Shopee

seaMoney

基于System Table监控查询和表

目录

System Tables 简介

System Tables 是 ClickHouse 提供的系统表.
用于记录:

- Server运行状态
- 内部Query处理情况
- 存储情况 等信息.

所有的System Tables 都存储在 system 这个数据库下面.
不可以被删除或修改.

system.query_log
system.tables
system.parts
...

目录

为什么要用 System Table 搭建看板

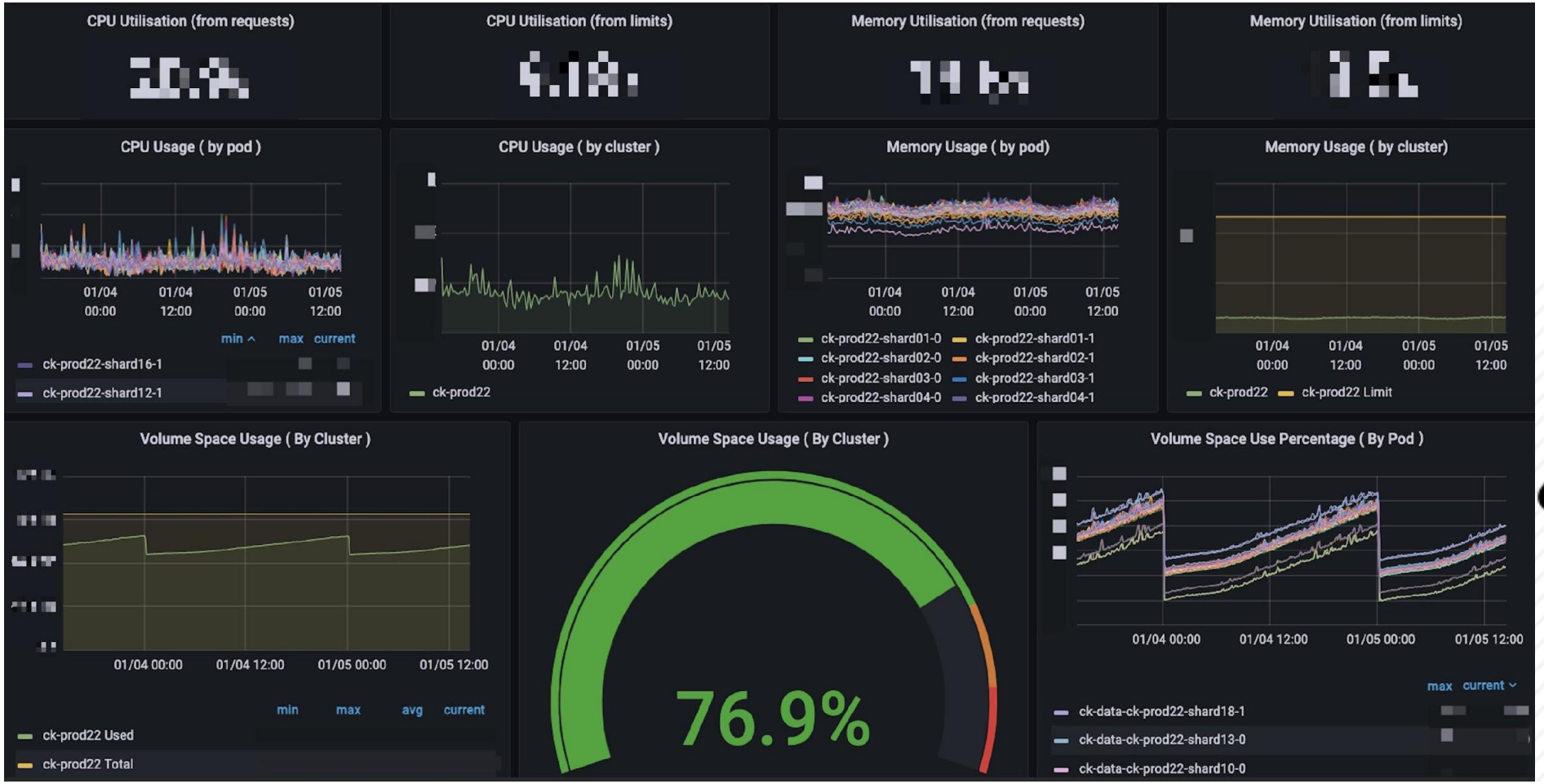
后端同学：服务不可用了，调用CK的API都超时了

分析同学：我着急出数，出报表，分析SQL跑不动了.

