

Tencent 腾讯

ClickHouse 应用实践

丁晓坤 & 熊峰

2019.10.20@tencent

目录

- Clickhouse 的部署与监控管理
- Clickhouse 的应用实践 iData

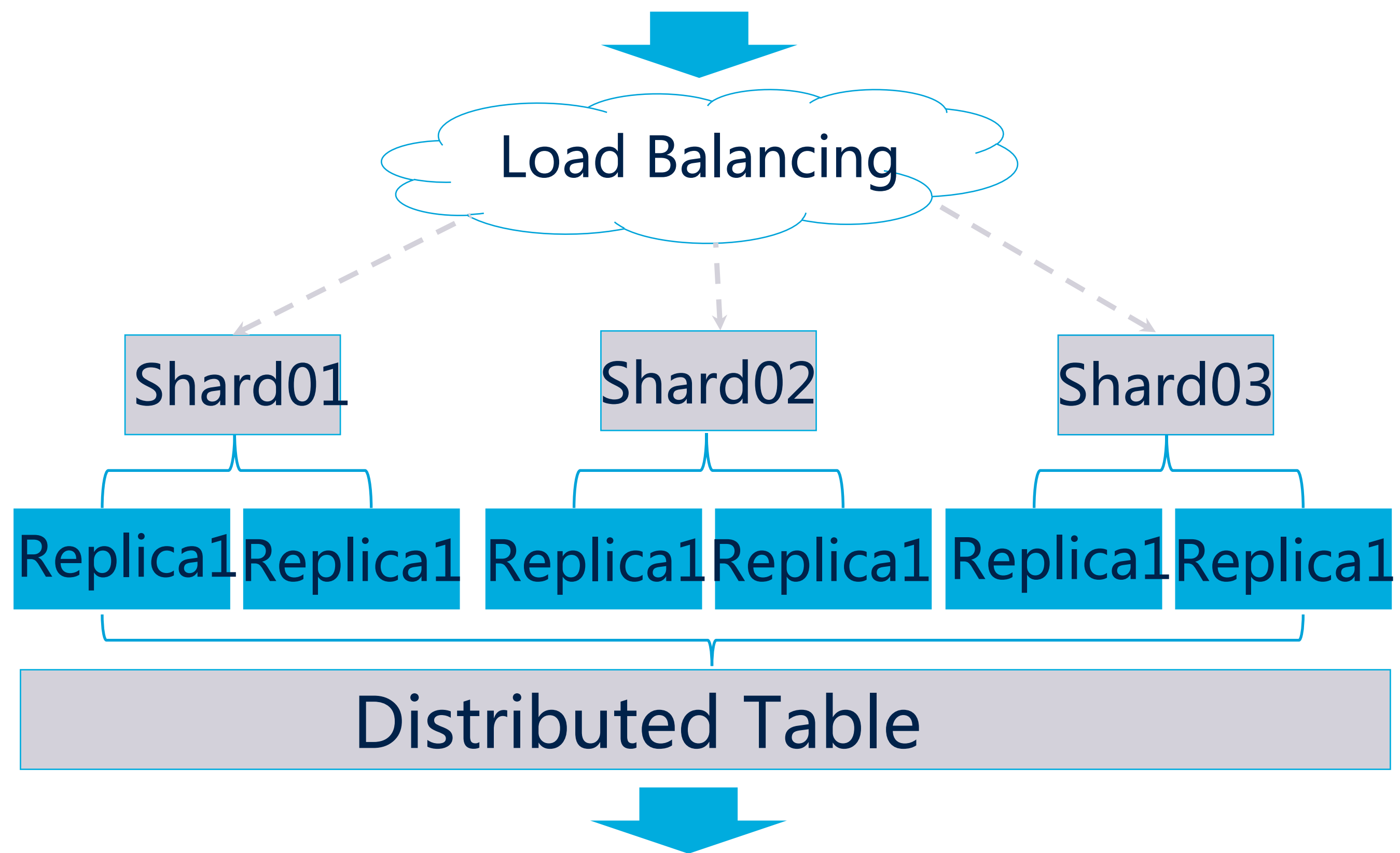
1 部署与监控管理

高内存，廉价存储：

单机配置:

Memory128G
CPU核数24
SATA20T , RAID5
万兆网卡

生产环境部署方案：



线性平滑扩容：

扩容：

- 1.安装新部署新的shard分片机器
- 2.新shard上创建表结构
- 3.批量修改当前集群的配置文件增加新的分片
- 4.名字服务添加节点

CSV 227 Gb, ~1.3 bln rows

```
SELECT passenger_count, avg(total_amount)
FROM trips GROUP BY passenger_count
```

