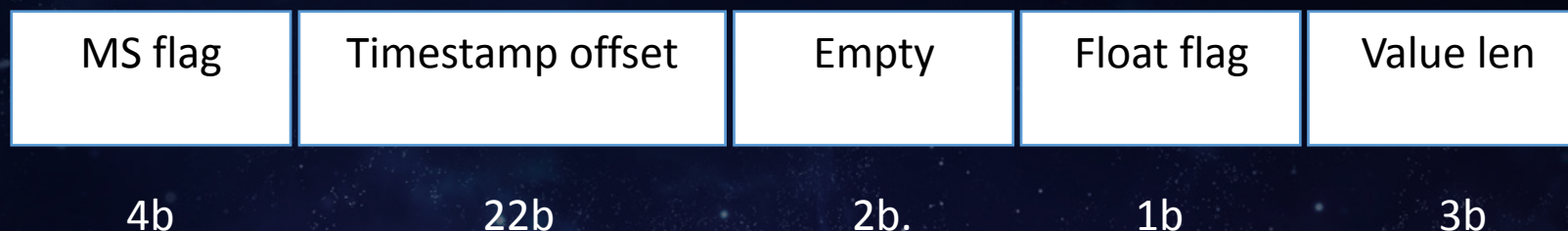
Tips: metric和tag的长度默认是3个字节, 即最多只能分配 16777216个UID。可以通过这些参数调整: `tsd.storage.uid.width.metric`, `tsd.storage.uid.width.tagk`, `tsd.storage.uid.width.tagv`。注意集群已经写过数据后就无法修改。

## Column qualifier 数据格式

秒级(2 Bytes)



毫秒级(4 Bytes)