运维同学：数据库存储马上要超过安全水位线，必须改一些表的TTL或者删除数据.

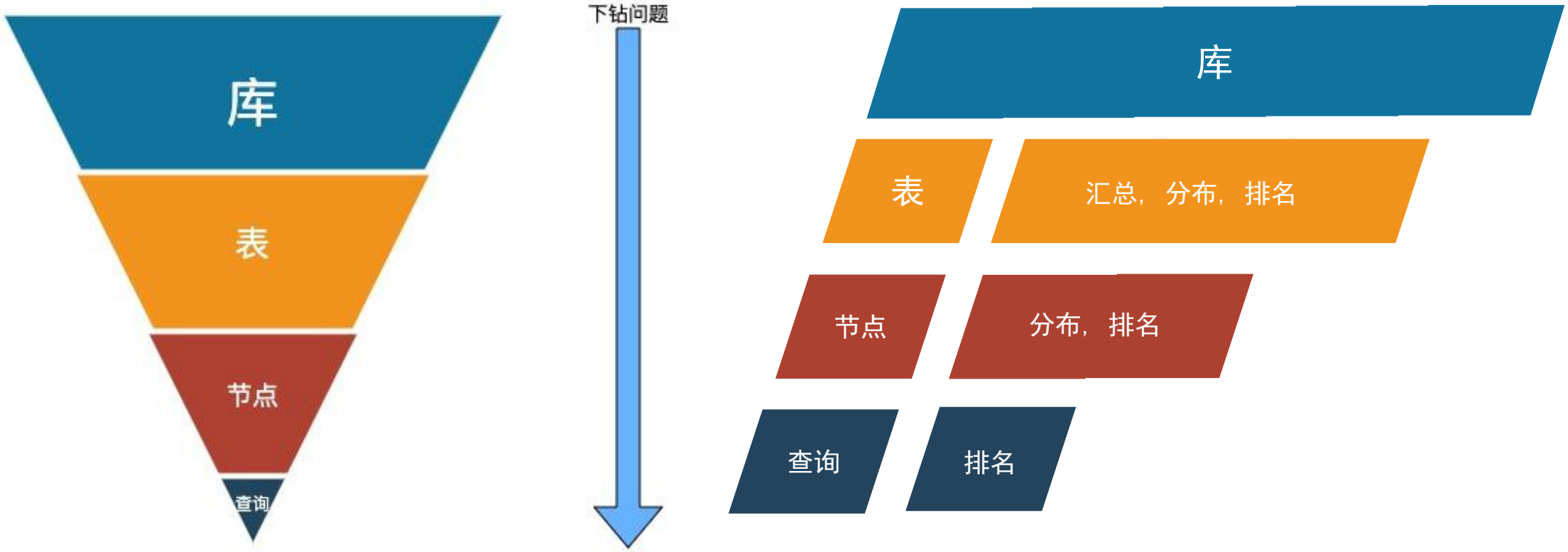
业务同学：我们也不知道哪张表重要，哪张不重要，没办法决定哪些表的TTL可以改或者删除.