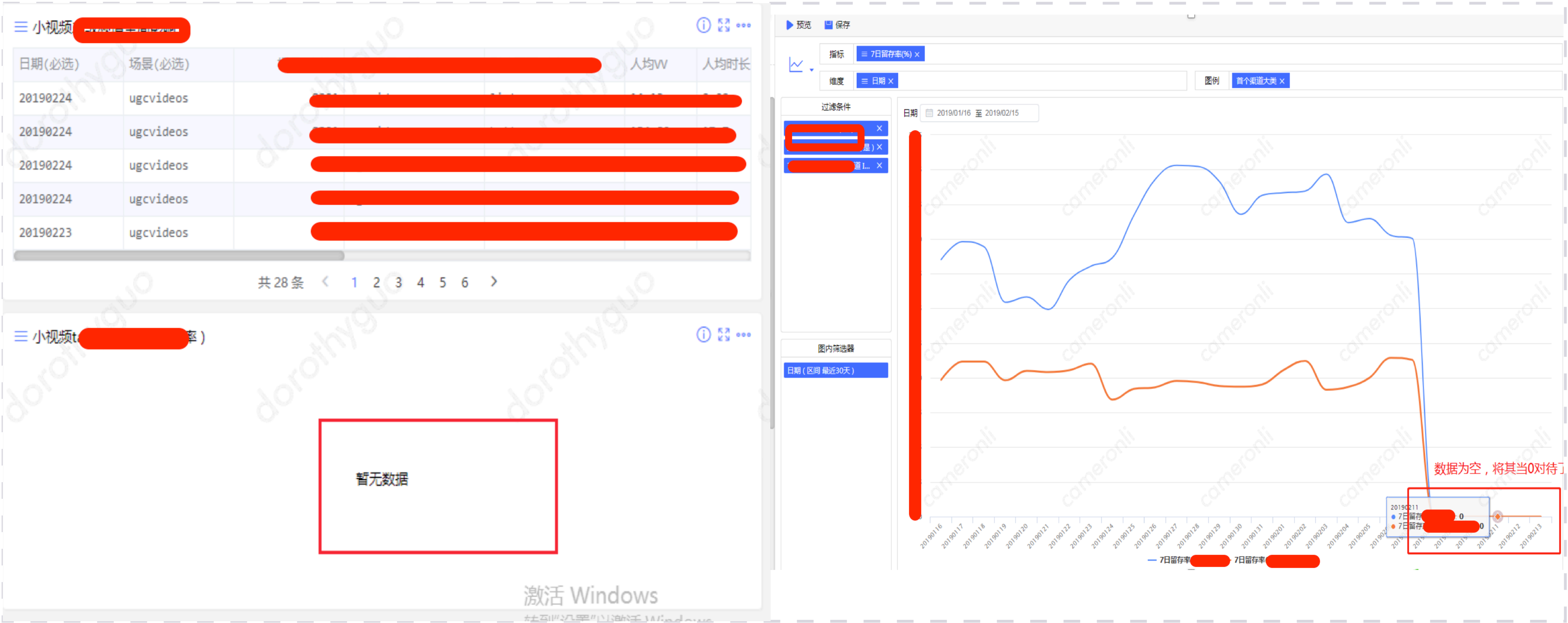
| Shards | 1 | 3 | 140 |
|----------|-------|-------|-------|
| Time, s. | 1,224 | 0,438 | 0,043 |
| Speedup | | x2.8 | x28.5 |

大批量，少批次

| WriteModel | BatchSize | RowLength | QPM | IOUtils | Partitions | FailedInserts |
|-------------|-----------|-----------|-----|---------|------------|---------------|
| SingleTable | 20000 | 1k | 65 | 40 | 680 | YES |
| SingleTable | 100000 | 1k | 21 | 29 | 215 | NO |
| SingleTable | 1000000 | 1k | 12 | 54 | 460 | NO |

| WriteModel | BatchSize | RowLength | QPM | IOUtils | Partitions | FailedInserts |
|------------|-----------|-----------|-----|---------|------------|---------------|
| MultiTable | 100000 | 100B | 390 | 51 | 602 | YES |
| MultiTable | 100000 | 1k | 21 | 29 | 215 | NO |
| MultiTable | 100000 | 10k | 9 | 49 | 413 | NO |