# 02

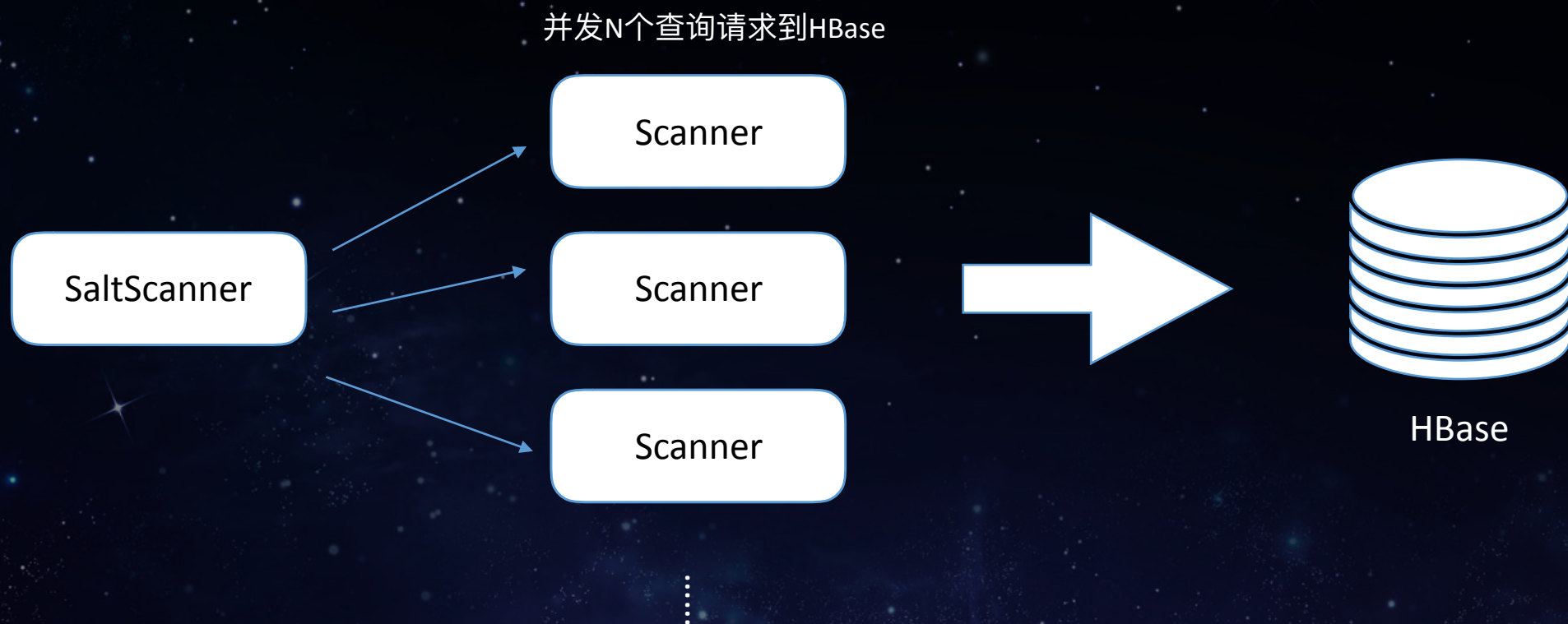
## OpenTSDB常见问题





## 2 OpenTSDB常见问题

### 写入热点与查询并发的权衡



Tips: 并发度取决于tsd.storage.salt.buckets（默认20），设置后不能修改。  
建议不要多于RS数量，并发太高容易打爆HBase。

## Tags的笛卡尔积影响查询效率

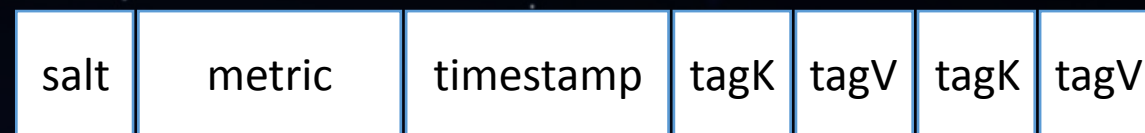
假设一个监控机器CPU情况的metric=host.cpu, 有3个tag: region, host, core

region: 机房区域, 假设有10个

host: 主机名, 假设每个机房有100台

core: cpu单核编号, 假设每个主机有8个核(0-7)

Tag无索引



salt	metric	timestamp	tagK	tagV	tagK	tagV
------	--------	-----------	------	------	------	------

那么这个metric会包含 $\text{region}(10) \times \text{host}(100) \times \text{core}(8) = 8000$  条时间线

当你查询最近1小时{**region=HangZhou**, **host=30.43.111.255**, **core=0**}数据的时候,  
实际上得访问**8000**条时间线的最近1小时数据。



## 一些查询的其他坑

- 1.数据点要全部攒到TSDB内存后才会开始聚合处理，容易OOM，不是流式处理
- 2.一个大查询只会落在一台TSDB上被处理，不能分布式处理
- 3.即使查最近15分钟数据，也会遍历最近1小时数据