为什么要用 System Table 搭建看板



目录

查询类看板介绍



基于 查询 的看板

如何获取相关指标


```
SELECT
 toStartOfxx (event_time) as t
  ,count(1) as select_query
 FROM clusterAllReplicas(cluster, system.query_log)
  where $timeFilter
  AND databases = [ 'data_base' ']
  AND query_kind = 'Select'
  AND type = 1/2
  AND is_initial_query = 1/2
  GROUP BY t
  ORDER BY t asc
```

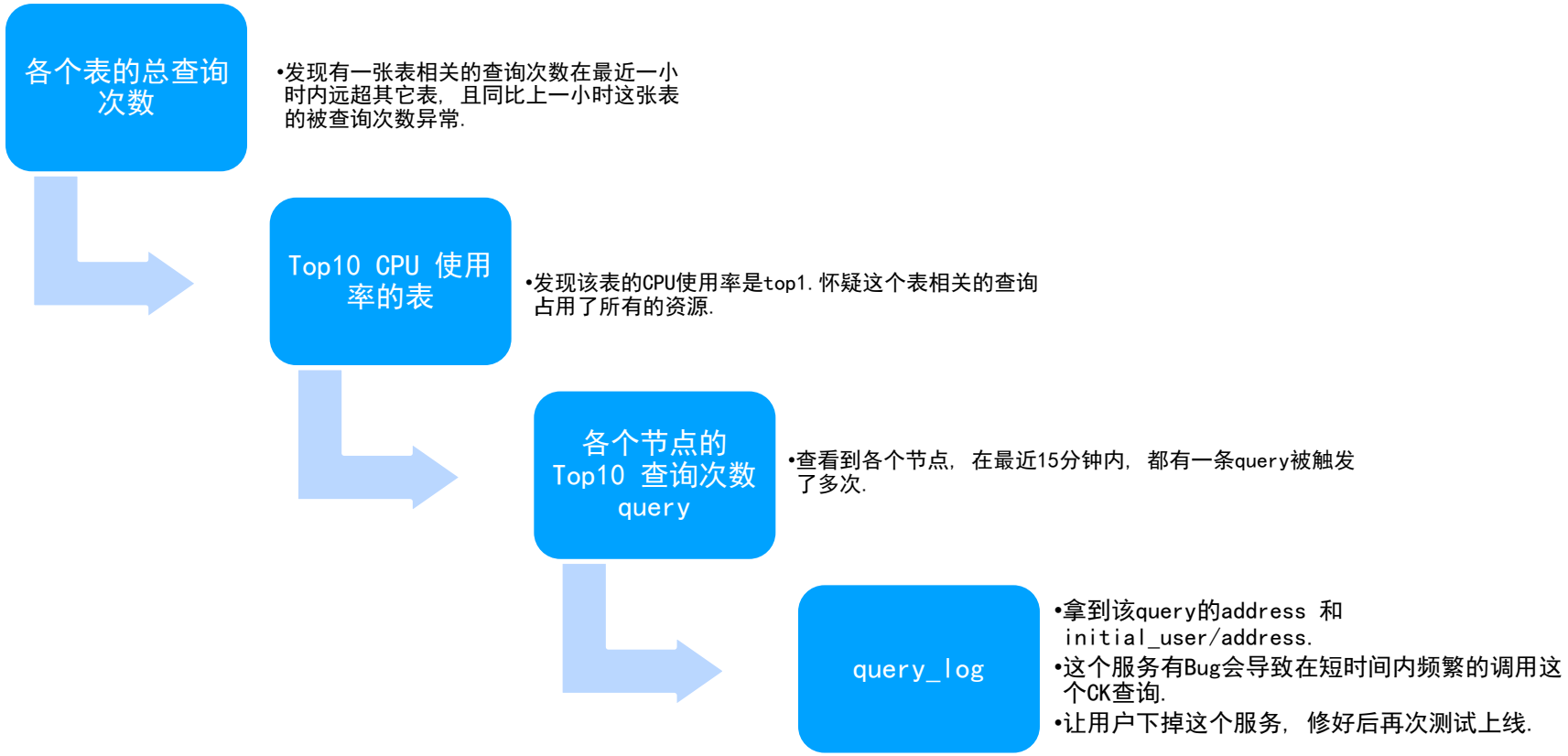
```
SELECT
FQDN(),
arrayJoin(tables) AS table_name,
sum(ProfileEvents['UserTimeMicroseconds']) AS
 total_cpu_time,
sum(read_bytes) AS total_memory
FROM clusterAllReplicas( 'cluster',
 system.query_log)
WHERE event_date = today()
   AND query_kind = 'Select'
   AND type = 'QueryFinish'
   AND is_initial_query = 1
 GROUP BY FQDN(), table_name
```

```
SELECT
 toStartOfxx (event_time) as t
  ,count(1) as select_query
 FROM clusterAllReplicas(cluster, system.query_log)
  where $timeFilter
  AND databases = [ 'data_base' ']
  AND query_kind = 'Select'
  AND type = 3/4
  AND is_initial_query = 1/2
  GROUP BY t
  ORDER BY t asc
```

```
SELECT
FQDN(),
arrayJoin(tables) AS table_name,
avg(query_duration_ms) AS avg_duration
FROM clusterAllReplicas( 'cluster',
 system.query_log)
WHERE event_date = today()
   AND query_kind = 'Select'
   AND type = 'QueryFinish'
   AND is_initial_query = 1
 GROUP BY FQDN(), table_name
```