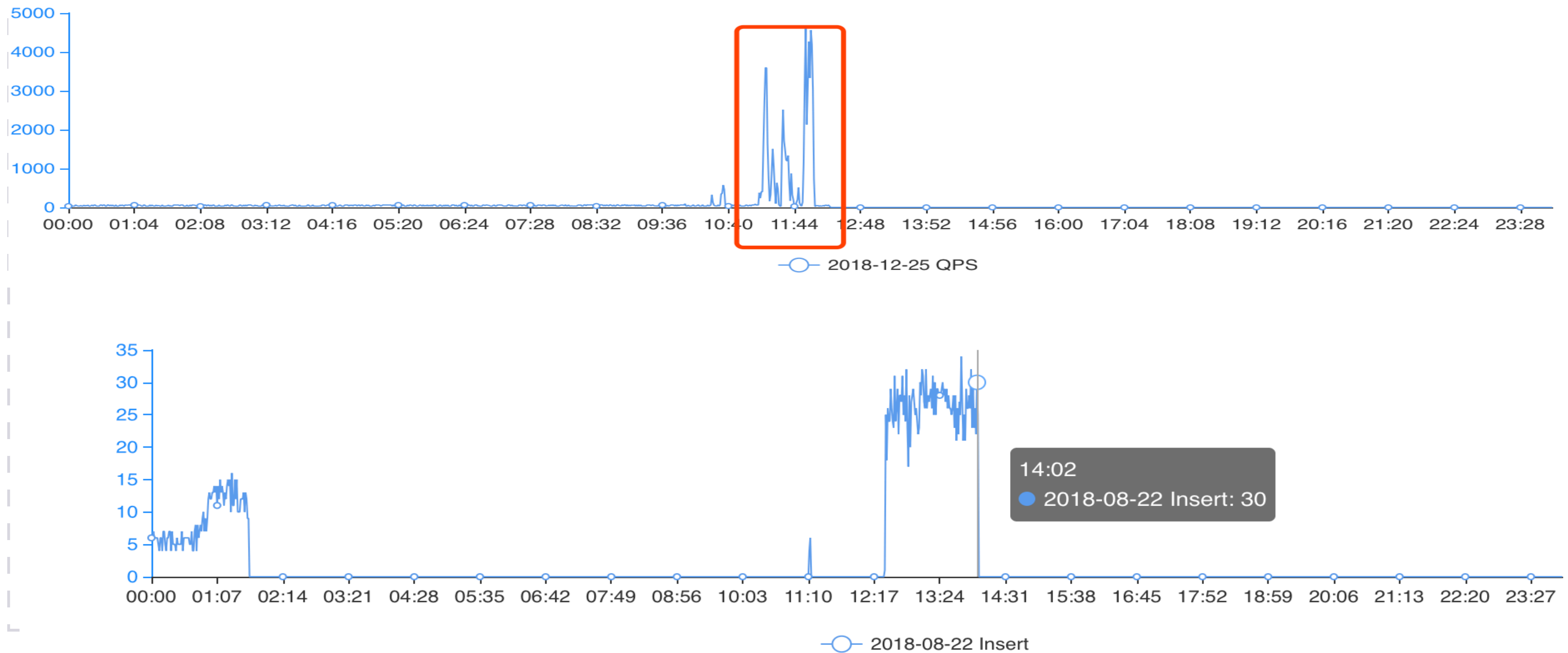
应用监控-业务指标：



服务监控-错误日志：



服务监控-请求指标：



服务监控-扫描详情：

| 节点 ▾ | User | 来源IP ▾ | ReadRows | TotalRows Approx | ReadBytes | MemoryUsage | PeakMemory | Query ▾ |
|---------------|---------|---------------|----------|------------------|-----------|-------------|------------|---|
| 192.168.20.24 | default | 148.204.7.148 | 3042 | 730240 | 0.31 GiB | 3.25 GiB | 3.28 GiB | select count(distinct if(start_play_num>0,omg_id,null)) as start_play_uv from where 1=1 and p_date between 2018-12-18 and 2018-12-24 and tab_name in (热点页卡) and p_date between 2018-12-18 and 2018-12-24 group by p_date,call_type FORMAT TabSeparatedWithNamesAndTypes |
| 192.168.20.24 | default | 148.204.7.148 | 4234 | 70240 | 0.93 GiB | 2.76 GiB | 3.03 GiB | select (distinct if(start_play_num>0,omg_id,null)) as start_play_uv from where 1=1 and p_date between 2018-12-18 and 2018-12-24 and tab_name in (热点页卡) and p_date between 2018-12-18 and 2018-12-24 group by p_date,platform FORMAT TabSeparatedWithNamesAndTypes |
| 192.168.20.24 | default | 192.168.20.16 | 3514 | 7264 | 0.03 GiB | 915.05 MiB | 2.49 GiB | SELECT p (exposure_num > 0, omg_id, NULL)) AS exposure_uv FROM WHERE (1 = 1) AND ((p_date >= 2018-12-18) AND (p_date <= 2018-12-24)) AND (tab_name IN 热点页卡) AND ((p_date >= 2018-12-18) AND (p_date <= 2018-12-24)) GROUP BY p_date, platform |
| 192.168.20.24 | default | 192.168.20.73 | 514 | 7264 | 0.01 GiB | 1.49 GiB | 2.60 GiB | SELECT p (exposure_num > 0, omg_id, NULL)) AS exposure_uv FROM WHERE (1 = 1) AND ((p_date >= 2018-12-18) AND (p_date <= 2018-12-24)) AND (tab_name IN 热点页卡) AND ((p_date >= 2018-12-18) AND (p_date <= 2018-12-24)) GROUP BY p_date, call_type |

服务监控-响应耗时：

| ip | por t | sql | 开始时间 | 结束时间 | 耗时ms | 状态 | 用户 | 查询应 用 | 错误日 志 | 操 作 |
|-------------|----------|---|------------------------|------------------------|-------|----|----|----------|----------|--------|
| 5.25. 67 | 81 23 | select p_date as p_date, video_music_id as video_music_id, sum(exposure_num) as exposure_num from kuaibao_analysis.yoo _content_model_use r_video_daily where 1=1 and p_date between '2018-12- 17' and '2018-12- 23' and p_date between '2018-12- 17' and '2018-12- 23' group by p_date,video_music_ id | 2018-12-24 10:52:16 | 2018-12-24 10:52:34 | 18029 | 成功 | | biPrs03 | - | |