2

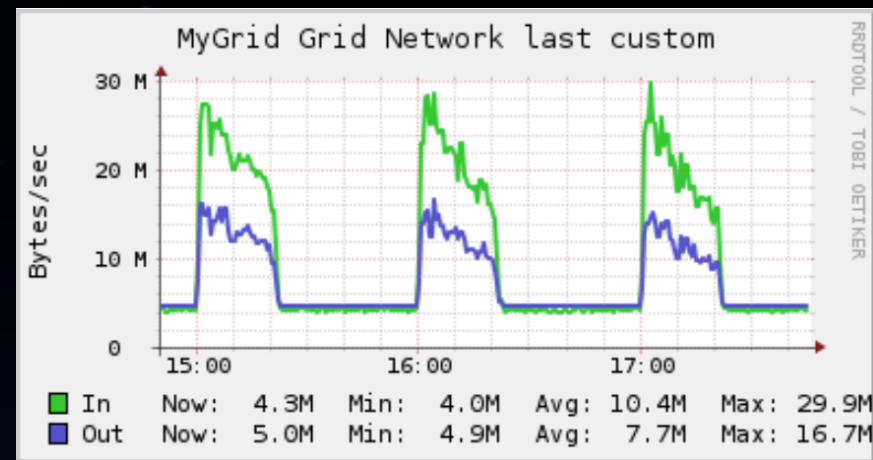
## OpenTSDB常见问题

OpenTSDB压缩问题：整点对HBase产生流量冲击

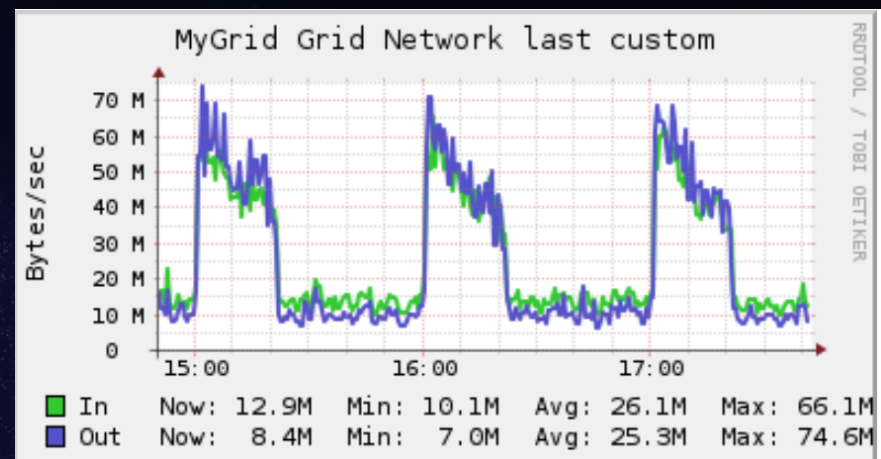
整点OpenTSDB对数据进行压缩会对HBase发起大量IO请求

导致长时间的数倍流量波峰，很容易打爆HBase

这样需要更高的规格抗波峰



OpenTSDB集群网络



HBase集群网络



## 2 OpenTSDB常见问题

### OpenTSDB压缩问题：整点对HBase产生流量冲击

OpenTSDB压缩原理是将离散的KV合成一个大KV，减少重复字段。

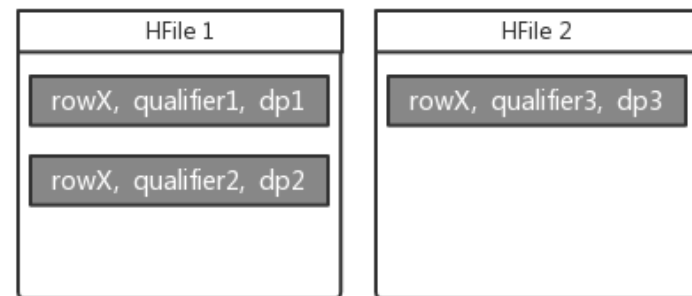
所以在整点会有读-压缩-重写-删  
这样一波流量产生。

OpenTSDB每小时的所有时序数据点都存在HBase的一行中。一行有许多列，每列代表一个时序的数据点(DataPoint, 记做dp)。

rowKey	qualifier1	qualifier2	qualifier3	qualifierN
rowX ( 时序行键, 某一小时的行键 )	dp1	dp2	dp3	dpN