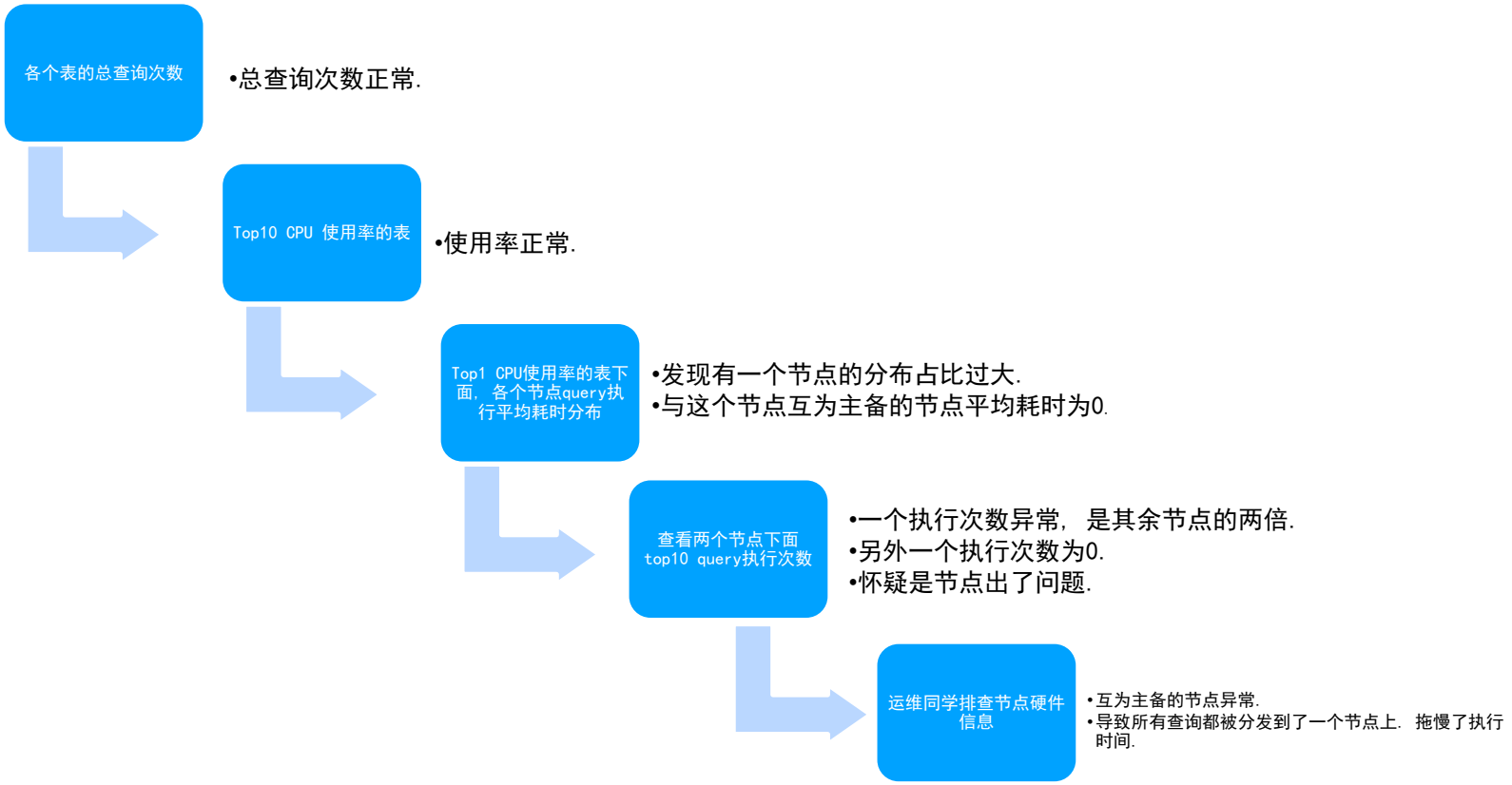
使用案例

1，集群CPU使用率飙升，调用CK的API SLA 出现大规模超时，分析同学反映SQL运行时间长.