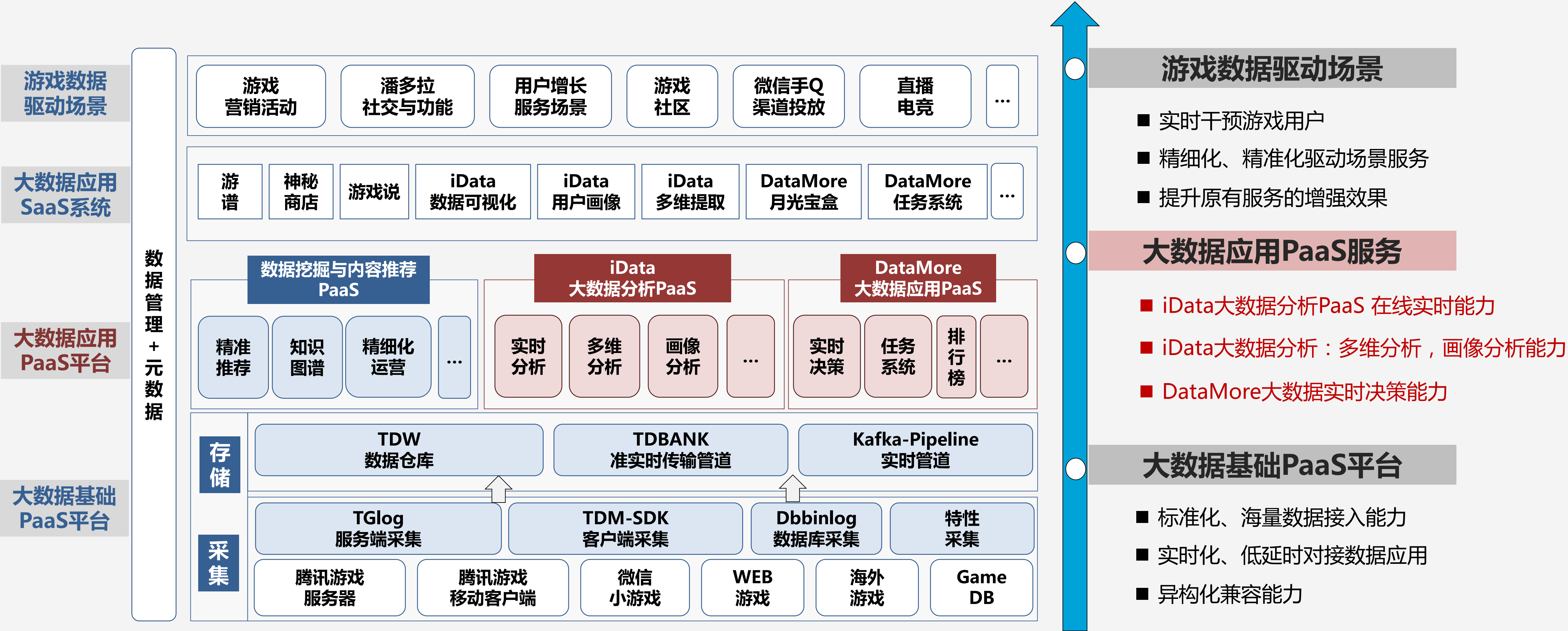
立体监控模型：

| 监控分层 | 监控项 | 敏感度 | 紧急度 |
|------|--------------|-----|-----|
| 应用层 | 业务指标，数据异常 | 低 | 高 |
| 服务层 | 错误日志 | 中 | 中 |
| | 请求指标 | | |
| | 扫描详情 | | |
| | 响应耗时 | | |
| 物理层 | 磁盘IO，持续负载，流量 | 高 | 低 |

2 业务应用实践 iData

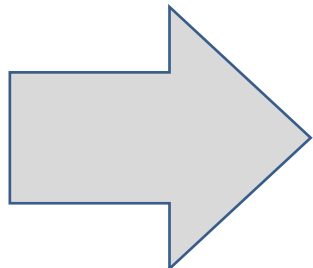
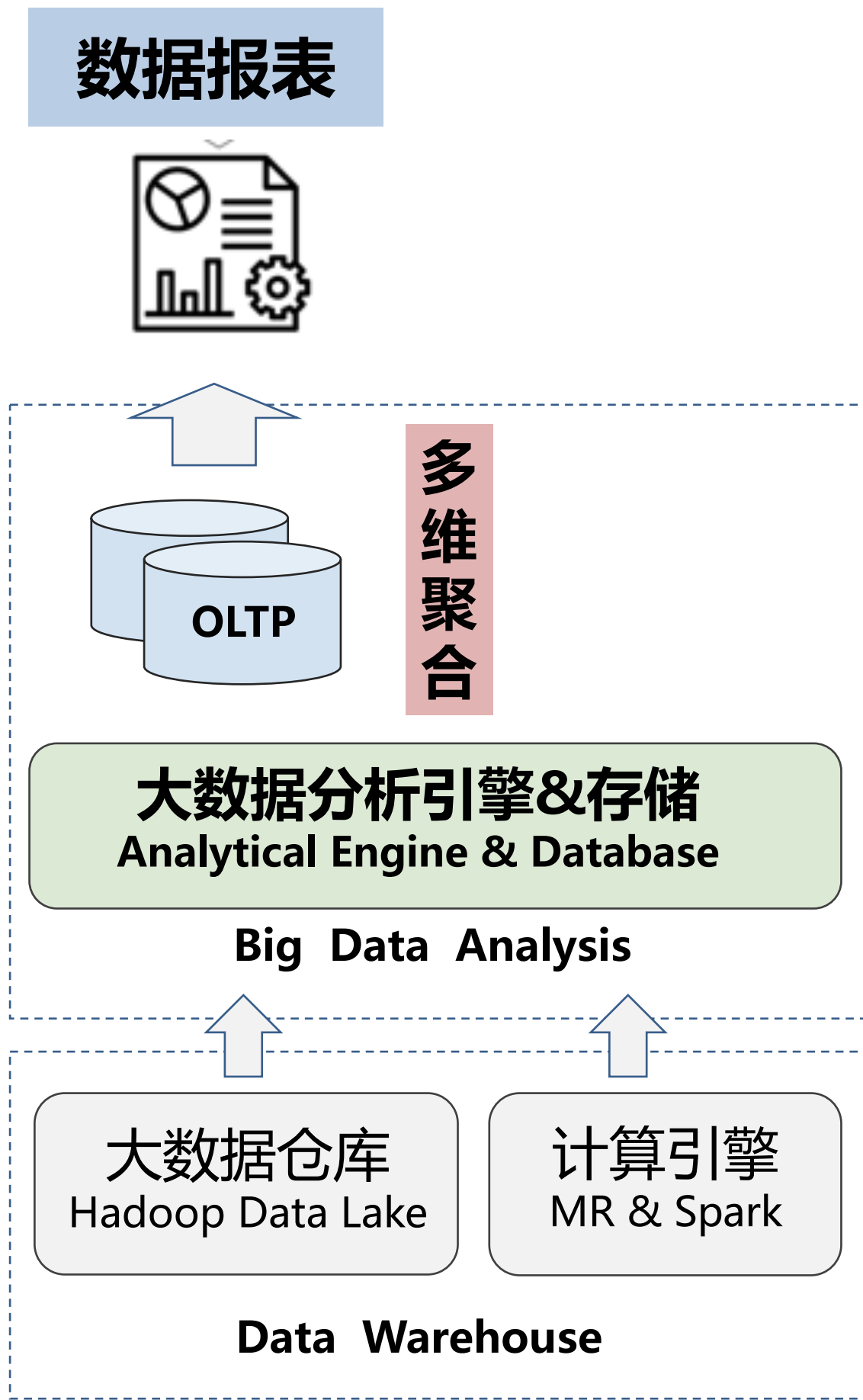
- 游戏数据分析的业务背景
- iData 数据分析引擎TGMars
- 为什么选用ClickHouse
- 平台在ClickHouse上的使用