Tips: 由于所有列都属于一个列族，所以图中省略了列族的概念

数据在HBase中的逻辑视图



数据在HBase中的物理视图

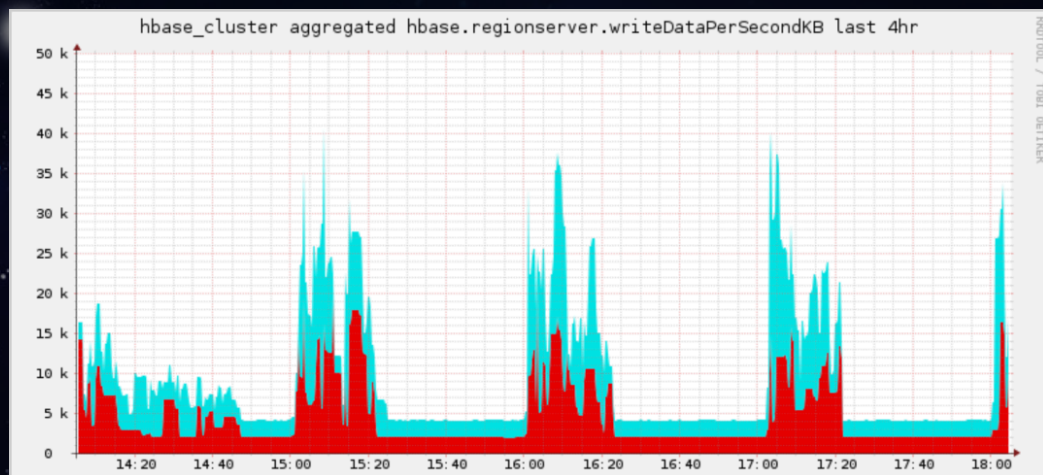




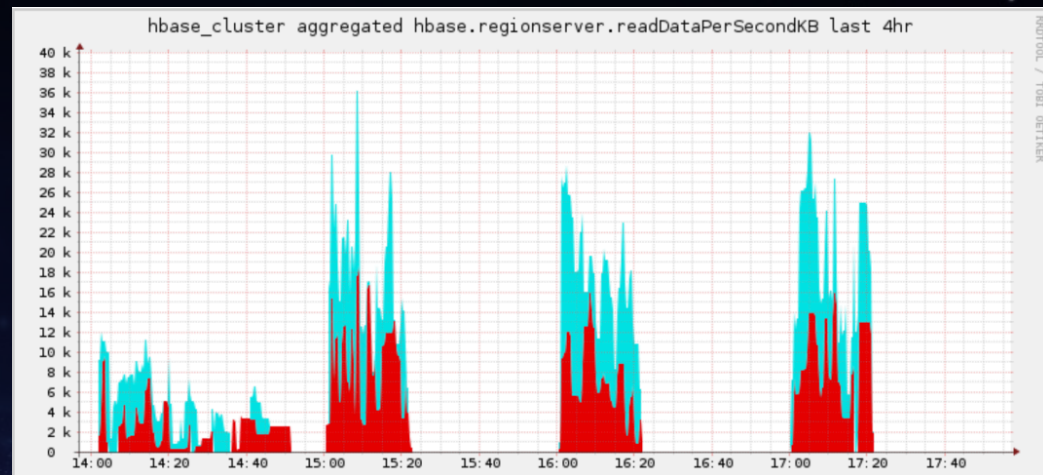
## 2 OpenTSDB常见问题

### OpenTSDB压缩问题：整点对HBase产生流量冲击

测试集群只做写操作，通过HBase具体的性能指标能观测到整点读写的波峰