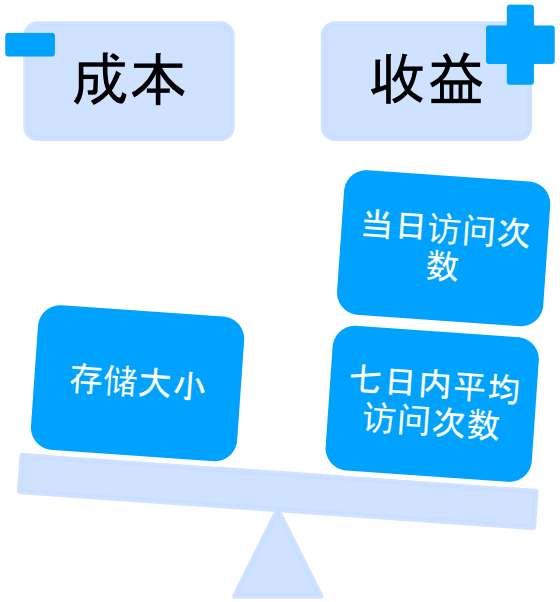
使用案例

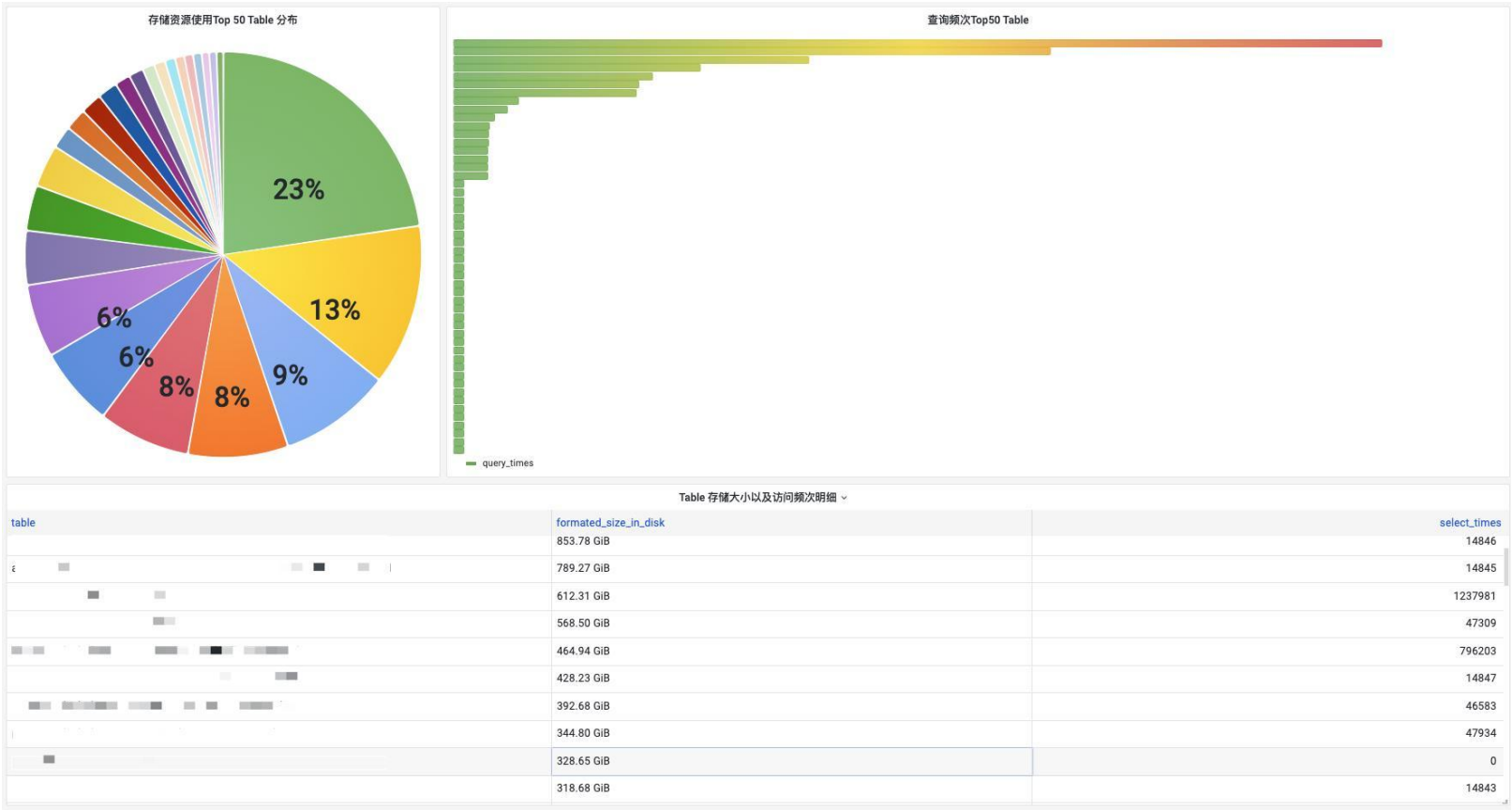
2，集群CPU使用率正常，调用CK的API SLA 出现大规模超时，分析同学反映SQL运行时间长.



目录

资源类看板介绍





收益： 当日查询总数 / 10,000 + 七日平均访问数 / 10,000.

成本： 存储大小 / 10G.

单表评分：收益 - 成本.