腾讯游戏 数据化驱动服务 场景视图：

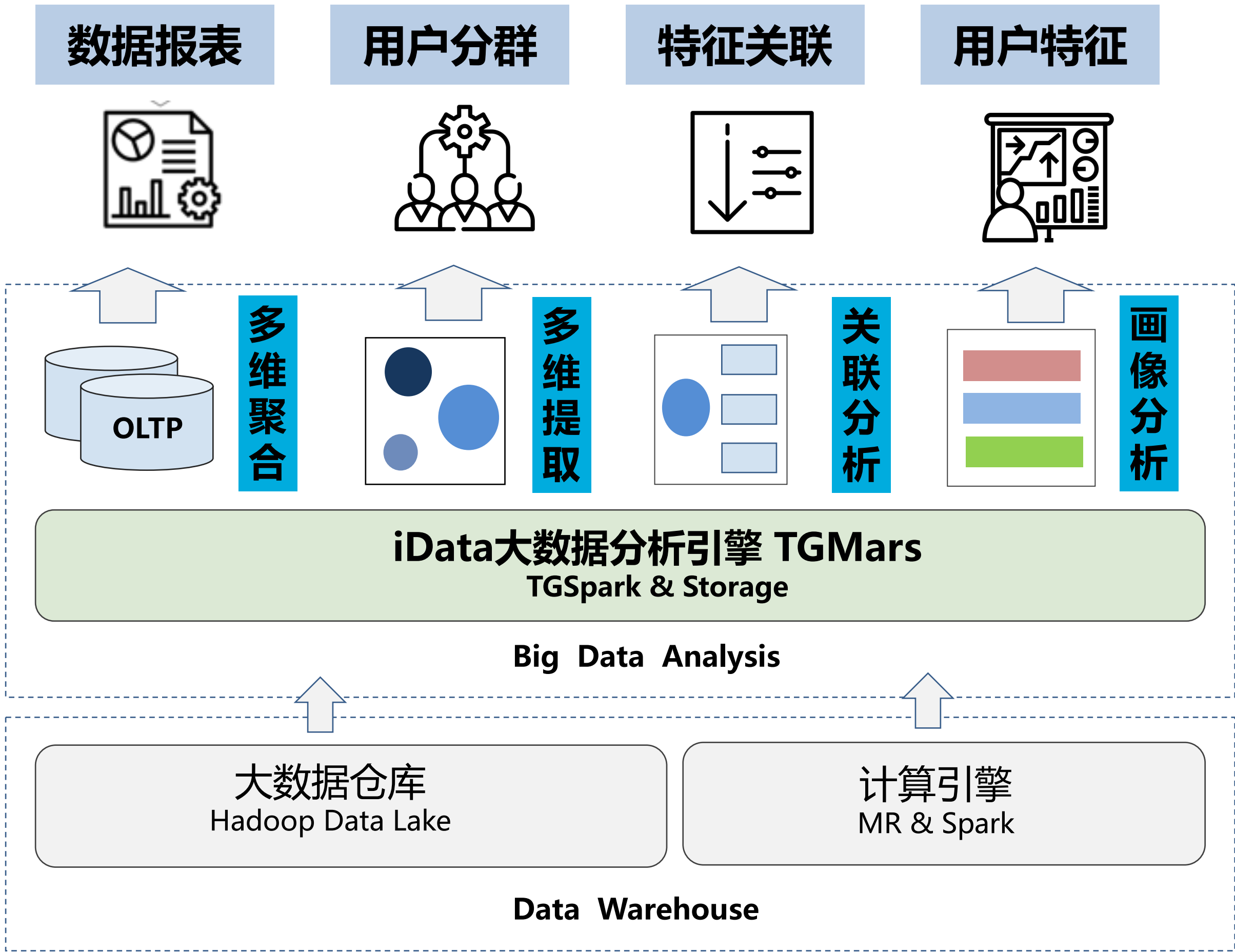


新大数据分析引擎2.0

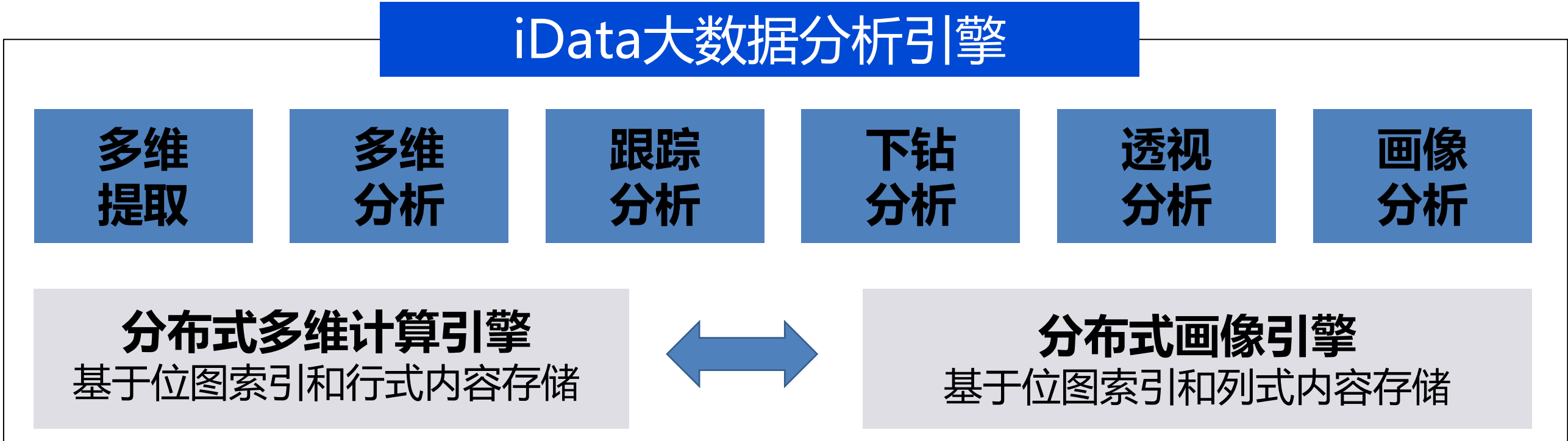
业界传统 大数据分析 引擎



iData New 超融合OLAP 大数据分析 引擎TGMars



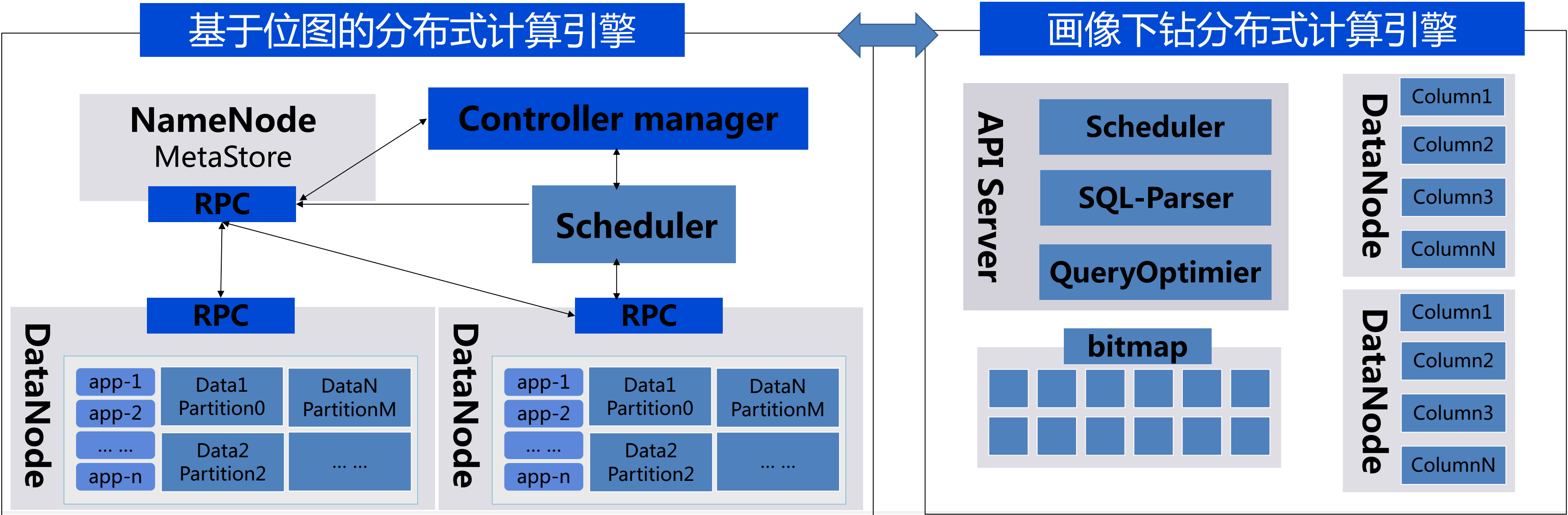
新大数据分析引擎2.0



分析，理应一气呵成
一键分析，提取/画像/跟踪

立即开始

| 任务名称 | 包类型 | 分析用户数 | 操作 | 创建时间 | 创建人 | 关注人 |
|---------------|-----|-------|--------------------|---------------------|----------|-----|
| 活跃用户_19060417 | QQ | 提取中 | 查看 | 2019-06-04 14:39:22 | vivianlv | 无 |
| 活跃用户_19060415 | QQ | 8,207 | 查看 回 画像/下钻 每日跟踪 下载 | 2019-06-04 14:38:26 | vivianlv | 无 |
| 活跃用户_19060412 | QQ | 8,207 | 查看 回 画像/下钻 每日跟踪 下载 | 2019-06-04 14:37:43 | vivianlv | 无 |
| 活跃用户_19060412 | QQ | 8,207 | 查看 回 画像/下钻 每日跟踪 下载 | 2019-06-04 14:37:40 | vivianlv | 无 |
| 活跃用户_19060409 | QQ | 8,207 | 查看 回 画像/下钻 每日跟踪 下载 | 2019-06-04 14:32:17 | vivianlv | 无 |
| 活跃用户_19060409 | QQ | 8,207 | 查看 回 画像/下钻 每日跟踪 下载 | 2019-06-04 14:32:14 | vivianlv | 无 |
| 新增用户_19060406 | QQ | 985 | 查看 回 画像/下钻 每日跟踪 下载 | 2019-06-04 14:29:42 | vivianlv | 无 |
| 新增用户_19060406 | QQ | 985 | 查看 回 画像/下钻 每日跟踪 下载 | 2019-06-04 14:29:39 | vivianlv | 无 |
| 活跃用户_19060404 | QQ | 8,207 | 查看 回 画像/下钻 每日跟踪 下载 | 2019-06-04 14:18:59 | vivianlv | 无 |
| 新增用户_19060401 | QQ | 172 | 查看 回 画像/下钻 每日跟踪 下载 | 2019-06-04 14:17:50 | vivianlv | 无 |