HBase写请求流量



HBase读请求流量





# 03

## OpenTSDB压缩优化

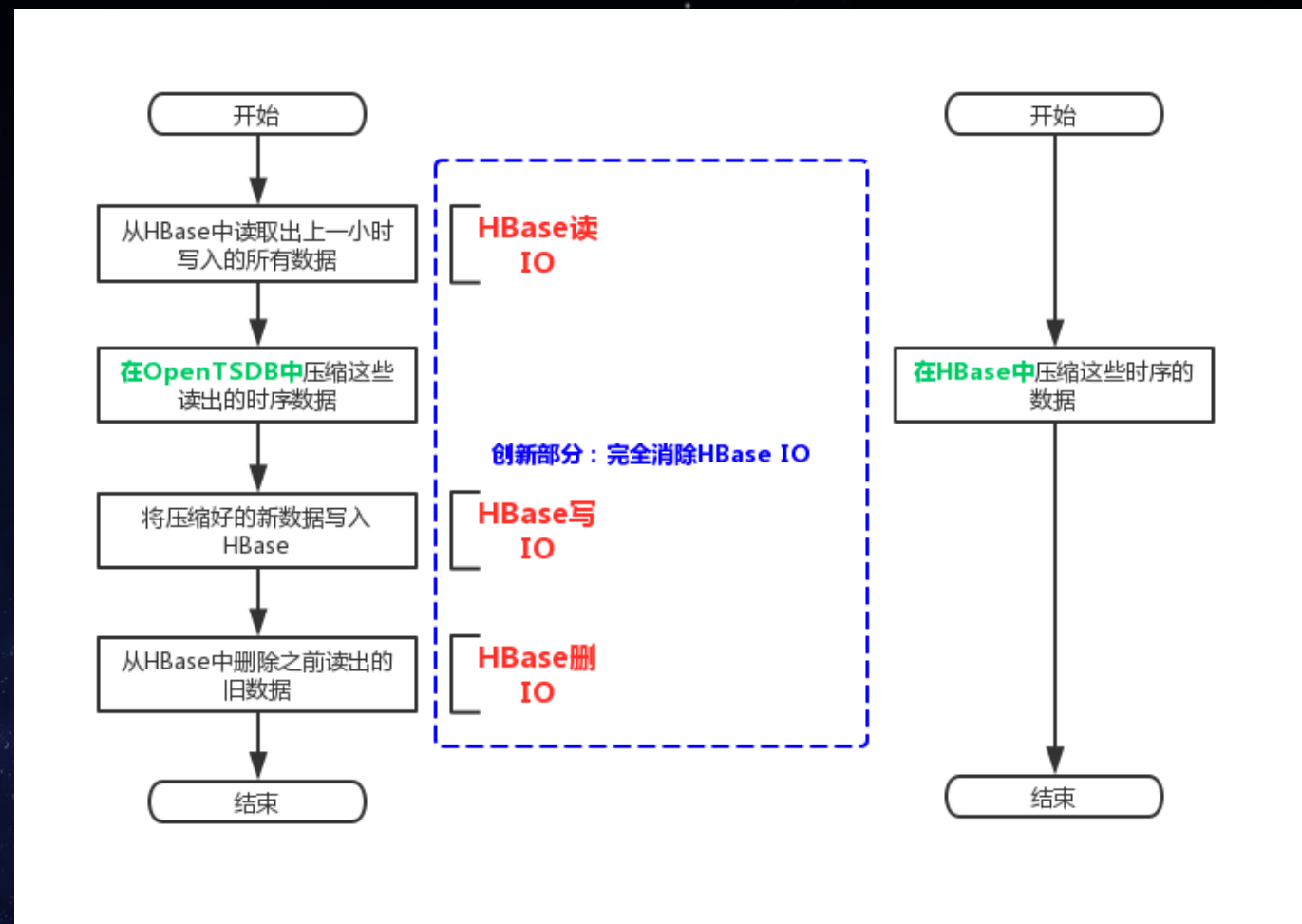


### 3 OpenTSDB压缩优化

## 云HBase团队针对OpenTSDB压缩问题的优化方案

将OpenTSDB压缩过程下沉到底层HBase

复用HBase压缩过程中的流量，本身HBase压缩就会对数据读写，在这时候做处理