如何获取相关指标

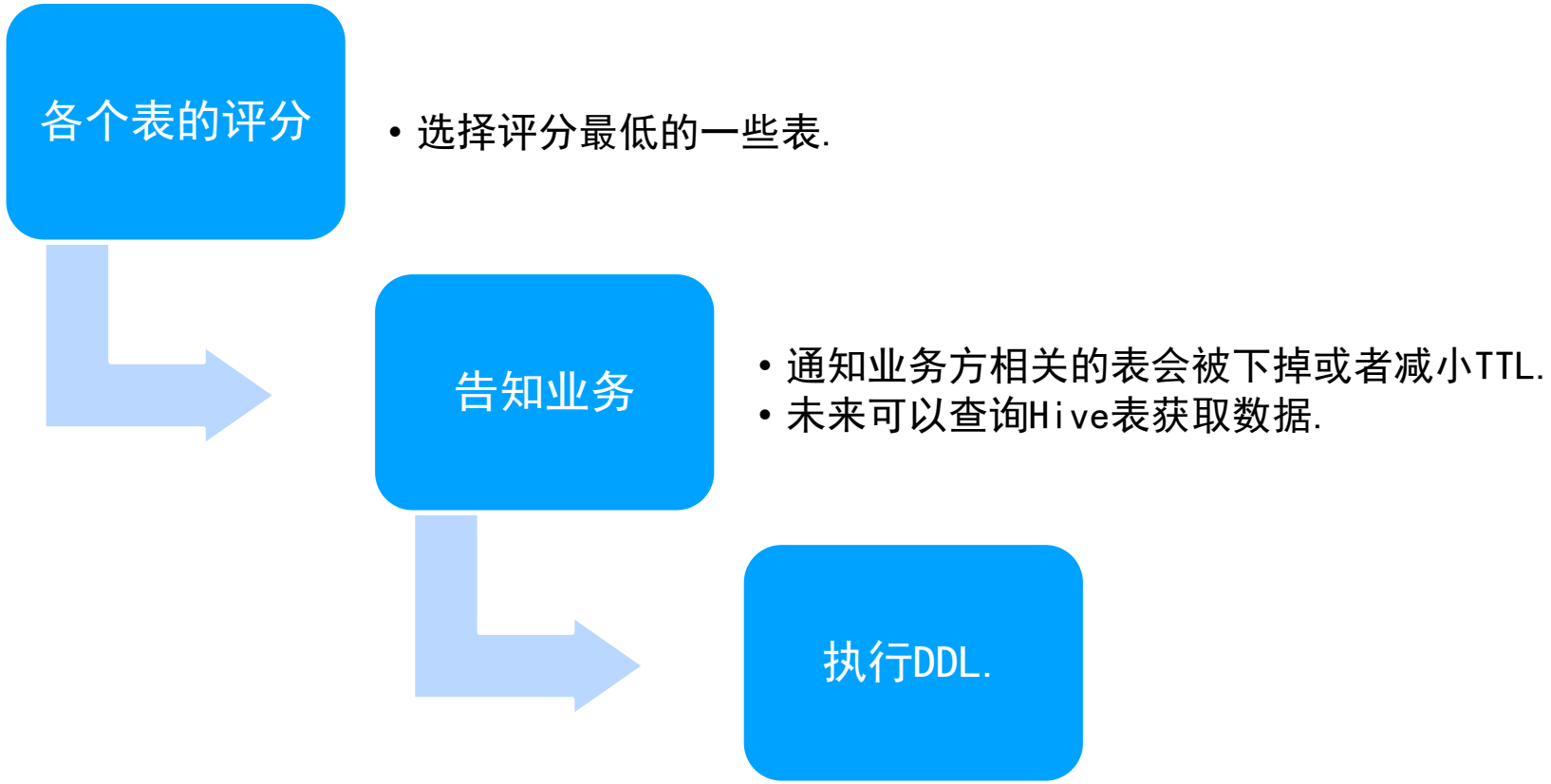
通过 system.query_log 来获取每张分布式表当天被查询的次数，以及七日平均查询次数.

通过 system.parts 来获取每张本地表表的存储大小.

通过 system.tables的dependencies_table来关联分布式表和本地表.

使用案例

1，集群磁盘使用率已经超过安全水位，运维同学要求调低一些表的TTL或者删除一些数据来保障集群的安全.



目录

SQL 调优分享

CK 有一个系统参数**distributed_product_mode, 有四种模式:**

1. **deny**
2. **local**
3. **global**
4. **allow**

以下面这个SQL为例来介绍这四种模式的不同:

```
SELECT A
FROM table_1_all
WHERE B in (
            SELECT distinct(B)
            FROM table_2_all
          )
```

distributed_product_mode = deny

会直接报错，不允许in / join 里面使用这种包含分布式表的查询.

distributed_product_mode = local

分发到所有节点上的sub query 子查询会变成:

```
SELECT A
FROM table_1_local
WHERE B in (
    SELECT distinct(B)
    FROM table_2_local
    )
```

导致查询结果不准.

distributed_product_mode = allow

分发到所有节点上的sub query 子查询会变成:

```
SELECT A
FROM table_1_local
WHERE B in (
    SELECT distinct(B)
    FROM table_2_ALL
    )
```

然后各个节点又会向所有节点发送:

```
SELECT distinct(B)
FROM table_2_local
```

的子查询, 导致读放大.

distributed_product_mode = global

自动使用了global 来修饰 in.

```
SELECT A
FROM table_1_all
WHERE B global in (
    SELECT distinct(B)
    FROM table_2_all
    )
```