任务名称: 活跃用户_19060404

用户数 (QQ号码数): 8,207

用户包条件: A【我的五福】手Q(iOS+安卓) 2019-05-28至2019-06-03登录过的用户

关注人ID: 1105323348
任务ID: 686848

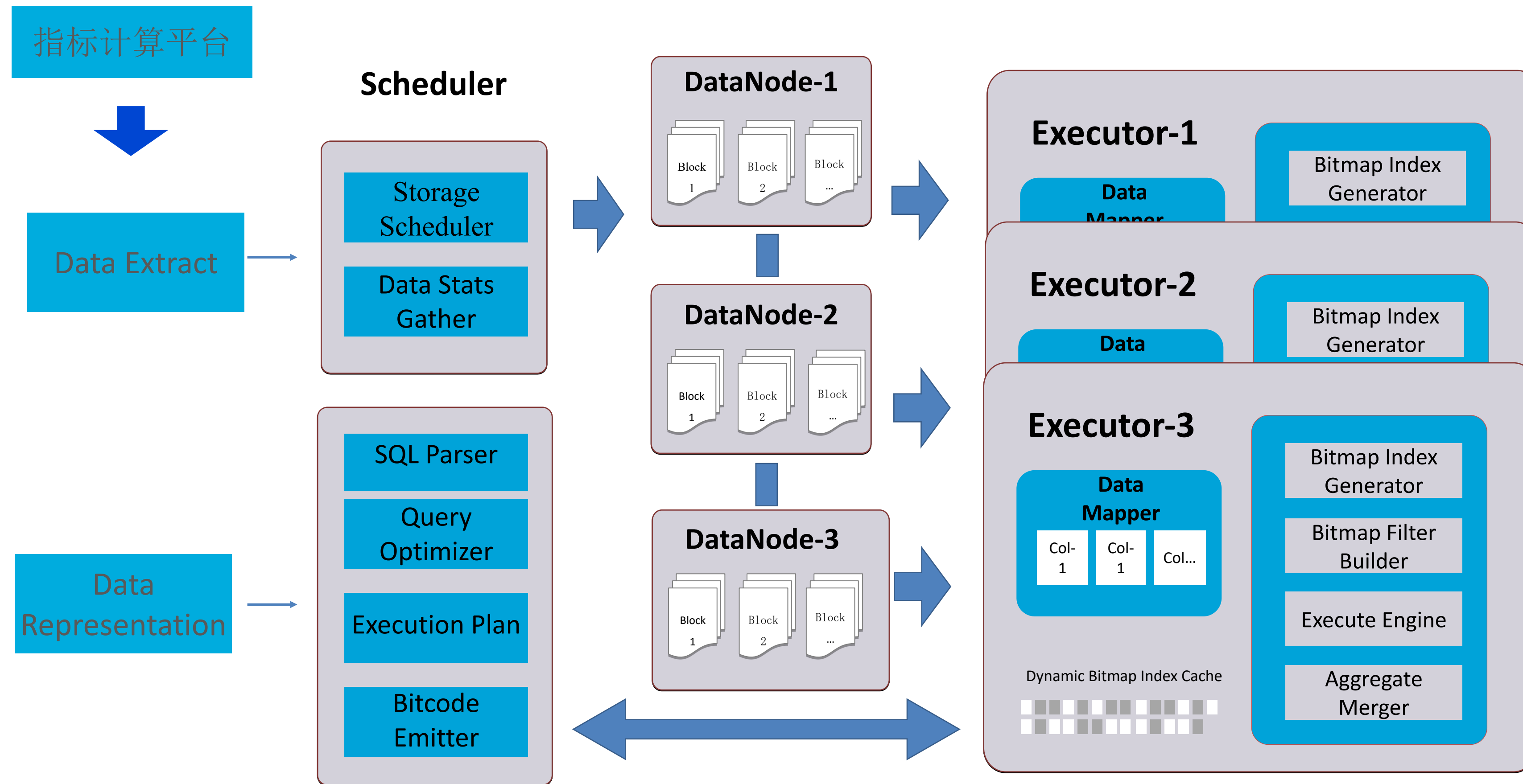
多维分析 2019-06-03 更新

分析用户数: 8,207

选择一个维度开始下钻分析

全部 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z 搜一下 +添加更多维度指标

旧画像系统



核心特点

- **分布式计算**
 - 并行计算
- **列式存储**
 - 按需加载减少IO
 - 可支持大量列
- **动态位图索引**
 - 缓存上次结果
 - 成本低、命中率高