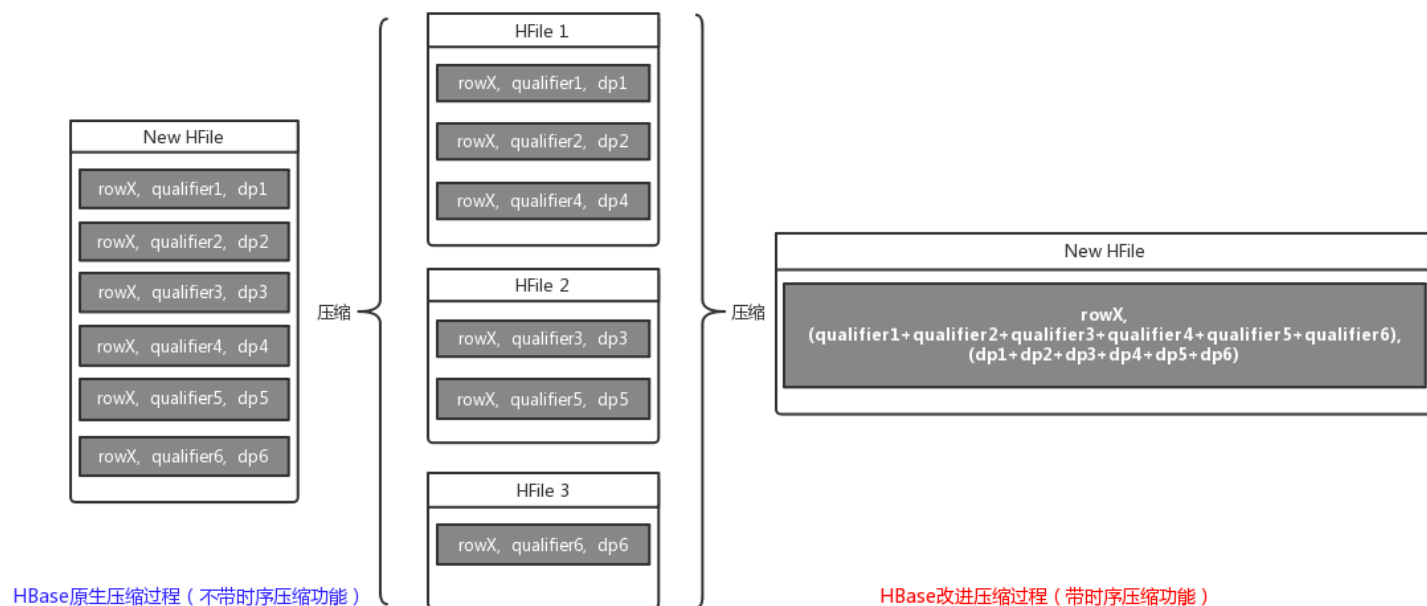




### 3 OpenTSDB压缩优化

## 云HBase团队针对OpenTSDB压缩问题的优化方案

**HBase压缩过程原理：**  
在合并HFile的时候将KV按  
OpenTSDB数据格式合并



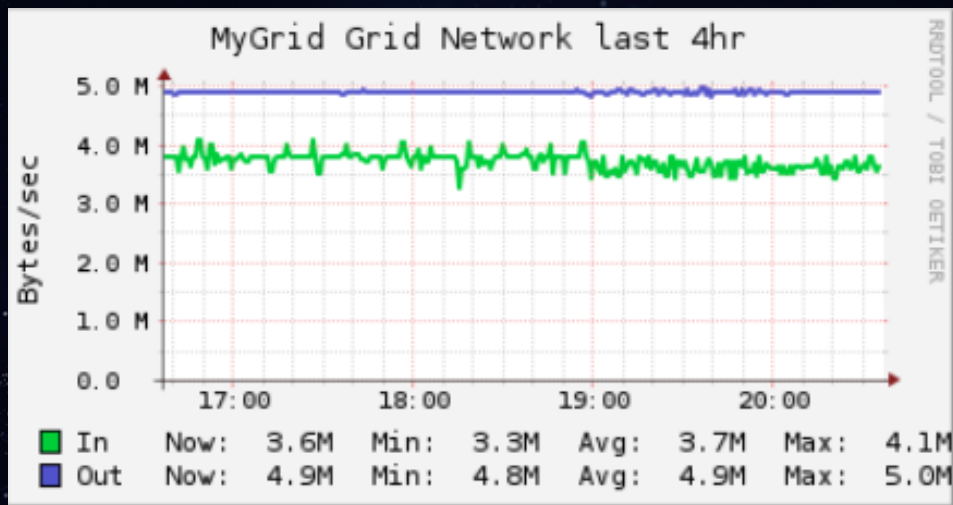




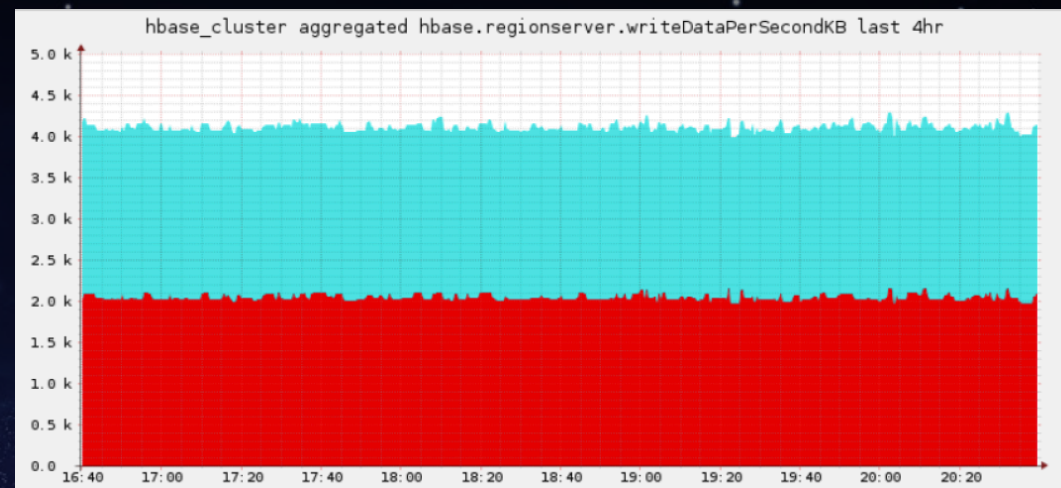