先向所有节点发送:

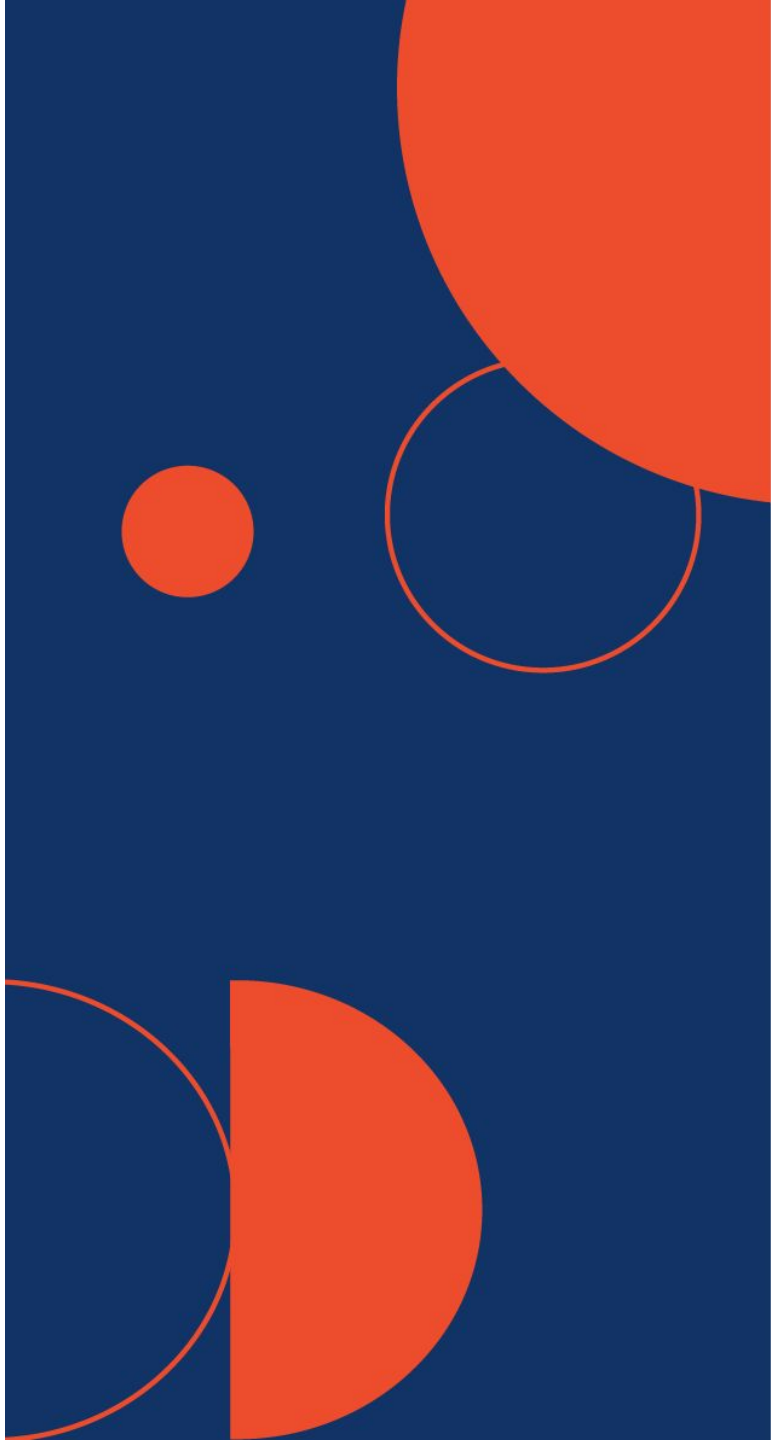
```
SELECT distinct(B)
FROM table_2_local 的子查询
```

将所有子查询的结果汇总到一个节点.
形成一个临时结果集: temp_results_123

然后向所有节点发送:

```
SELECT A
FROM table_1_local
WHERE B in (temp_results_123)
```

Q & A



Thank You!