iData画像服务需要升级

➤扩展性差

数据导入后结果不支持修改/追加

➤数据类型有限

数据类型只能支持数字类型

➤数据量有限

数据量达到10亿级以上查询效率有所降低

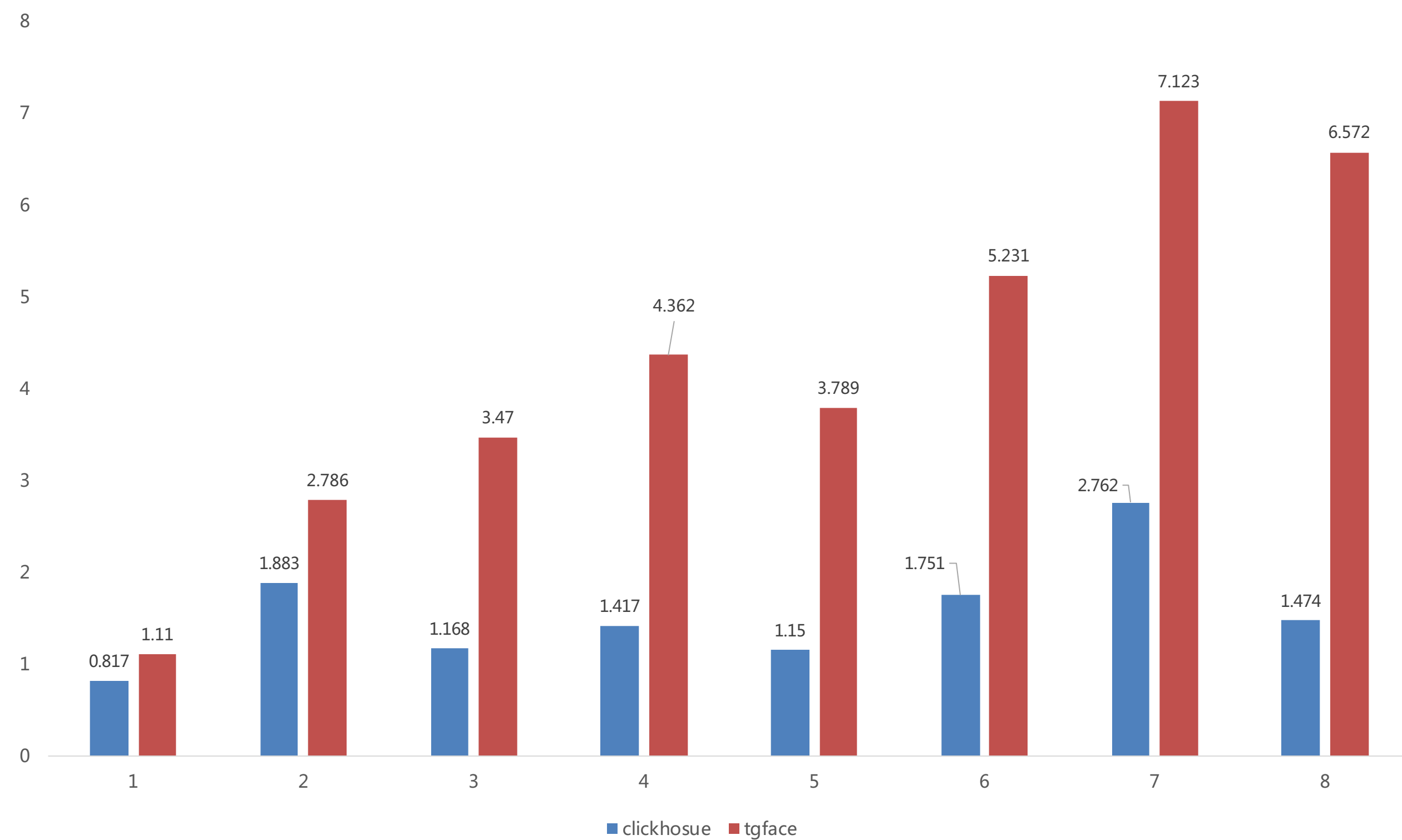
➤单表计算

不能进行多表关联计算

为什么选择ClickHouse

- SQL
- OLAP
- 超高性能
- 列式存储
- 统计函数
- 线性扩展
- 驱动丰富

4亿数据下钻耗时(单机)



如何使用ClickHouse满足特殊需求

- TDW HIVE SQL `{"2000352":"2","2":"4","50":"1","26":"28","31":"42"}`
- 转换成拓展的列 `{"26":"16"}`
- 嵌套数据类型 `{"26":"32","1":"2","2000209":"1"}`
- array 数据类型

| 2000352 | 2 | 30 | 26 | 31 | 26 | 1 | 2000209 |
|---------|---|----|----|----|----|---|---------|
| 2 | 4 | 1 | 28 | 42 | | | |
| | | | 16 | | | | |
| | | | 32 | | | 2 | 1 |

Map类数据处理方式

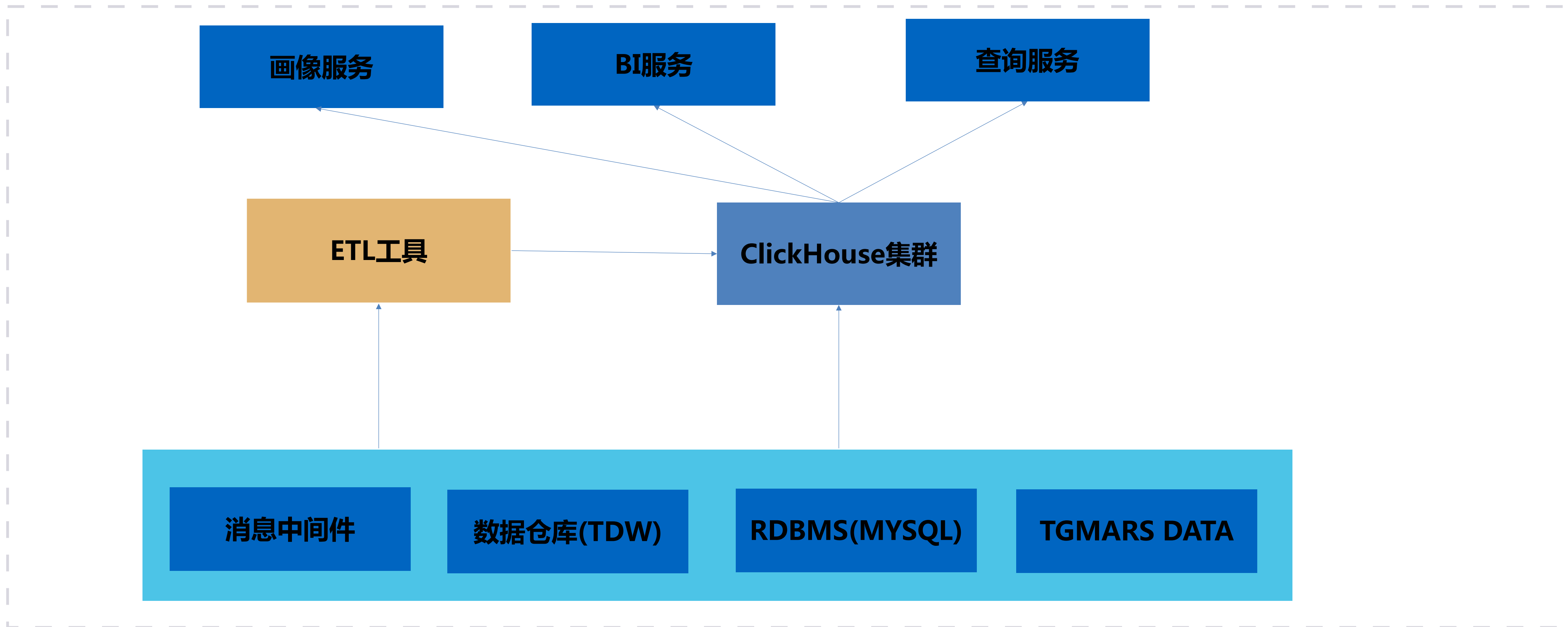
```
SELECT
  Goals.play_times_key AS key,
  sum(Goals.play_times_value) AS value
FROM wegame
ARRAY JOIN Goals
GROUP BY key
ORDER BY value DESC LIMIT 10
```

```
SELECT
  play_times_key AS key,
  sum(play_times_value) AS value
FROM wegame
ARRAY JOIN
  play_times_key,
  play_times_value
GROUP BY key
ORDER BY value DESC LIMIT 10
```

| key | value |
|---------|-----------|
| 26 | 997318243 |
| 1 | 270874452 |
| 2 | 121733714 |
| 30 | 25988721 |
| 2000352 | 20221399 |
| 17 | 12508995 |
| 43 | 11683279 |
| 31 | 9726032 |
| 48 | 9262996 |
| 50 | 7276426 |

10 rows in set. Elapsed: 0.320 sec. Processed 1.14 billion rows, 19.68 GB (3.58 billion rows/s., 61.55 GB/s.)

如何使用ClickHouse



Future

- 支持更多的机器学习算法
- 支持explain 查看整个sql 执行计划
- 集群管理

Thanks



腾讯招聘



腾讯大数据