### 3 OpenTSDB压缩优化

## 云HBase团队针对OpenTSDB压缩问题的优化方案

测试集群只做写操作，完全消除对HBase的流量波峰



OpenTSDB网络



HBase写入流量



# 04

## 云OpenTSDB使用模式

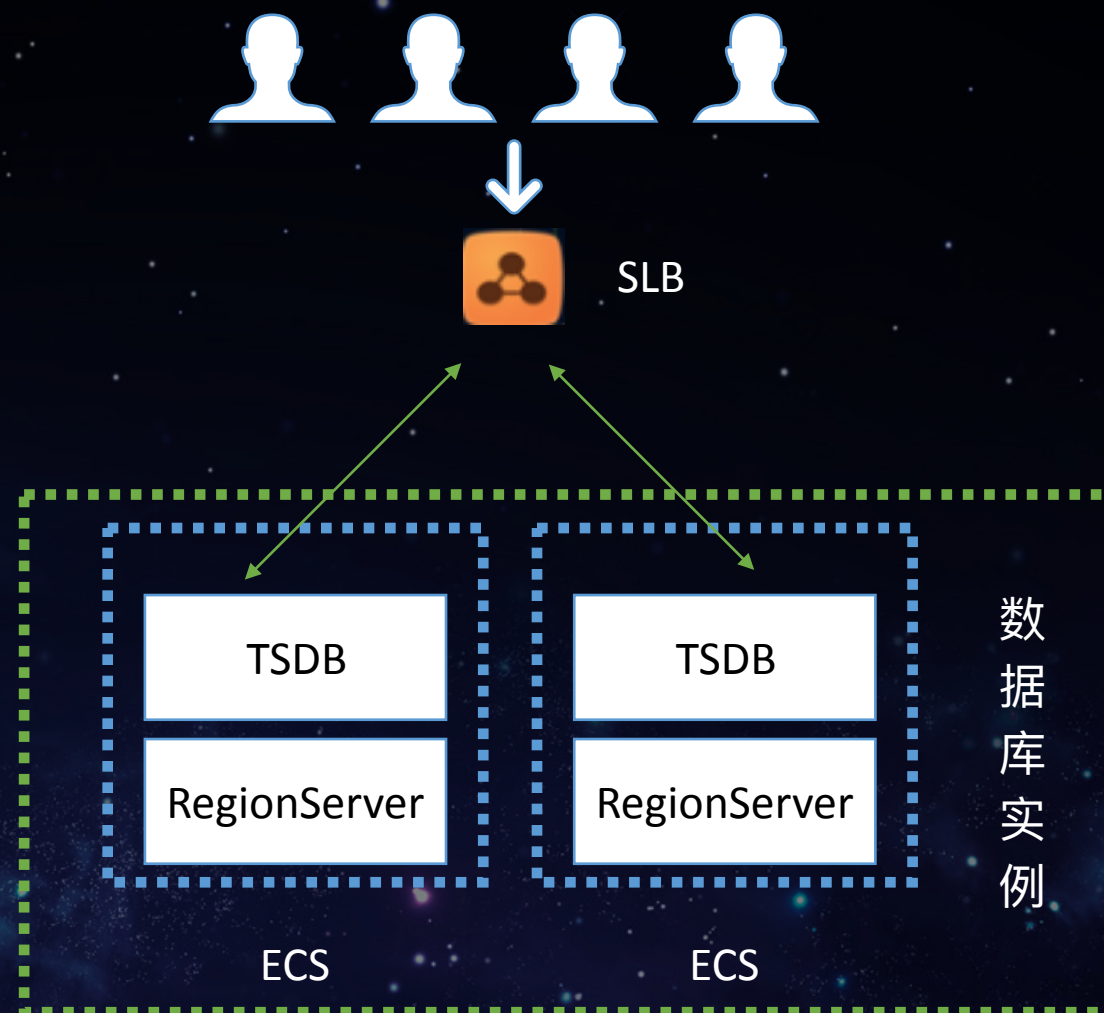


4

# 云OpenTSDB使用模式

**独享模式：** 完全独立部署的时序数据库

适合时序业务较重，需要分离部署



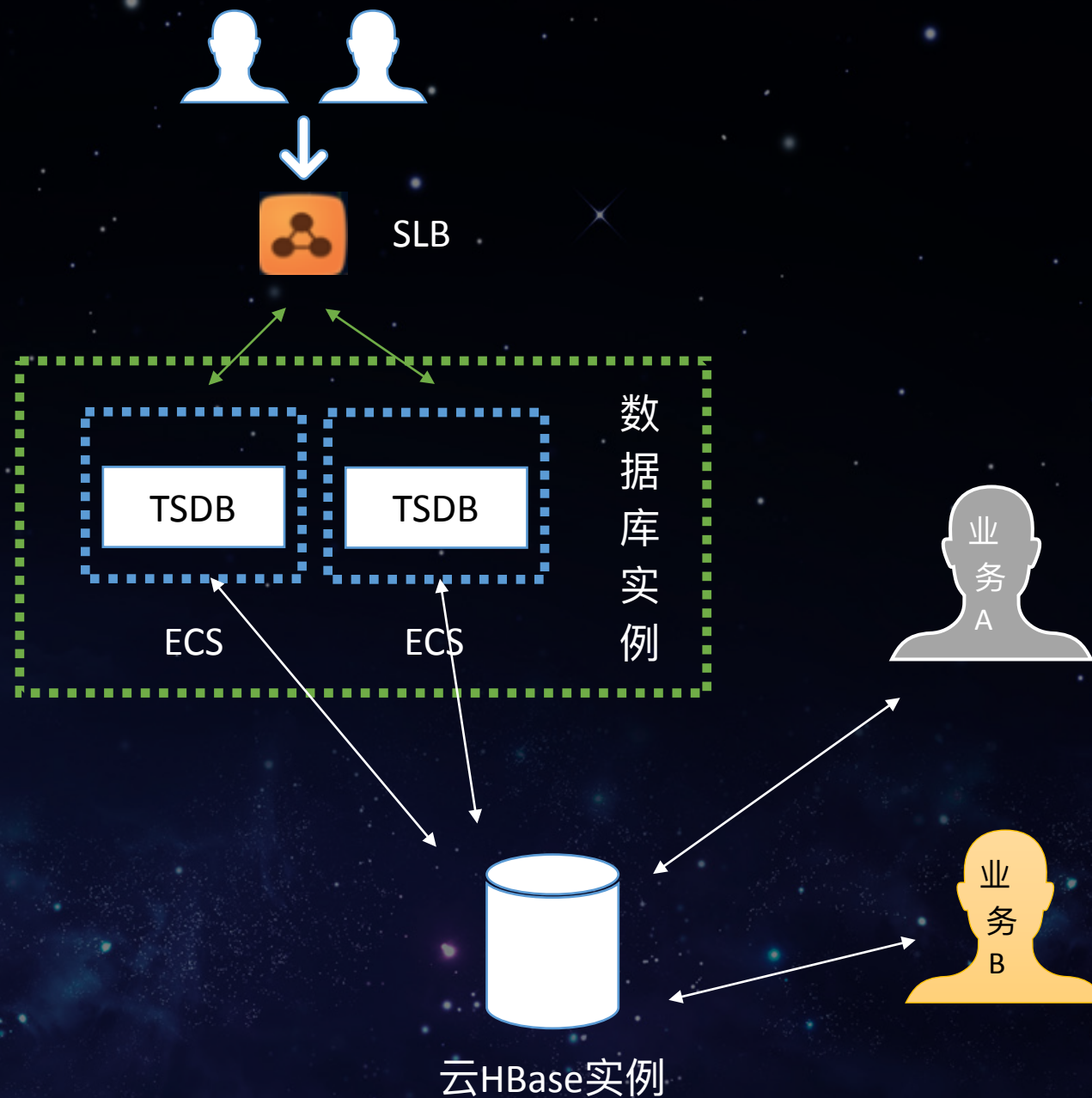




## 4 云OpenTSDB使用模式

**共享模式：** 复用已购买的云HBase

适合时序业务较小，或者用不满云HBase资源





CHTC  
中国HBase技术社区

云栖社区

# THANK YOU



社区管理员



HBase 技术社区公众号



HBase + Spark 钉钉社区大群

中国 HBase 技术社区网站：<http://hbase.group>